BOOTSTRAPPING SINGLE LQUATION REGRESSION MODELS: SOMĿ FINITE SAMPLE RESULTS

Ah Boon Sim

A Thesis

in

The Department

of

Economics

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

April 1989

ISBN 0-315-51323-3

ABSTRACT

BOOTSTRAPPING SINGLE EQUATION REGRESSION MODELS:
SOME FINITE SAMPLE RESULTS

Ah Boon Sim, Ph.D.
Concordia University, 1989

The bootstrap is becoming a powerful tool in regression analysis. It can be used as a bias correction procedure. It can also be used to construct confidence intervals for the unknown parameters. However, many studies have found that the empirical coverages of these intervals are significantly smaller than their nominal values. This study provides an insight into the above problem and proposes a few solutions. The concept of a 'selection' matrix is introduced in this dissertation. Throughout this study, various properties of the selection matrix are derived and applied to different regression problems. Several important results concerning the bootstrap estimates of regression coefficients are also obtained.

DEDICATION

This dissertation is dedicated to my parents.

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to all those who have helped toward the writing of this dissertation. In particular, my sincere thanks to the Late Professor Balvir Singh for bringing to my attention the concept of bootstrap and its applications in econometric problems. My participation in a project which applied the bootstrap to asset pricing models, using both Canadian and American financial data reinforced my interest in the bootstrap method. This project was funded by a SSHRC grant held jointly by Professor Lawrence Kryzanowski and Dr. Minh Chau To. I also benefited from working as a research assistant to Professor Muni S. Srivastava, who was working jointly with Professor Singh on applications of the bootstrap to econometric problems. I would also like to express my sincere gratitude to my thesis supervisor, Professor Gordon Fisher, who in early 1986 suggested that this thesis should focus on theoretical aspects of the bootstrap.

Many thanks also go to Dr. Yogendra P. Chaubey for his many valuable comments. I would also like to thank Professor Lawrence Kryzanowski, Dr. Jon Breslaw and Dr. Ian Irvine for their comments. For encouragements and support, my sincere appreciation goes to the graduate students, faculty and staff members of the Economics department with special thanks to the graduate secretary, Mrs. Helen Twerdun and to fellow graduate students Andy Kwan, David Olesik and Aminu Said.

I am deeply grateful to my wife, Maria Auxiliadora Chong de Sim, who typed this difficult manuscript with incredible swiftness and accuracy. Finally, I would like to thank Dr. Abdul Rahman, Dr. Zeng Rung Liu, my parents and parents-in-law for their moral support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION AND AN OVERVIEW

1.1 Bootstrap: The Euphoria

During the earlier years when the bootstrap method was first introduced, there was much euphoria over its general applicability to a wide range of statistical and econometric problems. Efron (1979), in an effort to show the superiority of the bootstrap method in comparison with the jackknife, labels the bootstrap method as being more widely applicable and more dependable than the jackknife. This pioneering paper started the initial euphoria and there was much enthusiasm in applications of Efron's bootstrap.

The first reaction to this new development, which was just another by-product of the computer revolution, also led to investigations into its asymptotic properties. Asymptotically, under general conditions, it was found to have nice properties. However, its finite sample properties are generally unknown. In many applications, researchers utilized Monte Carlo simulations to obtain some estimates of its finite sample properties. Much to their disappointment, the bootstrap seems to perform poorly. Many suggestions were subsequently put forward, mainly as a result of Monte Carlo simulation studies. Nevertheless, only a few of these suggestions have a sound theoretical basis.

Unlike the literature on the jackknife, which is well developed and understood, the literature on the bootstrap is often confusing. This is noted in Wu (1986). Although the euphoria has since faded away, there is still much interest in applications of the bootstrap. The bootstrap must be recognized as a potentially very powerful tool for statistical analyses. However, in order for the bootstrap to be useful, one must find appropriate ways to apply it. This should be the direction for future research. This dissertation represents an attempt to work in this direction.

1.2 Bootstrap and Regression Models

Efron (1979) advocates the use of a sampling-with-replacement method in statistical computations, and this gave birth to the concept known as the "bootstrap". It was presented as a refinement of the Quenouille-Tukey jackknife. Why the term "bootstrap"? The connotation of this name for a statistical method may be as astounding as its predecessor, the "jackknife". In statistics, the "jackknife" gets its name from the household jackknife, which is a simple household tool. This name is used to reflect the fact that the "jackknife" is a simple statistical tool. In bootstrapping, only the observed data are required and no other extraneous data are needed. Thus, one is in fact pulling oneself up by the bootstraps. This is how the "bootstrap" gets its name.

The bootstrap method provides a useful tool for estimating the sampling properties of a given statistic. This is done without prior knowledge of the parent distribution of an observed sample. To a certain extent, this is similar to the jackknife method. For this reason, both the jackknife and the bootstrap methods are often referred to as "distribution-free methods". A statistical method is distribution-free provided its application is valid, regardless of the underlying distribution. However, the accuracy of Efron's bootstrap does vary with the class of statistics and with the underlying probability distribution.

Asymptotic properties of the bootstrap were first studied by Bickel and Freedman (1981), Freedman (1981) and Singh (1981). While both Bickel and Freedman (1981) and Singh (1981) focussed on the general properties for a wide class of statistics, Freedman (1981) focussed on the application of bootstrap to regression models. Various extensions and applications have been made. Recent papers include Wu (1986), Diciccio and Tibshirani (1987), Efron (1987), Loh (1987), Rao and Wu (1988) and Tibshirani (1988).

Inferences on the application of bootstrap to regression models are mainly based upon Monte Carlo simulations. These can be found in Freedman and Peters (1984a and 1984b) and Wu (1986). On the other hand, a coherent finite sample theory of the bootstrap in the context of regression model is nonexistent.

In Chapter 2, some finite sample properties of the means and variances of bootstrap estimates of regression coefficients are obtained analytically. This is done for the case of a simple linear regression (LR) model. In most applications, OLS residuals are used for bootstrapping. Probably, this is because these residuals are simple to compute. In the literature, Freedman and Peters (1984a) have suggested the use of either BLUS [see e.g., Theil (1965)] or inflated OLS residuals for bootstrapping a LR model. Inflated OLS residuals are OLS residuals multiplied by a factor depending on the sample size and the number of regression coefficients to be estimated, such that the second sample moment of the inflated residuals is an unbiased estimate of the second moment of the true errors. On the other hand, BLUS is the acronym for best linear unbiased with a scalar variance-covariance matrix. However, this suggestion goes unheeded, as no coherent finite sample theory is available to support (or reject) the use of either or both types of residuals. The results in Chapter 2 support the use of either BLUS or inflated OLS residuals for bootstrapping, depending on the sample size and on the number of regression coefficients. These results reject the use of OLS residuals, especially when the sample size is small.

1.3 Bootstrap and the Selection Matrix

The LR model is defined in Section 2.2 and the results in Chapters 2 through to 4 are based upon this model. The derivations of these finite sample results are made possible with the use of a selection matrix. This selection matrix is defined in Section 2.4. Basically, an mxn selection matrix is a binary matrix which randomly selects m elements with replacement from a set of n elements, and each of the n elements has the same probability (n^{-1}) that it will be selected. Each element takes on the value 0 or 1, and it has probability (n^{-1}) that its value will be 1.

The purpose of the selection matrix is to select, just as its name suggests, at random an element from a set of n elements. The selected element is then returned to the original set and at the same time, that element is recorded in the new set. The column in the first row of the selection matrix corresponding to the location of the selected element in the original set is then assigned the value 1, and zero is assigned to the remaining columns of the first row. The same procedure is repeated for the second row. For an mxn selection matrix, the above procedure must be repeated m times.

Various properties of the selection matrix are derived, as they are needed, in Chapters 2 through to 6; except for Chapter 4, which makes use of theorems already obtained in Chapters 2 and 3. Although all lemmas and theorems relating to the selection matrix are equally important with respect to the main results, the fundamental ones can be found in Chapters 2 and 3.

Application of the selection matrix to bootstrapping is one innovation of this disertation, a term coined during a private discussion with my thesis supervisor, Professor Gordon Fisher. Its properties may be difficult to understand and for this reason, alternative proofs for two of its main lemmas are given in Appendix A. It is also discovered that some of its properties are related to that of a binomial probability distribution. It must be mentioned that the selection matrix is an integral part of the results obtained in the following chapters.

1.4 Bootstrap Estimates of Regression Coefficients

One of the objectives of Chapter 2 is to investigate whether OLS residuals are suitable for bootstrapping when the sample size is small. Asymptotically, Freedman (1981) has shown that OLS residuals can be used for bootstrapping. It is now common knowledge among many authors that

OLS residuals are not suitable when the sample size is small. However, most of the inferences come from Monte Carlo simulations. Thus, the cause of the problem(s) associated with the use of OLS residuals for bootstrapping is still unknown. A few authors have suspected that the problem stems from the fact that the OLS residuals do not have a scalar variance-covariance matrix. [See e.g., Stine (1985).] Nevertheless, it is shown that this conjecture is incorrect. On the other hand, the suggestion that BLUS residuals be used for bootstrapping is valid.

When OLS residuals are used for bootstrapping, it is shown in Chapter 2 that the arithmetic mean of bootstrap estimates of β (a regression coefficient) is an unbiased estimate of its exact value, provided that $\hat{\beta}$, the OLS estimate of β, is also unbiased and provided that the regression model has an intercept. When the regression model does not have an intercept, the OLS residuals must be centered to mean zero such that the sample mean of these residuals is zero. This is true for any type of regression residuals. It is also shown in Chapter 2, that whenever the sample mean of regression residuals used for bootstrapping is zero, the arithmetic means of bootstrap estimates of regression coefficients approach their OLS estimates as the number of bootstrap replications goes to infinity. This observation applies to a general class of regression residuals. Thus, the arithmetic mean of bootstrap estimates of a regression coefficient is not sensitive to the type of regression residuals used for bootstrapping.

On the other hand, higher sample moments of bootstrap estimates of regression coefficients are not invariant to the type of regression residuals used for bootstrapping. For a general class of error distributions, bootstrapping based upon OLS residuals will in general lead to underestimation of higher moments of OLS estimates of the regression coefficients. In particular, if K is the number of regression coefficient, the sample variance of a bootstrap estimate of a regression coefficient will underestimate its exact value by the factor (Kn^{-1}). Also, bootstrapping based upon either BLUS or inflated OLS residuals leads to unbiased estimates of the second moment of $\hat{\beta}$. This is also shown in Chapter 2.

The main cause of the poor performance associated with OLS residuals can be attributed to the fact that the sample variance and higher sample moments of the OLS residuals underestimate the error variance and higher moments of the error distribution, respectively. This problem can be partially alleviated by multiplying the OLS residuals by the factor $\left[n(n-K)^{-1}\right]^{1/2}$. With this correction factor, one would be able to obtain an unbiased estimate of the error variance. In fact, it is shown that the bootstrap estimate of $D(\hat{\beta})$, the dispersion matrix of $\hat{\beta}$, approaches $D(\hat{\beta})$ as the number of bootstrap replications goes to infinity.

The preceding correction procedure is only valid up to the second moment of $\hat{\beta}$. This alleviates slightly the underestimation of higher moments of $\hat{\beta}$ but the underestimation that remains can cause serious problems. Another alternative is to use Theil's (1965) BLUS residuals. When the objective is to obtain bootstrap estimates of the first two moments of $\hat{\beta}$, bootstrapping based upon either inflated OLS residuals or the class of BLUS residuals yields identical results. On the other hand, when the objective is to obtain bootstrap estimates of higher moments of $\hat{\beta}$, it is better to use BLUS residuals for bootstrapping provided that the number of regression coefficients is large relative to n. When K is small, the difference between the two sets of bootstrap estimates is negligible. However, BLUS residuals are cumbersome to obtain. Thus, when K is small and when n is moderate to large, it may be preferable to use inflated OLS residuals for bootstrapping.

Chapter 3 examines closely the relative effectiveness of BLUS and inflated OLS residuals, when used for bootstrapping. This is done by comparing the sample moments of these residuals with the exact moments of the true errors. Analytical results are obtained for the third and fourth moments. A Monte Carlo simulation study is also conducted to compare the performances of the first ten sample moments of a few selected regression residuals.

The main lemma in Chapter 3 is Lemma 3.10. It demonstrates that sample moments of bootstrap estimates of regression coefficients can be expressed as linear functions of sample moments of the underlying residuals. This forms the basis for all the main theorems in Chapters 3 and 4. Lemma 3.10 is important because of the following. By applying least-squares theory, it can easily be shown that moments of $\hat{\beta}$ are linear functionals of the underlying error moments. Thus, one only needs to examine the sample moments of the regression residuals when the objective is to examine the sample moments of bootstrap estimates of β. Using this procedure, one is able to reduce the computational costs for a Monte Carlo study considerably.

1.5 Bootstrap Confidence Intervals of Regression Coefficients

Bootstrap distributions have been studied by, among others, Singh (1981), Babu and Singh (1983), Abramovitch and Singh (1985), Wu (1986), and Hall (1987). However, most of the results were obtained for the case when the statistic of interest is the arithmetic mean of n observations on a random variable. Most of these studies rely on the Edgeworth expansions. The focus of Chapter 3 is on the bootstrap distributions of $\hat{\beta}$. In particular, the focus is on the sample moments of bootstrap estimates of β and on the moments of $\hat{\beta}$. It must be mentioned that this type of analysis is based upon the assumption that the probability density function of $\hat{\beta}$ is uniquely determined by its moments and that these moments exist. It is also based upon a similar assumption that the probability density function of the bootstrap estimates of β is uniquely determined by its moments and that these moments exist.

When the true errors are observable and when a sample of size n of these errors can be obtained for bootstrapping, the naive bootstrap confidence interval (BCI) of β is similar to those BCI's considered by Efron (1985, 1987). Improvements can be made to this BCI by adopting the method of either Diciccio and Tibshirani (1987) or Loh (1987). The first method consists of the composition of a variance-stabilizing transformation and a skewness-reducing transformation. On the other hand, the second method involves computer simulation and density estimation.

The approach adopted in Chapter 4 is based upon an idea similar to that of Babu and Singh (1983). One problem pertaining to distributions of bootstrap estimates of regression coefficients is the lack of a coherent finite-sample theory. It is shown that it would be inappropriate to apply existing theorems to bootstrap estimates of regression coefficients. This is true especially when the true errors are nonobservable and when regression residuals are used for bootstrapping. The discrepancy can be large, and this happens when K is large relative to n.

Three types of BCI's are studied in Chapter 4. These are BCI0, BCI1 and BCI2. BCI0 is the 'naive' BCI obtained by ordering the bootstrap estimates of β. Let σ^2_β be the variance of $\hat{\beta}$, and let s^2_β be an unbiased estimate of σ^2_β. Also, let $t_1=\sigma^{-1}_\beta(\hat{\beta}-\beta)$ and $t_2=s^{-1}_\beta(\hat{\beta}-\beta)$. Then, BCI1 and BCI2 are based upon t_1 and t_2, respectively. When the sample size is small to moderate, BCI2 would be the appropriate interval. Nevertheless, BCI2 is still shorter than the exact confidence interval and the difference is at most $O(n^{-1})$. The terms $O(n^{-1})$ and $O_p(n^{-1})$ are used interchangeably throughout this dissertation.

In the case when a sample of size n of the true errors is used for obtaining the bootstrap estimates of the regression coefficients, it is shown that both bootstrap distributions of t_1 and t_2 admit $O(n^{-1})$ errors. This is consistent with the results of Efron (1979, 1985), Singh (1981) and Abramovitch and Singh (1985). The bootstrap approximations are more accurate when the error distribution is symmetric.

However, for the case when OLS, inflated OLS or BLUS residuals are used for bootstrapping, the results in Chapter 4 are no longer consistent with existing results in the literature. This suggests that one should treat bootstrap distributions of regression estimates as a separate and distinct problem. It would be erroneous to associate the regression estimates with a general class of statistics. The performances of bootstrap approximations in the regression framework can be very poor, especially when either OLS or inflated OLS residuals are used for bootstrapping and when K is large relative to n. These poor performances can be improved slightly by using BLUS residuals. Nevertheless, the better of these bootstrap approximations of t_1 and t_2 can still admit errors which may exceed $O(n^{-1})$.

1.6 An Application of the Bootstrap to a Multiplicative Model

In Chapter 5, the methodology developed in Chapters 2 through to 4 is applied to a problem associated with multiplicative lognormal models. An important difficulty in this type of regression models is the estimation of the constant term and its standard error. Let β be the coefficient to be estimated and $\hat{\beta}$ be an unbiased estimate of β. A nonlinear function of $\hat{\beta}$, say $g(\hat{\beta})$, is generally a biased estimate of $g(\beta)$.

A bias-correction procedure based upon Finney's (1951) g-function has been suggested by Bradu and Mundlak (1970). The g-function requires extensive tables and may sometimes yield unacceptable negative values for some of its arguments (see Teeken and Koerts, 1972 p.84). A simpler procedure which does not require the use of Finney's g-function is given by Srivastava and Singh (1989).

The estimation of a confidence interval for the constant term is often a difficult task. One alternative is to apply asymptotic theory and use the normal or t-tables, depending on the sample size. The jackknife method has been considered by Chaubey and Singh (1988), and the bootstrap method is proposed in Chapter 5.

It is shown that bootstrap confidence intervals obtained by ordering bootstrap estimates of the constant term are biased and should be avoided. This result applies to both the Bradu-Mundlak and Srivastava-Singn estimates. Alternative bootstrap confidence intervals are proposed, and the one based upon the bootstrap t-distribution is found to have the best empirical coverage. However, this method may sometimes yield an infinite upper bound.

The lognormality assumption is crucial for the results in Chapter 5. The validity of both Bradu-Mundlak and Srivastava-Singh estimates depend on this assumption. In this study, the bootstrap method is only considered for the construction of confidence intervals. Nevertheless, the role of the bootstrap can be enhanced. When the lognormality assumption is violated, both estimates will no longer be reliable and better estimates need to be constructed. Two feasible alternatives are the jackknife and bootstrap estimates considered in Chaubey and Sim (1988).

1.7 Application of the Bootstrap to an AR(1) Process

In Chapter 6, an almost unbiased estimate is obtained for the parameter of an AR(1) process. This estimate is based upon the results of White (1961) and Marriott and Pope (1954). The distribution of this estimate (or a similar estimate) had been studied by several authors including Tanaka (1983), Phillips (1984), Durbin (1986) and Phillips and Reiss (1986). Both Tanaka (1983) and Phillips (1984) showed that Edgeworth approximations to the exact distribution of the above estimate perform poorly, especially in the tails, when the model is close to the border of nonstationarity. The normal approximation is not very satisfactory for sample sizes of less than or equal to 20. Phillips (1984) proposed a method based upon extended rational approximants (ERA's), which yields very close estimates for the parameter of an AR(1) model, when the sample size is greater than or equal to 5. Nevertheless, this method requires that the distribution of the time-series be known.

The bootstrap method is used in Chapter 6 to approximate the exact distribution of the above estimate. This latter method does not require the underlying distribution of the time-series to be known. Some results are obtained for the mean and variance of the bootstrap estimates of the parameter.

Simulation results are also obtained for comparing the bootstrap with conventional confidence intervals when the underlying time-series is a first-order stationary Gaussian process. The results suggest that the conventional confidence intervals are not reliable and should not be used when the true value of the AR(1) parameter exceeds 0.5. On the other hand, bootstrap confidence intervals provide fairly good approximations to the exact confidence interval when the parameter value is positive.

CHAPTER TWO

BOOTSTRAPPING LINEAR REGRESSION MODELS: SOME FINITE SAMPLE RESULTS

2.1 Introduction

Prior to the introduction of the bootstrap by Efron (1979), the jackknife had been applied by Miller (1974) and Hinkley (1976) to linear regression (LR). Efron had shown that the jackknife ignores a crucial assumption of LR analysis; namely, the assumption that the errors are independently and identically distributed (i.i.d.). On the contrary, the bootstrap presumes that errors are i.i.d. When applied to a LR problem, therefore, bootstrapping is much more efficient than jackknifing, provided the errors are indeed i.i.d. However, the bootstrap does have its own disadvantages; for example: (1) it is computationally intensive and hence, for some problems, it may be rather costly; (2) it may not be appropriate for small samples, because it is accurate only up to $O(n^{-1})$ in probability; and (3) the true sampling distribution of the errors cannot be observed.

The fact that the bootstrap is a computationally intensive method is not in itself a problem. It is the availability of a relatively less costly alternative that makes the bootstrap unattractive. One such alternative is an existing analytical (conventional least-squares) method. Yet another alternative is the jackknife method. With the advent of modern electronic computers, the bootstrap has become more practicable for many applications. Its fundamental attraction is the ability to transform an otherwise complicated problem into a comparatively simple one.

In the application of Efron's bootstrap to LR analysis, the main drawback is in problem (3) above. Nevertheless, Freedman (1981) has derived some asymptotic results for bootstrapping a LR model with respect to (w.r.t.) ordinary least-squares (OLS) residuals, these being the practical substitutes for the unobservable errors. Although this method is asymptotically valid, it may not be appropriate when the sample size is small. This problem was addressed by Stine (1985), who suggested a rescaling scheme for the OLS residuals to correct for some of the deficiencies. However, Stine's rescaling scheme, in general, leads to further complications. This becomes evident in Theorem 1.11 below.

Other related papers on bootstrap theory are Singh (1981), Bickel and Freedman (1981) [hereinafter B&F (1981)], Freedman and Peters (1984a, b), Berar (1984), Efron (1985) and Wu (1986).

In Section 2.2, the LR model is introduced to establish notation and the underlying assumptions. A Monte Carlo interpretation of the classical results will be given in Section 2.3. The purpose of this section is to serve as a link between the classical and the bootstrap results. It may also assist in understanding the fundamental objective of applying the bootstrap to the LR model.

Much of the bootstrap literature has been based on the assumption that a random sample of size n can be obtained from the population under study. In Section 2.4, the procedure for bootstrapping the LR model when the true errors are observable is illustrated. This is compared with the Monte Carlo simulation described in Section 2.3. Section 2.4 also serves as a bench-mark for comparison with the case when the true errors are not observable and must be replaced by estimates. This case will be discussed in both Sections 2.5 and 2.6 below. Two estimators of the true errors that are considered in this chapter are OLS and Theil's (1965) BLUS residuals. The problems associated with the use of OLS residuals for bootstrapping the LR model are highlighted in Section 2.5. In Section 2.6, the derivation and use of BLUS residuals for bootstrapping the same model are demonstrated. Section 2.4, 2.5 and 2.6 contain the main theoretical contributions of this chapter.

In the literature, Freedman and Peters (1984a) have suggested the use of either BLUS or inflated OLS residuals for bootstrapping the LR model. Inflated OLS residuals are OLS residuals multiplied by a factor depending on the sample size and number of coefficients to be estimated, such that the second sample moment of the inflated residuals is an unbiased estimate of the second moment of the true errors. However, no coherent finite sample theory is available to support (or reject) the use of either or both types of residuals. In this chapter, a finite sample theory is developed, w.r.t. the use of bootstrapping in LR models. This is done in Sections 2.5 and 2.6 below. Some finite sample results were obtained for the first two moments of the bootstrap estimates of the regression coefficients, when OLS, inflated OLS and BLUS residuals were used for bootstrapping. The main results in Sections 2.5 and 2.6 reject the use of OLS residuals but support the use of either BLUS or inflated OLS residuals, when the objective of applying bootstrap to the LR model is to obtain unbiased estimates of the regression coefficients and their dispersions. Since BLUS residuals are computationally more burdensome to obtain, preference is given to use of inflated OLS residuals. Moreover, when n is large, the cost of computing BLUS residuals will be prohibitive.

When the sample mean of the residuals is non-zero, it will be shown below that bootstrapping based upon these residuals leads to biased estimates of the regression coefficients. In some situations, it may also lead to biased estimates of the dispersions of these coefficients. Thus, the task of centering the residuals (to sample mean zero), before bootstrapping, is crucial for a successful application of Efron's bootstrap. However, one degree of freedom is lost due to centering and this must be corrected. The error is $O(n^{-1})$ and this can be corrected by scaling the centered residuals by the factor $\left[n(n-1)^{-1}\right]^{1/2}$, although this procedure is effective only for the first two moments. It is partially effective for the higher moments (see Chapter 3 below). The results obtained in this chapter are consistent with the asymptotic results of Freedman (1981).

Finally, it is shown below that both the bootstrap and conventional least-squares approaches yield the same estimate of β. Also, the bootstrap estimate of the dispersion of the OLS estimate of β is exactly the same as its OLS estimate, provided that an infinite number of bootstrap replications and the correct residuals are used. When mathematical expressions for the statistics of interest concerning the regression coefficients are difficult to obtain, Efron's bootstrap may prove to be a useful (and powerful) technique for econometric analyses. A few such examples can be found in nonlinear regression, seemingly unrelated regression and two-stage least-squares regresssion models. However, for the LR model below, one does not need to apply the bootstrap method to obtain the bias and dispersion of the OLS estimate of β. This is true regardless of the type of error distribution, provided that its second moment exists.

2.2 The Model

Let the LR model to be considered be

$$y = X\beta + \varepsilon \tag{2.2.1}$$

in which y is a nx1 vector of observations on the dependent variable, X is nxK matrix of observations on K exogeneous variables, β is Kx1 vector of unknown coefficients to be estimated from the data, and ε is a nx1 vector of unobservable random disturbances.

The assumptions underlying the model are those of Freedman (1981) which are:

A.2.1: The exogeneous observations, x_{ij} $(i=1,\dots,n; j=1,\dots,K)$ are assumed to be nonstochastic such that the matrix $X = [x_{ij}]$ has full column rank K for all n and

$$n^{-1}[X^T X] \xrightarrow[n]{a.e.} M_{xx} \, ,$$

the KxK matrix M_{xx} being positive-definite (p.d.) of finite elements. This assumption excludes variables which grow indefinitely, as n goes to infinity (see e.g. White, 1984 p.42).

A.2.2: The components of ε are i.i.d. with common unknown distribution function, *F*, having mean zero and variance σ^2. The error variance is assumed to be finite.

As a consequence of A.2.1. and A.2.2,

$$n^{-1}[X^T\varepsilon] \xrightarrow[n]{a.s.} 0.$$

In regression models, the regressors are not necessarily fixed; they are assumed fixed here, for simplicity. Thus, the finite sample bootstrap results below are only applicable to LR models with fixed regressors. No attempt will be made here to extend the results to models with stochastic regressors.

2.3. Classical Results

Attention is restricted to the OLS estimate (OLSE) of the LR model. An estimate $\hat{\beta}$ of β is selected according to

$$\hat{\beta}: \quad \underset{\beta}{\text{Min}} \left\{ \varepsilon^T \varepsilon \right\} = \underset{\beta}{\text{Min}} \left\{ (y - X\beta)^T (y - X\beta) \right\}$$

yielding

$$\hat{\beta} = (X^TX)^{-1}X^Ty. \tag{2.3.1}$$

Substituting (2.2.1) into (2.3.1),

$$\hat{\beta} = \beta + (X^TX)^{-1}X^T\varepsilon. \tag{2.3.2}$$

The OLS predictor of ε is $\hat{\varepsilon}$, given by

$$\hat{\varepsilon} = y - X\hat{\beta} = M\varepsilon \tag{2.3.3}$$

in which $M = \left[I_n - X(X^TX)^{-1}X^T \right]$

Taking expectations of (2.3.2),

$$E(\hat{\beta}) = \beta \tag{2.3.4}$$

by A.2.2. The dispersion of $\hat{\beta}$ is $D(\hat{\beta})$ and

$$D(\hat{\beta}) = \sigma^2(X^TX)^{-1} \tag{2.3.5}$$

by A.2.1 and A.2.2. Thus, for given n and X, $\hat{\beta}$ is distributed with mean β and variance $\sigma^2(X^TX)^{-1}$. Subject to the normal regularity conditions (see e.g. White, 1984 p.2), $\hat{\beta}$ is a consistent estimate of β and $\left\{ n^{1/2}\left[\hat{\beta}-\beta\right] \right\}$ is asymptotically normal having mean zero and variance $\left\{ \sigma^2 M_{XX}^{-1} \right\}$ (see e.g. White, 1984 p.14). Finally, an unbiased estimator of σ^2 is given by

$$\sigma_n^2 = \left\{(n-K)^{-1}\left[\hat{\varepsilon}^T\hat{\varepsilon}\right]\right\} \tag{2.3.6}$$

and $\left[n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}\right]$ is a maximum likelihood estimator of σ^2, provided that the errors are normally distributed.

Results (2.3.4) through (2.3.6) can be confirmed by Monte Carlo simulation for fixed n and a sufficiently large number of trials. Monte Carlo simulation is a method that "consists in drawing a large number of samples from the population in question, computing the value of the statistic of interest for each sample, and recording the empirically observed sampling distribution of that statistic" [Christ (1958, p.513)].

Let j be the index corresponding to the j'th trial. For n finite and all $j=1,2,\ldots,J$, $\varepsilon_{(j)}^T=\left[\varepsilon_{(j)1},\ldots,\varepsilon_{(j)n}\right]$ is drawn from the known distribution F. The vector $y_{(j)}$ is then constructed for given β and X as

$$y_{(j)} = X\beta + \varepsilon_{(j)}.$$

From (2.3.1), (2.3.3) and (2.3.6), we have

$$\text{(i)} \quad \hat{\beta}_{(j)} = (X^TX)^{-1}X^Ty_{(j)} ,$$

$$\text{(ii)} \quad \hat{\varepsilon}_{(j)} = y_{(j)} - X\hat{\beta}_{(j)} ,$$

$$\text{(iii)} \quad s_{(j)}^2 = \left\{(n-K)^{-1}\left[\hat{\varepsilon}_{(j)}^T\hat{\varepsilon}_{(j)}\right]\right\}.$$

The classical results then imply that for n finite,

$$\text{(i)} \quad \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\hat{\beta}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \beta,$$

(ii) $\left\{\frac{1}{J}\sum_{j=1}^{J}\left[s^2_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \sigma^2,$

(iii) $\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\beta}_{(j)}-\beta)(\hat{\beta}_{(j)}-\beta)^T\right]\right\} \xrightarrow[J]{a.s.} \sigma^2(X^TX)^{-1}.$

Thus, for J very large, Monte Carlo results must coincide with the classical results for finite n.

2.4 Bootstrapping the LR Model: The Case when ε is Known

Let $(\varepsilon_1,\ldots,\varepsilon_n)$ be a known random sample of size n from a population with distribution F and let F_n denotes the distribution that puts mass (n^{-1}) at each points $\varepsilon_1,\ldots,\varepsilon_n$. If $B(\varepsilon_1,\ldots,\varepsilon_n:F)$ is the specified random variable of interest, then Efron's bootstrap can be used to approximate the distribution of $B(\varepsilon_1,\ldots,\varepsilon_n:F)$ under F by that of $B(\varepsilon^*_1,\ldots,\varepsilon^*_n:F_n)$ under F_n, in which $(\varepsilon^*_1,\ldots,\varepsilon^*_n)$ denotes a bootstrap sample of size n from F_n. The bootstrap procedure for the LR model was first described by Efron (1979, Section 1.7) and, recently, by Wu (1986). Singh (1981) gives asymptotic results when the bootstrap is applied to the LR model.

The bootstrap method is similar to the Monte Carlo method, except that in the bootstrap experiment, $\varepsilon^*_{(j)}$ is drawn from F_n instead of F. The reason for using F_n is that it is attainable, whereas F is neither known nor observable. For all $j=1,2,\ldots,J$, $\varepsilon^*_{(j)}=\left[\varepsilon^*_{(j)},\ldots,\varepsilon^*_{(j)n}\right]$ is drawn from F_n, and $y^*_{(j)}$ is constructed as

$$y^*_{(j)} = X\beta + \varepsilon^*_{(j)}. \tag{2.4.1}$$

By analogy with (2.3.1), (2.3.3) and (2.3.6):

$$\beta^*_{(j)} = (X^TX)^{-1}X^Ty^*_{(j)}, \tag{2.4.2}$$

$$\hat{\varepsilon}^*_{(j)} = y^*_{(j)} - X\beta^*_{(j)}, \tag{2.4.3}$$

$$s^{*2}_{(j)} = \left\{(n-K)^{-1}\left[\hat{\varepsilon}^{*T}_{(j)}\hat{\varepsilon}^*_{(j)}\right]\right\}. \tag{2.4.4}$$

Corresponding to (2.3.2),

$$\beta^*_{(j)} = \beta + (X^TX)^{-1}X^T\varepsilon^*_{(j)}. \tag{2.4.5}$$

In order to investigate the properties of $\beta^*_{(j)}$, some new notation is required.

Definition 2.1: Let $S_{(j)}$, a mxn matrix, be the selection matrix corresponding to the j'th bootstrap replication, where $j=1,\ldots,J$. Each row of $S_{(j)}$, denoted by $S_{(j)r.}$ $(r=1,\ldots,m)$, has zero everywhere except in one position which is unity. Let this one position be denoted by $S_{(j)ri}$ in which i is an integer randomly selected with replacement from the set $S=(1,2,3,\ldots,n)$. For each r, $S_{(j)rc}(c=1,2,\ldots,n)$ is a random variable which takes on the values 0 or 1 and each point has probability (n^{-1}) that its value will be 1. In essence, $S_{(j)}$ is a binary matrix which randomly selects m elements with replacement from a set of n and each of the n elements has the same probability (n^{-1}) that it will be selected on each draw.

Definition 2.2: Let the transposes of $S_{(j)r.}$ and $S_{(j)}$ be $(S_{(j)r.})^T$ and $S^T_{(j)}$, respectively. Also, let $S^T_{(j).r}$ denote the r'th column of $S^T_{(j)}$.

It is straightforward to show that $(S_{(j)r.})^T=S^T_{(j).r}$. Hereinafter, the transpose of $S_{(j)r.}$ will be written as $S^T_{(j).r}$.

Lemma 2.1: Let J be the number of bootstrap replications and n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{n^{-1}\left[E(m,n)\right]\right\}$$

and E(m,n) represents an mxn matrix with unity everywhere.

Proof:

Note that for each $r(r=1,\ldots,m)$, $S_{(j)r.}$ randomly selects an element from a set of n elements and each of the n elements has probability (n^{-1}) that it will be selected. Consequently, each of the elements $S_{(j)rc}$ $(c=1,\ldots,n)$ equals zero and unity with probabilities $(1-n^{-1})$ and (n^{-1}), respectively. When this selection process is repeated J times, the observed frequency of the c'th element, from the set of n, in the r'th element of a bootstrap sample of size m is $\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)rc}\right]\right\}$. As J goes to infinity, the latter quantity converges a.s. to its theoretical value of (n^{-1}), for all r,c. *Q.E.D.*

Lemma 2.2: Let both m and n be finite. Then

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}S_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[(n^{-1}m).I_{n}\right].$$

Proof:

Let $S^{T}_{(j)} = \left[S^{T}_{(j).1},\ldots,S^{T}_{(j).m}\right].$

Then

$$S^{T}_{(j)}S_{(j)} = \left\{\sum_{r=1}^{m}\left[S^{T}_{(j).r}S_{(j)r.}\right]\right\}.$$

Note that $\left[S^{T}_{(j).r}S_{(j)r.}\right]$, $r=1,\ldots,m$, are nxn matrices with zero everywhere except for one of its main diagonal elements which is unity. Note that for each r, the 1xn vector of the main diagonal elements of $\left[S^{T}_{(j).r}S_{(j)r.}\right]$ is exactly the vector $S_{(j)r.}$. The theoretical probability that a particular diagonal element is unity is (n^{-1}). It is then straightforward from Lemma 2.1 that

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j).i}S_{(j)i.}\right]\right\} \xrightarrow[J]{a.s.} \left[(n^{-1}).I_{n}\right]$$

for all $i=1,\ldots,m$ and

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{m}\left[S^{T}_{(j).i}S_{(j)i.}\right]\right\} \xrightarrow[J]{a.s.} \left[(n^{-1}m).I_{n}\right]. \quad \textit{Q.E.D.}$$

An alternative proof for Lemma 2.2 is also given in Appendix A. The following corollary states the result of Lemma 2.2 when n elements are selected with replacement from a set of n. This is common when the bootstrap method is applied to regression models.

Corollary 2.1: When $m=n$ and n is finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}S_{(j)}\right]\right\} \xrightarrow[J]{a.s.} I_{n}.$$

This is a direct result from Lemma 2.2.

Lemma 2.3: Let both m and n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}S^{T}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[n^{-1}(n-1)\right].I_{m} + (n^{-1}).E(m,m)\right\}.$$

Proof:

First note that the mxm matrices $\left[S_{(j)}S^{T}_{(j)}\right]$, $(j=1,\ldots,J)$, can also be written as

$$(S_{(j)}S^{T}_{(j)}) = \left[S_{(j)hi}\right], \qquad (h,i=1,\ldots,m).$$

$$= \left[S_{(j)h.}\,S^{T}_{(j).i}\right]$$

Since $S_{(j)h.}$ $(h=1,\ldots,m)$ is a binary vector, comprising zero elements except for one which is unity, the inner product of one such vector and its transpose is unity. That is, $S_{(j)hi}=1$ when h=i. When two such vectors $S_{(j)h.}$ and $S_{(j)i.}$ are different, the inner product of one vector and the transpose of the other will be zero. Also, when h is different from i, the probability that $S_{(j)h.}$ and $S_{(j)i.}$ are identical is (n^{-1}) and the probability that they are different is $(1-n^{-1})$. Thus, $S_{(j)hi}(h,i=1,\ldots,m)$ equal unity and zero with probabilities (n^{-1}) and $(1-n^{-1})$, respectively. Consequently, when both m and n are finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)hi}\right]\right\} = 1 \quad \text{when } h=i$$

and

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)hi}\right]\right\} \xrightarrow[J]{a.s.} (n^{-1}) \quad \text{when } h\neq i.$$

The remainder of the proof is then straightforward. *Q.E.D.*

An alternative proof for Lemma 2.3 can also be found in Appendix A. Corollary 2.2 below gives the result of Lemma 2.3 when a finite number (say, m) of elements is being selected from an infinite set. When m=n and both are infinite, the result of Lemma 2.3 is given in Corollary 2.3.

Corollary 2.2: Let m be finite. Then, as n goes to infinity,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}S_{(j)}^{T}\right]\right\} \xrightarrow[J]{a.s.} I_m.$$

This corollary is straightforward from Lemma 2.3.

Corollary 2.3: When m=n, as n goes to infinity,

$$\left[\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}S_{(j)}^{T}\right]\right\} - \left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}^{T}S_{(j)}\right]\right\}\right] \xrightarrow[J]{a.s.} 0.$$

This result is apparent from Lemma 2.2 and 2.3.

Lemma 2.4: Let there exist a matrix A, such that $A=[a_{hi}]$ and let the arguments a_{hi} $(h,i=1,\ldots,n)$ be real. Also, let n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}AS^{T}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left\{\sum_{i=1}^{n}\left[n^{-1}(a_{ii})\right]\right\} I_m + \left\{\sum_{h=1}^{n}\sum_{i=1}^{n}\left[n^{-2}(a_{hi})\right]\right\}\left[E(m,m)-I_m\right]\right\}.$$

Proof:

Note that the mxm matrix

$$\left[S_{(j)}AS^{T}_{(j)}\right] = \left[S_{(j)v.}AS^{T}_{(j).w}\right], \qquad (v,w=1,\ldots,m).$$

Let the 1xm vectors Y_v $(v=1,\ldots,m)$, comprising elements $y_{v1}, y_{v2}, \ldots, y_{vn}$, be defined as

$$Y_v = \left[S_{(j)v.}A\right], \qquad v=1,\ldots,m.$$

For any $(v=1,\ldots,m)$, the relative frequency of a_{hi} $(h=1,\ldots,n)$ in y_{vk} of each bootstrap sample is (n^{-1}) when i equals $k(k=1,\ldots,n)$ and zero when i is different from k. When v equals w and $y_{vi}=a_{hi}$ $(i=1,\ldots,n)$, the relative frequency of y_{vi} in $\left[Y_vS^{T}_{(j).w}\right]$ is 1 when i=h and zero when i is different from h. When v is different from w, the relative frequency of y_{vi} in $[Y_vS^{T}_{(j).w}]$ is (n^{-1}) for all $i=1,\ldots,n$. Consequently, when v equals w, the probability that $(S_{(j)v.}AS^{T}_{(j).w})$ equals a particular element of A is (n^{-2}). Thus, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)v}AS^{T}_{(j)w}\right]\right\} \xrightarrow[J]{a.s.} \left\{\sum_{i=1}^{n}\left[n^{-1}(a_{ii})\right]\right\} \qquad \text{when } v=w$$

or

$$\left\{\sum_{h=1}^{n}\sum_{i=1}^{n}\left[n^{-2}(a_{hi})\right]\right\} \qquad \text{when } v\neq w. \quad \textit{Q.E.D.}$$

With the above lemmas, it is now possible to state the following theorems which will help illustrate some of the finite sample properties of bootstrap estimates of the regression coefficients.

Theorem 2.1: Let $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)$ and the sample mean of ε be $\bar{\varepsilon}=\left\{\sum_{i=1}^{n}\left[n^{-1}\varepsilon_i\right]\right\}$. Also, let both m and n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\right]\right\} \xrightarrow[J]{a.s.} \left[\bar{\varepsilon}.E(m,1)\right].$$

Proof:

Upon applying Lemma 2.1,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\right]\right\} \xrightarrow[J]{a.s.} \left\{(n^{-1})\left[E(m,n)\varepsilon\right]\right\}.$$

However,

$$\left[E(m,n)\varepsilon\right] = \left\{(n\bar{\varepsilon})\left[E(m,1)\right]\right\}.$$

Consequently,

$$\left\{(n^{-1})\left[E(m,n)\varepsilon\right]\right\} = \left\{(n^{-1})(n\bar{\varepsilon})\left[E(m,1)\right]\right\} = \left\{\bar{\varepsilon}\left[E(m,1)\right]\right\}. \quad Q.E.D.$$

Theorem 2.2: Let both m and n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[X^TS_{(j)}\varepsilon\right]\right\} \xrightarrow[J]{a.s.} \bar{\varepsilon}\left[X^TE(m,1)\right].$$

Proof:

The proof is straightforward from Theorem 2.1.

Theorem 2.3: Let the notation be that of Theorem 2.1 and $s_n^2=\left[\varepsilon^T\varepsilon/n\right]$. Then, for both m and n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(S_{(j)}\varepsilon)^T(S_{(j)}\varepsilon)\right]\right\} \xrightarrow[J]{a.s.} \left[ms_n^2\right].$$

Proof:

Note that

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(S_{(j)}\varepsilon)^T(S_{(j)}\varepsilon)\right]\right\} = \varepsilon^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}^T S_{(j)}\right]\right\}\varepsilon.$$

The remainder of the proof becomes straightforward upon application of Lemma 2.2. Q.E.D.

Let the bootstrap errors be defined as $\varepsilon^*_{(j)} = S_{(j)}\varepsilon$. Then, Theorem 2.2 states that the expectation of $X^T\varepsilon^*_{(j)}$ is zero, provided that the sample mean of ε is zero. Theorem 2.3 states that the second moment of $\varepsilon^*_{(j)}$ converges a.s. to the second sample moment of ε, as J goes to infinity. The following corollary states that the same moment of $\varepsilon^*_{(j)}$ converges a.s. to the exact second moment of ε, as both n and J go to infinity.

Corollary 2.4: Let m be finite. Then, as n goes to infinity,

(i) $s^2 \xrightarrow[n]{a.s.} \sigma^2$,

(ii) $\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(S_{(j)}\varepsilon)^T(S_{(j)}\varepsilon)\right]\right\} \xrightarrow[n]{a.s.} m\sigma^2.$

Statement (i) is true by the Strong Law of Large Numbers (SLLN) and (ii) follows from (i) and Theorem 2.3. [C.f. B&F's (1981 p. 1197) Theorem 2.1, and Singh (1981)].

Theorem 2.3 states that as J goes to infinity, the variance of the bootstrap errors approaches the sample variance of the observed true errors. As for the empirical distribution of the observed errors, B&F's (1981, p. 1212) Lemma 8.4 shows that this distribution converges almost everywhere to the actual error distribution. This is possible because of the SLLN. Thus, statement (i) of the above corollary is implied by this lemma. The main statement of Theorem 2.3 and corollary 2.4 is that the variance of bootstrap errors converges almost surely to the sample variance of the observed errors, as J goes to infinity; and, when n also goes to infinity, it also converges almost everywhere to the actual variance of the true errors.

Theorem 2.4: Let the matrix A in Lemma 1.4, be replaced by the error sample variance-covariance matrix, $[\varepsilon\varepsilon^T]$. Also, let both m and n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\varepsilon^T S^T_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[(s_n^2)I_m\right] + \bar{\varepsilon}^2\left[E(m,m) - I_m\right]\right\}.$$

Proof:

Upon application of Lemma 2.4,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\varepsilon^T S^T_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[\left\{\sum_{i=1}^{n}\left[n^{-1}(\varepsilon_i^2)\right]\right\}I_m + \left\{\sum_{h,i}^{n}\left[n^{-2}(\varepsilon_h\varepsilon_i)\right]\right\}\left[E(m,m) - I_m\right]\right].$$

However,

$$\text{(i)} \quad \left\{\sum_{i=1}^{n}\left[n^{-1}(\varepsilon_i^2)\right]\right\} = s_n^2 \quad ,$$

$$\text{(ii)} \quad \left\{\sum_{h,i}^{n}\left[n^{-2}(\varepsilon_h\varepsilon_i)\right]\right\} = \bar{\varepsilon}^2.$$

Statement (ii) is true because

$$\left\{\sum_{h,i}^{n}\left[n^{-2}(\varepsilon_h\varepsilon_i)\right]\right\} = \left\{\sum_{h=1}^{n}\left[n^{-1}\varepsilon_h\right]\right\}\left\{\sum_{i=1}^{n}\left[n^{-1}\varepsilon_i\right]\right\} = \bar{\varepsilon}^2. \quad Q.E.D.$$

Corollary 2.5: Let m be finite and let $\bar{\varepsilon}\neq 0$. Then, as n goes to infinity,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\varepsilon^{T}S_{(j)}^{T}\right]\right\}\xrightarrow[J]{a.s.}\sigma^{2}I_{m}.$$

This corollary follows from Theorem 2.4 and applications of the SLLN and B&F's Lemma 8.4.

Theorem 2.4 states that the variance-covariance matrix of the bootstrap errors converges a.s. to $\left[(s_n^2)I_m\right]$, as J goes to infinity, provided that the sample mean of ε is zero. Otherwise, it will only do so when n goes to infinity and this is stated in Corollary 2.5. The following theorem concerns the mean of the bootstrap estimates of a regression coefficient. It must be noted that, hereinafter, only bootstrap samples of size n will be considered.

Theorem 2.5: Given a finite sample of size n,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}\right]\right\}\xrightarrow[J]{a.s.}\beta+(X^{T}X)^{-1}X^{T}\left[\bar{\varepsilon}.E(n,1)\right].$$

Proof:

First, let m=n in Theorem 2.1. The proof then becomes apparent upon application of Theorem 2.1. *Q.E.D.*

Let e_i be the equiangular line in $\mathbb{R}^i$, i=1,2,...,n. Then, $e_m e_n^T=E(m,n)$ and $e_n=E(n,1)$. The sample mean of ε can also be written in terms of e_n as $\bar{\varepsilon}=n^{-1}e_n^T\varepsilon=(e_n^Te_n)^{-1}e_n^T\varepsilon$, because $(e_n^Te_n)^{-1}=n^{-1}$. Since $\bar{\varepsilon}$ is a scalar, $(\bar{\varepsilon}.E(n,1)$ in Theorem 2.5 can also be written as $(E(n,1).\bar{\varepsilon})=e_n(e_n^Te_n)^{-1}e_n^T\varepsilon=\mathbb{P}_n\varepsilon$, where $\mathbb{P}_n=e_n(e_n^Te_n)^{-1}e_n^T$ is an orthogonal projection onto the equiangular line. Thus, the R.H.S. of Theorem 2.5 can also be written as $\left[\beta+(x^Tx)^{-1}x^T\mathbb{P}_n\varepsilon\right]$.

Theorem 2.6: For n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\beta)(\beta^*_{(j)}-\beta)^T\right]\right\} \xrightarrow[J]{a.s.} D(\hat{\beta}) + (X^TX)^{-1}X^T\left[\bar{\varepsilon}^2.E(n,n)\right]X(X^TX)^{-1}$$

in which $D(\hat{\beta}) = \left[\hat{\sigma}^2(X^TX)^{-1}\right]$, and $\hat{\sigma}^2 = \left[n^{-1}\varepsilon^T\varepsilon - \bar{\varepsilon}^2\right]$.

Proof:

From equation (2.4.5), $\left[\beta^*_{(j)}-\beta\right] = (X^TX)^{-1}X^TS_{(j)}\varepsilon$.

Consequently,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\beta)(\beta^*_{(j)}-\beta)^T\right]\right\} = \left[(X^TX)^{-1}X^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\varepsilon^TS^T_{(j)}\right]\right\}X(X^TX)^{-1}\right].$$

Applying Theorem 2.4 and setting m=n, for n finite

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\varepsilon\varepsilon^TS^T_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[(s^2_n).I_n\right] + \bar{\varepsilon}\left[E(n,n) - I_n\right]\right\}.$$

Let $\hat{\sigma}^2 = (s^2-\bar{\varepsilon}^2)$, then for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\beta)(\beta^*_{(j)}-\beta)^T\right]\right\} \xrightarrow[J]{a.s.}$$

$$\left[(X^TX)^{-1}X^T\left\{\left[\hat{\sigma}^2I_n\right] + \bar{\varepsilon}^2\left[E(n,n)\right]\right\}X(X^TX)^{-1}\right]. \quad Q.E.D.$$

Corollary 2.6: Let the sample mean of $(\varepsilon_1,\ldots,\varepsilon_n)$ be zero,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\beta)(\beta^*_{(j)}-\beta)^T\right]\right\} \xrightarrow[J]{a.s.} \hat{\sigma}^2(X^TX)^{-1}.$$

This is a direct result of Theorem 2.6.

Note that the R.H.S. of Corollary 2.6 is exactly the OLS estimate of $D(\hat{\beta})$. Thus, Corollary 2.6 states that the bootstrap estimate of $D(\hat{\beta})$ converges a.s. to $D(\hat{\beta})$, as J goes to infinity, provided that the sample mean of ε is zero. When $\hat{\sigma}=\sigma$ and J is very large, another result of

Theorem 2.6 can also be derived. This result states that the bootstrap estimate of $D(\hat{\beta})$ converges a.s. to the Monte Carlo estimate of the same quantity.

When n is finite, the sample mean of ε is often non-zero. This problem can be overcome by subtracting from each element of ε its sample mean. In the sense of Efron (1979), this procedure is known as centering. The transformed sample of ε will be referred to as the centered sample of ε.

Nevertheless, by centering ε at its sample mean, another problem is being created. The reason is that one degree of freedom is lost due to centering. If $\hat{\varepsilon}$ is the centered sample of ε, then $\hat{\sigma}^2$ will underestimate σ^2 because $E\left[n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}\right]=\left[n^{-1}(n-1)\right]\sigma^2$. Thus, $\hat{\varepsilon}$ will not be suitable for bootstrapping when an unbiased bootstrap estimate of $D(\hat{\beta})$ is required. However, by an appropriate transformation, $\hat{\varepsilon}$ can be made suitable for obtaining an unbiased bootstrap estimate of $D(\hat{\beta})$. A Monte Carlo simulation study is also required to determine the effect of using $\hat{\varepsilon}$ on the higher moments of the bootstrap estimate of β This will be done in Chapter 3.

Theorem 2.7: Let $\bar{\beta}^{*}=\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}\right]\right\}$, then for finite n

$$E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}-\bar{\beta}^{*})(\beta^{*}_{(j)}-\bar{\beta}^{*})^{T}\right]\right\} \xrightarrow[J]{a.s.} \left[z.D(\hat{\beta})\right] ,$$

in which $z=\left[n^{-1}(n-1)\right]$.

Proof:

In Theorem 2.5, it has been established that for n finite

$$\left[\bar{\beta}^{*}-\beta\right] \xrightarrow[J]{a.s.} (X^TX)^{-1}X^T\left[\bar{\varepsilon}.E(n,1)\right] = (X^TX)^{-1}X^T\mathbb{P}_n\varepsilon.$$

Further,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}-\bar{\beta}^{*})(\beta^{*}_{(j)}-\bar{\beta}^{*})^T\right]\right\} =$$

$$\left[(X^TX)^{-1}X^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}(\varepsilon-\bar{\varepsilon})(\varepsilon-\bar{\varepsilon})^TS^T_{(j)}\right]\right\}X(X^TX)^{-1}\right]$$

and

$$E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}(\varepsilon-\bar{\varepsilon})(\varepsilon-\bar{\varepsilon})^TS^T_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[(z.\hat{\sigma}^2)I_n\right].$$

Hence,

$$E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^{*}_{(j)}-\bar{\beta}^{*})(\beta^{*}_{(j)}-\bar{\beta}^{*})^T\right]\right\} \xrightarrow[J]{a.s.} (X^TX)^{-1}X^T\left[(z.\hat{\sigma}^2)I_n\right]X(X^TX)^{-1}$$

and the R.H.S. of this latter expression reduces to $\left[z.D(\hat{\beta})\right]$. *Q.E.D.*

Theorem 2.5 states that when the sample mean of ε is not zero, the bootstrap estimate of β will contain finite sample bias. Theorem 2.6 establishes an estimator of $D(\hat{\beta})$ which will have zero finite sample bias, provided that the sample mean of ε is zero. When the sample mean of ε is not zero, Theorem 2.7 shows that an unbiased bootstrap estimate of $D(\hat{\beta})$ can still be obtained. This can be done by scaling the bootstrap estimate of $D(\hat{\beta})$ in Theorem 2.7 by a factor $\left[n(n-1)^{-1}\right]$.

The subsequent two sections deal with the case when the true errors are unknown but are replaced by estimates. Section 2.5 investigates the case when OLS residuals are used as estimates of the true errors for bootstrapping. The case when BLUS residuals are used for bootstrapping will be discussed in Section 2.6. Both OLS and BLUS residuals are derived from the fitted model. Thus, both Sections 2.5 and 2.6 assumed that the current model is being fitted. No attempt will be made here to incorporate model misspecification into the results below.

2.5 Bootstrapping the LR Model: The Case of OLS Residuals

In applications of Efron's bootstrap to the LR model using real world data, both β and ε are not observable. However, the fact that β is unknown is not a problem because it could be replaced by its OLS estimate, $\hat{\beta}$ as defined in (2.3.1), without distorting the main results in the preceding section. If β is replaced by $\hat{\beta}$ in equations (2.4.1) and (2.4.5), Theorems 2.5, 2.6, and 2.7 and Corollary 2.6 still hold. On the other hand, when ε is replaced by $\hat{\varepsilon}$, bootstrapping the OLS residuals leads to some difficulties. Freedman (1981) has shown that the bootstrap procedure using OLS residuals is valid asymptotically for the LR model when m=n or when the OLS residuals are centered. However, small-sample properties of the bootstrap estimates of the regression coefficients are still unknown. In this section, it is shown that for the LR model, the bootstrap estimates of a regression coefficient and its dispersion are identical to those obtained by the conventional least-squares method, provided that the bootstrap method is correctly applied.

In the literature, difficulties associated with the use of OLS residuals for bootstrapping the LR model have been partly attributed to the fact that, although $\hat{\varepsilon}$ is asymptotically an appropriate estimator of ε, it has a covariance matrix which depends on the design matrix X. Stine (1985) has proposed a rescaling scheme which allegedly produces better results. In this section, it is demonstrated that Stine's rescaling scheme can lead to further complications. It is also demonstrated that by inflating the OLS residual by a factor of $\{n/(n-K)\}^{1/2}$ bootstrapping based upon the inflated residual will yield reliable estimates of the bias and dispersion of $\hat{\beta}$, provided that the sample mean of the OLS residual is zero.

The bootstrap algorithm for the LR model, whose purpose is to generate $\beta^*_{(j)}$ for $j=1,\ldots,J$, is described as follows:

Algorithm 2.1:

Purpose; To obtain $\beta^*_{(j)}$ for $j=1,\ldots,J$, where J is usually between 200 and 10,000.

Step 1. Obtain $\hat{\beta}$ and ε according to the specifications of (2.3.2) and (2.3.4) respectively.

Step 2. Choose an arbitrary double precision SEED. Then, set this value in the system through RANSET(), an internal subroutine of FORTRAN V.

Step 3. Generate n random integers, drawn with replacements from the set $(1,2,3,\ldots,n)$, by using the FORTRAN's random generator RANF(). This returns a different seed on each subsequent call.

Step 4. Construct $S_{(j)}$ according to definition 2.1.

Step 5. Reconstruct the linear responses as

$$y^*_{(j)} = X\hat{\beta} + S_{(j)}\hat{\varepsilon}. \tag{2.5.1}$$

Step 6. Compute

$$\beta^*_{(j)} = (X^TX)^{-1}X^Ty^*_{(j)}. \tag{2.5.2}$$

Step 7. Repeat steps 3,4,5, and 6 for $j=1,\ldots J$.

Attention is now directed to the bootstrap estimates,

$(\beta^*_{(j)}: j=1,\dots J)$. The following theorems concern the finite sample properties of bootstrap estimates of the regression coefficients, when OLS rsiduals are used for bootstrapping.

Theorem 2.8: Let $\hat{\varepsilon} = M\varepsilon$ be the OLS residual and its sample mean be given by $\bar{\hat{\varepsilon}} = \left\{\sum_{i=1}^{n}\left[n^{-1}\hat{\varepsilon}_i\right]\right\}$. Also, let n be finite. Then,

$$\hat{\beta}^* = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^*_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \hat{\beta} + (X^TX)^{-1}X^T\left[\bar{\hat{\varepsilon}}.E(n,1)\right].$$

Proof:

Upon substitution of (2.5.1) into (2.5.2),

$$\beta^*_{(j)} = \hat{\beta} + (X^TX)^{-1}X^TS_{(j)}\hat{\varepsilon}. \tag{2.5.3}$$

Consequently,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^*_{(j)}\right]\right\} = \hat{\beta} + (X^TX)^{-1}X^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\hat{\varepsilon}\right]\right\}.$$

The remainder of the proof is then straightforward upon application of Theorem 2.1. [See also Freedman (1981).] *Q.E.D.*

Generally, when the regression model has an intercept, $\bar{\hat{\varepsilon}}=0$. For any regression model without an intercept, $\bar{\hat{\varepsilon}}\neq0$. Corollary 2.8 illustrates the consequences for the mean of bootstrap estimates of β when $\bar{\hat{\varepsilon}}\neq0$ and when $\bar{\hat{\varepsilon}}=0$.

Corollary 2.7: Let $B(\hat{\beta}^*)$ be the sample bias of the bootstrap estimates, $(\beta^*_{(j)}; j=1,\dots J)$. Then, if the OLS residual is centered at mean zero, $B(\hat{\beta}^*)=0$. This is apparent from Theorem 2.8. Otherwise,

$$B(\hat{\beta}^*) = (X^TX)^{-1}X^T\left[\bar{\hat{\varepsilon}}.E(n,1)\right].$$

Theorem 2.9: Let n be finite and let

$$D_1(\beta^*) = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\hat{\beta})(\beta^*_{(j)}-\hat{\beta})^T\right]\right\}.$$

Then,

$$D_1(\beta^*) \xrightarrow[J]{a.s.} \left[n^{-1}(n-K)\right]D(\hat{\beta}) + \left\{(X^TX)^{-1}X^T\left[\bar{\varepsilon}^2.E(n,n)\right]X(X^TX)^{-1}\right\},$$

in which $D(\hat{\beta}) = \sigma^2(X^TX)^{-1}$ and $\hat{\sigma}^2 = \left[\hat{\varepsilon}^T\hat{\varepsilon}/(n-K)\right]$.

Proof:

From equation (2.5.3),

$$(\beta^*_{(j)}-\hat{\beta}) = (X^TX)^{-1}X^TS_{(j)}\hat{\varepsilon}.$$

Hence,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\hat{\varepsilon}\hat{\varepsilon}^TS_{(j)}\right]\right\} = \left[(X^TX)^{-1}X^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\hat{\varepsilon}\hat{\varepsilon}^TS^T_{(j)}\right]\right\}X(X^TX)^{-1}\right].$$

By applying Theorem 2.4 with m=n,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\hat{\varepsilon}\hat{\varepsilon}^TS_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[(n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}).I_n + (\bar{\varepsilon}^2)\left[E(n,n) - I_n\right]\right].$$

However, $(\hat{\varepsilon}^T\hat{\varepsilon})=\varepsilon^TM\varepsilon$, and $tr(M)=(n-K)$. This means that $\hat{\sigma}^2 = \left[(\hat{\varepsilon}^T\hat{\varepsilon}/(n-K)\right]$ is an unbiased estimator of σ^2. Consequently, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\hat{\beta})(\beta^*_{(j)}-\hat{\beta})^T\right]\right\} \xrightarrow[J]{a.s.} (X^TX)^{-1}X^T$$

$$\left\{\left[(n-K)n^{-1}\hat{\sigma}^2\right].I_n + (\bar{\varepsilon}^2).E(n,n)\right\}X(X^TX)^{-1}$$

and the R.H.S reduces to

$$\left[(n-K)n^{-1}\right].\hat{\sigma}^2(X^TX)^{-1} + \left\{(X^TX)^{-1}X^T\left[(\bar{\varepsilon}^2).E(n,n)\right]X(X^TX)^{-1}\right\}. \quad \textit{Q.E.D.}$$

Corollary 2.8: When the sample mean of the OLS residual is zero and OLS residuals are used for bootstrapping, $D_1(\beta^*)$ will underestimate $D(\hat{\beta})$ almost surely. Specifically, the ratio

$$\left\{\left[D_1(\beta^*) - D(\hat{\beta})\right]/D(\hat{\beta})\right\} \xrightarrow[J]{a.s.} (-K/n).$$

This result is straightforward from Theorem 2.9.

Theorem 2.10: Let n be finite. Then, using the notation of Theorem 2.8,

$$D_2(\beta^*) = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\bar{\beta}^*)(\beta^*_{(j)}-\bar{\beta}^*)^T\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[(n-K)n^{-1}\right]D(\hat{\beta})\right\},$$

in which $\bar{\beta}^* = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)})\right\}$.

Proof:

Using Theorems 2.1 and 2.9, the proof for this theorem is similar to the proof of Theorem 2.7. *Q.E.D.*

Theorem 2.8 states that when the sample mean of the OLS residuals is zero, bootstrapping based upon these residuals yields unbiased estimates of β. This happens when the regression model has an intercept. However, the bootstrap estimate of $D(\hat{\beta})$ tends to underestimate $D(\hat{\beta})$ almost surely. When the number of coefficients is small relative to n, this underestimation will be a negligible proportion of $D(\hat{\beta})$. On the other hand, when number of coefficients is large relative to n, the above underestimation of $D(\hat{\beta})$ can be significant. This underestimation of $D(\hat{\beta})$ by either $D_1(\beta^*)$ or $D_2(\beta^*)$ becomes more significant when the model does not have an intercept and when the OLS residuals are centered before bootstrapping.

Stine (1985) suggests rescaling the OLS residuals by dividing each residual by the square root of the corresponding main diagonal element of M. That is, let

$$\hat{r}_i = \hat{\varepsilon}_i(1-h_{ii})^{-1/2}$$

for $i=1,\ldots,n$, in which

$$h_{ij} = \left[\delta_{ij} - x_i(X^TX)^{-1}x_j^T\right]$$

where $i,j=1,\ldots,n$; and δ_{ij} is the Kronecker delta. Alternatively, the nx1 vector $\hat{r}$ comprising elements $\hat{r}_1,\hat{r}_2,\ldots,\hat{r}_n$ may be written as

$$\hat{r} = R\varepsilon \tag{2.5.4}$$

in which R comprises M with its r'th row divided by a factor $(1-h_{ii})^{1/2}$ $(i=1,\ldots,n)$. Unlike M, R is neither idempotent nor symmetric.

The bootstrap algorithm for the LR model using rescaled OLS residuals can be set out as follows:

Algorithm 2.2:

Purpose: The purpose of this algorithm is the same as that of algorithm 2.1, except that $\hat{\varepsilon}$ is now replaced by $\hat{r}$.

Steps 1 to 7 are the same as those of algorithm 2.1, except for step 5 which has the following modification.

Step 5. Reconstruct the linear responses as

$$y^*_{(j)} = X\hat{\beta} + S_{(j)}\hat{r}.$$

Note that the sample mean of $\hat{r}$ is non-zero, even when the sample mean of the OLS residual is zero. Thus, $\hat{r}$ has to be centered before bootstrapping and this can be done by subtracting from each element of $\hat{r}$

its sample mean. The following theorem considers the first and second moments of the bootstrap estimates of β, when the centered $\hat{r}$ is used for bootstrapping.

Theorem 2.11: Let r be centered such that its sample mean is zero and let

$$\beta^*_{(j)} = \hat{\beta} + (X^TX)^{-1}X^TS_{(j)}\hat{r}.$$

Further, let n be finite. Then,

$$\text{(i)} \quad \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^*_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \hat{\beta}\ ,$$

$$\text{(ii)} \quad D_3(\beta^*) = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\hat{\beta})(\beta^*_{(j)}-\hat{\beta})^T\right]\right\} \xrightarrow[J]{a.s.}$$

$$\left[\left[(n-K)n^{-1}\right] + 2\left\{\sum_{i>s}^{n}\left[n^{-1}R_{is}R_{si}\right]\right\}\right]D(\hat{\beta})$$

in which

$$R_{is} = \left\{(\delta_{is}-h_{is})(1-h_{ii})^{-1/2}\right\}$$

for $i,s=1,\ldots,n$.

Proof:

The proof for part (i) is straightforward upon application of Theorem 2.8. For part (ii), it is necessary to know that

$$tr(R^TR) = \left[(n-K) + 2\left\{\sum_{i>s}^{n}\left[R_{is}R_{si}\right]\right\}\right].$$

For any real nxn matrix R,

$$tr(R^TR) = \left\{\sum_{i=1}^{n}\sum_{s=1}^{n}\left[R_{is}R_{si}\right]\right\} = \left\{\sum_{i=1}^{n}\left[R^2_{ii}\right]\right\} + 2\left\{\sum_{i>s}^{n}\left[R_{is}R_{si}\right]\right\}.$$

However,

$$R^2_{ii} = \left[(1-h_{ii})^2(1-h_{ii})^{-1}\right] = (1-h_{ii})$$

and

$$\left\{\sum_{i=1}^{n}\left[1-h_{ii}\right]\right\} = tr(M) = (n-K).$$

The remainder of the proof is then straightforward upon application of Theorem 2.9. *Q.E.D.*

Corollary 2.9: When n is finite and J goes to infinity, $D(\beta^*_3)$ will underestimate (overestimate) $D(\hat{\beta})$ almost surely, when the sign of

$$\left[2\left\{\sum_{i>s}^{n}\left[R_{is}R_{si}\right]\right\} - K\right]$$

is negative (positive). Thus, Stine's rescaling scheme does not, in general, correct the underestimation problem of $D_2(\beta^*)$. This is obvious from part (ii) of Theorem 2.11.

Let the OLS residual be scaled by a factor $\{n(n-K)^{-1}\}^{1/2}$ and the transformed residual will be referred to, hereinafter, as the inflated OLS residual. The following theorem concerns the bootstrap estimate of $D(\hat{\beta})$ when inflated OLS residuals are used for bootstrapping. Nevertheless, the result of this theorem is strictly restricted to regresssion models with intercepts.

Theorem 2.12: Let the sample mean of ε be zero and

$$\beta^*_{(j)} = \hat{\beta} + \left[(X^TX)^{-1}X^TS_{(j)}\right]\left\{\left[n(n-K)^{-1}\right]^{1/2}\hat{\varepsilon}\right\}.$$

Then,

$$D_4(\beta^*) = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\beta^*_{(j)}-\hat{\beta})(\beta^*_{(j)}-\hat{\beta})^T\right]\right\} \xrightarrow[J]{a.s.} D(\hat{\beta}).$$

Proof:

The proof for this theorem is straightforward from the proof of Theorem 2.9. *Q.E.D.*

Theorem 2.2 states that when inflated OLS residuals are used for bootstrapping, the bootstrap estimate of $D(\hat{\beta})$ converges almost surely to its conventional least-squares estimate, as J goes to infinity.

Although OLS residuals can be used for bootstrapping, one needs to be careful when applying these residuals. For a regression model which has an intercept, these residuals have to be multiplied by the factor $\left\{n(n-K)^{-1}\right\}^{1/2}$. Additional care should be taken when the model does not have an intercept. For this type of regression model, the residuals have to be centered by subtracting from their original values their sample mean and the resulting values have to be multiplied by the factor $\left\{n^2(n-1)^{-1}(n-K)^{-1}\right\}^{1/2}$. When the original OLS residuals are used for bootstrapping, the consequence will be an underestimation of the dispersion of $\hat{\beta}$. The severity of the consequence for not following the correct procedure depends on both the sample size and the number of coefficients to be estimated. For example, when n=100 and K=2, the underestimation is only two percent. On the other hand, when n=10 and K=5, the underestimation will be an enormous fifty percent.

2.6 Bootstrapping the LR model: The Case of BLUS Residuals

In the preceding section, bootstrap results based upon untreated OLS residuals have been shown to have poor properties. When the sample size is small, tests based upon such results are misleading. However, bootstrap results based upon well-conditioned OLS residuals are equivalent to the conventional least-squares results, at least for the first two moments of $\hat{\beta}$. These residuals can be obtained by applying simple transformations to the OLS residuals. Alternatively, one can also use any class of residuals that satisfy A.2.2 and A.2.3. One of these is the class of BLUS residuals (see e.g., Theil, 1965). The acronym BLUS stands for Best Linear Unbiased with Scalar covariance matrix. The BLUS residual vector is linear in y and its second moment is an unbiased estimator of the exact second moment of ε (see e.g., Koerts, 1967 p.170).

The objective of the BLUS approach is to construct a partial isometry of order (n-K)xn such that, (i) for every x lying orthogonal to the span of X, the norm of Bx equals the norm of x, i.e. $x^TB^TBx = x^Tx$, and (ii) for every x in the span of X, $Bx=0$. Then, $BB^T=I_{n-k}$ and $B^TB=M$. Thus, B has rank (n-K).

Let X be partitioned as

$$X^T = \left[X_0^T \vdots X_1^T\right]$$

where X_0 consists of the first K rows of X and let X_1 be the remainder of X. X_0 is assumed to be nonsingular. Subsequently, B can be constructed as [see e.g., Neudecker (1969, p.949)]

$$B = \left[0_{n,k} B_1\right]$$

in which

$$B_1^T = \left[M_{11}^{-1/2} M_{10} M_{11}^{-1/2}\right] ,$$

$$M_{10} = \left[-X_1(X^TX)^{-1}X_0^T\right] ,$$

$$M_{11} = \left\{I_{n-k} - X_1 (X^T X)^{-1} X_1^T\right\}.$$

Attention is now focussed on By, the (n-K)x1 vector of BLUS residuals. It is shown below that this residual is suitable for bootstrapping the LR model, provided that its sample mean is zero. However, its sample mean is usually non-zero, even when there is an intercept in the regression model. When the sample mean of the BLUS residuals is not zero, it will also be shown below that BLUS residuals will still be suitable for bootstrapping, provided that an appropriate transformation is applied to them prior to bootstrapping. For notational simplicity, $\hat{\varepsilon}$ is again used here but it should not be confused with $\hat{\varepsilon}$ in the previous section. Note that there are only (n-K) BLUS residuals, as compared to n OLS residuals. However, bootstrap samples of size n are still required and these have to be selected with replacement from a set of (n-K) elements.

The following theorems examine the finite sample properties of bootstrap estimates of the regression coefficients, based upon BLUS residuals.

Theorem 2.13: Let $\hat{\varepsilon}$=By and $\bar{\hat{\varepsilon}} = \left[\frac{1}{L}\sum_{i=1}^{L}(\hat{\varepsilon}_i)\right]$ in which L=(n-K). Then,

$$\hat{\beta}^*_{(j)} = \hat{\beta} + \left\{\left[(X^T X)^{-1} X^T S_{(j)} \hat{\varepsilon}\right]\right\}.$$

Then, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\hat{\beta}^*_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \hat{\beta} + (X^T X)^{-1} X^T \left[\bar{\hat{\varepsilon}}.E(n,1)\right].$$

Proof:

There is a similarity between this theorem and Theorem 2.8. The only difference is that $S_{(j)}$ is now nxL instead of nxn. Upon application of Lemma 2.1 and letting n=L and m=n,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[(L^{-1}).E(n,L)\right]. \quad Q.E.D.$$

Corollary 2.10: When the sample mean of the BLUS residuals is zero, the finite sample bias of the bootstrap estimates, $\left[\hat{\beta}^{*}_{(j)}; j=1,\ldots,J\right]$, will also be zero. This is a direct result from Theorem 2.13.

Theorem 2.14: Let n be finite. Then, using the notation of Theorem 2.13,

$$D_{S}(\beta^{*}) = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\beta}^{*}_{(j)}-\hat{\beta})(\hat{\beta}^{*}_{(j)}-\hat{\beta})^{T}\right]\right\} \xrightarrow[J]{a.s.} D(\hat{\beta})$$

$$+ (X^{T}X)^{-1}X^{T}\left\{\bar{\hat{\varepsilon}}\left[E(n,n)-I_{n}\right]\right\} X(X^{T}X)^{-1}.$$

Proof:

The proof for this theorem is similar to that of Theorem 2.9, except that

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\hat{\varepsilon}\hat{\varepsilon}^{T}S^{T}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left[L^{-1}(\hat{\varepsilon}^{T}\hat{\varepsilon}).I_{n}\right] + \bar{\hat{\varepsilon}}^{2}\left[E(n,n)-I_{n}\right].$$

Also, $\hat{\sigma}^{2}=(\hat{\varepsilon}^{T}\hat{\varepsilon}/L)$ is an unbiased estimator of σ^{2}. Thus, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\beta}^{*}_{(j)}-\hat{\beta})(\hat{\beta}^{*}_{(j)}-\hat{\beta})^{T}\right]\right\} \xrightarrow[J]{a.s.} \hat{\sigma}^{2}(X^{T}X)^{-1}$$

$$+ (X^{T}X)^{-1}X^{T}\left\{\bar{\hat{\varepsilon}}^{2}\left[E(n,n)-I_{n}\right]\right\}X(X^{T}X)^{-1}. \quad Q.E.D.$$

Theorem 2.14 states that when the sample mean of the BLUS residuals is zero and, when n is finite, the bootstrap estimate of the dispersion of β approaches $D(\hat{\beta})$ almost surely, as J goes to infinity. However, the sample mean of the BLUS residuals is usually non-zero. Both Theorem 2.5 and 2.6 below consider the bootstrap estimation of $D(\hat{\beta})$ when the sample mean of the BLUS residuals is non-zero. The results of both theorems are identical, despite the following differences. Theorem 2.5 assumes that the original BLUS residuals are being used for bootstrapping and that the OLS estimate of β in Theorem 2.14 is replaced by the mean of the bootstrap estimates of β. On the other hand, Theorem 2.16 assumes that the centered BLUS residuals are being used for bootstrapping and that the OLS estimate of β is retained.

Theorem 2.15: Let $\bar{\beta}^{*} = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}\right]\right\}$. Then, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\beta}^{*}_{(j)}-\bar{\beta}^{*})(\hat{\beta}^{*}_{(j)}-\bar{\beta}^{*})^{T}\right]\right\} \xrightarrow[J]{a.s.} zD(\hat{\beta})$$

in which $z=\left[L^{-1}(L-1)\right]$.

Proof:

The proof is straightforward from the proof of Theorem 2.7

Q.E.D.

Theorem 2.16: Let ε be the centered BLUS residuals, having sample mean zero and let $\hat{\varepsilon}$ be used for bootstrapping. Further, let n be finite. Then,

$$E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\beta}^{*}_{(j)}-\hat{\beta})(\hat{\beta}^{*}_{(j)}-\hat{\beta})^{T}\right]\right\} \xrightarrow[J]{a.s.} zD(\hat{\beta}) \text{ in which } z=\left[L^{-1}(L-1)\right].$$

Proof:

The proof is straightforward from the proof of Theorem 2.6, by taking note that

$$E\left[L^{-1}(\hat{\varepsilon}^T\hat{\varepsilon})\right] = z\sigma^2. \qquad Q.E.D.$$

The results of Theorems 2.15 and 2.16 indicate that, when the BLUS residuals having non-zero sample mean are used for bootstrapping, bootstrap estimates of $D(\hat{\beta})$ will underestimate $D(\hat{\beta})$ almost surely, as J goes to infinity. However, an unbiased bootstrap estimate of $D(\hat{\beta})$ can still be obtained by scaling the bootstrap estimate of $D(\hat{\beta})$ in either Theorem 2.15 or 2.16 by a factor $\left[L(L-1)^{-1}\right]$. Alternatively, it can easily be shown that the same result can be obtained by either (i) scaling the bootstrap estimates of β by a factor $\left[L(L-1)^{-1}\right]^{1/2}$ or (ii) scaling the centered BLUS residuals by the same factor.

2.7 Summary

In this chapter a link has been established between the classical, Monte Carlo and bootstrap results for the LR model as defined in Section 2.2. In the LR model, the true disturbances are unobservable in empirical applications to real world data, but the OLS residuals can easily be computed. If the sample size is sufficiently large, the use of OLS residual for bootstrapping the LR model poses no great problem. This is evident in Theorems 2.9 and 2.10. On the contrary, when the sample size is small, the same theorems indicate that the bootstrap results based upon OLS residuals will lack the properties to yield reliable inferences. Hence, OLS residuals should not be used for bootstrapping.

Stine (1985) recognized that OLS residuals are not suitable for bootstrapping. However, Stine attributed the problem to the fact that OLS residuals do not have a scalar covariance matrix and, hence, that they are not i.i.d. Consequently, Stine introduced a rescaling scheme which he claimed would produce better results. This rescaling scheme has been used by Wu (1986, p.1282), when applying the bootstrap technique to regression analyses. In Theorem 2.11 and Corollary 2.9, it is shown that this rescaling scheme can lead to further complications. Thus, Stine's rescaling scheme should also be avoided.

Freedman and Peters (1984a) have advocated the use of either BLUS or inflated OLS residuals, but neither of these residuals has actually been used in the literature. For the BLUS residuals, one likely reason is the computational burden; another is the absence of any ready made BLUS subroutine. A common reason regarding the non-usage of both residuals in bootstrapping is the lack of a coherent finite sample theory for bootstrapping LR models. This leads to the notion that the errors caused by the use of OLS residuals in bootstrapping the LR model are not serious. Yet, from Theorems 2.9 and 2.10, it is clear that when the sample size is small, the problem can be serious. It has been shown in Theorem 2.12, that by scaling the OLS residuals by a factor $\left[n(n-K)^{-1}\right]^{1/2}$, the bootstrap estimates of the regression coefficients will have the desirable first and second moments. The use of BLUS residuals for bootstrapping has also been considered here. The results of Theorems 2.14, 2.15 and 2.16 suggest that bootstrap results based upon BLUS residuals also have desirable properties, provided that appropriate adjustments are made prior to bootstrapping.

In this chapter, attention has been focussed on the means and variances of bootstrap estimates. Higher moments of bootstrap estimates of the regression coefficients will be studied in Chapter 3. Bootstrap prediction (and confidence) intervals are difficult to evaluate analytically. Hence, a simulation study is needed to complete the analysis of the application of Efron's bootstrap method to LR model, when the sample size is small. This will be done in Chapter 4 below. In Chapter 5, the results of this chapter and the following two chapters are applied to a multiplicative Cobb-Douglas regression model. AR(1) models are common in applied econometric works and Chapter 6 examines the problems of applying bootstrap methods to these models. In both Chapters 5 and 6, problems relating to the construction of bootstrap confidence intervals and the bootstrap confidence intervals themselves will be discussed and alternative solutions will be suggested.

The bootstrap method is useful especially in cases when analytical formulas are not readily available. For example, in seemingly unrelated and nonlinear regression models, the exact variance-covariance matrices of least-squares estimates of the regression coefficients are difficult to obtain. These difficulties can easily be overcome by applying the bootstrap method, provided that the appropriate regression residuals are used for bootstrapping. However, the application of bootstrap methods to these problems is left for future research.

Applications of the above results are not restricted to regression models. For examples, these results can be applied to the estimation of tail indices as in Kryzanowski, Rahman and Sim (1986) and to the weighted-mean problem. Finally, the selection matrix introduced in this chapter can be used to obtain extensions of the above results for accommodating other interesting problems. For examples, extensions can be made to study the bootstrap distributions of sample correlation coefficients, sample means, sample variances and covariances, and robust estimators.

CHAPTER THREE

BOOTSTRAP DISTRIBUTION OF $\hat{\beta}$ AND HIGHER MOMENTS OF THE BOOTSTRAP ESTIMATES OF β

3.1 Introduction

In Chapter 2, it is shown that bootstrapping leads to unbiased estimates of the dispersion of $\hat{\beta}$, provided that either inflated OLS or Theil's (1965) BLUS residuals are used for bootstrapping. When the mean of the OLS residuals is zero, bootstrapping always leads to unbiased estimates of β. In this type of situation, the mean of bootstrap estimates of β approaches $\hat{\beta}$ when J, the number of bootstrap replications, goes to infinity.

As a sequel to Chapter 2, the focus is now on the bootstrap distribution of $\hat{\beta}$. In order to determine the closeness of the bootstrap distribution of $\hat{\beta}$ to its exact distribution, higher moments of the bootstrap estimates of β are examined. The accuracy of this approach is subject to the assumption that the probability density function of $\hat{\beta}$ is completely determined by its moments. Theoretically, a set of moments determines a distribution uniquely only under certain conditions. Fortunately, these conditions are met by most of the distributions commonly arising in statistical practice, with the exception of the lognormal distribution (see Kendall and Stuart, 1977, pp. 89 and 192). The assumption that the probability density function of $\hat{\beta}$ is uniquely determined by its moments is maintained throughout this chapter. Thus, in order for the bootstrap distribution of $\hat{\beta}$ to qualify as an exact distribution of $\hat{\beta}$, all moments of the bootstrap estimates of β must be exactly the same as the corresponding true moments of $\hat{\beta}$.

Bootstrap distributions had been studied by among others, Singh (1981), Babu and Singh (1983), Abramovitch and Singh (1985) and Hall (1987). Let θ be the parameter of interest and $\hat{\theta}_n$ be its sample estimate. Further, let υ_n^2 be a consistent estimate of σ^2, where σ^2 is the variance of $\hat{\theta}_n$. Singh (1981) gives assymptotic results for the general bootstrap distribution of $\hat{\theta}_n$, whereas Babu and Singh (1983) compare the bootstrap distributions of $(\hat{\theta}_n-\theta)$ and $\upsilon_n^{-1}(\hat{\theta}_n-\theta)$. Abramovitch and Singh (1985) used information from the bootstrap distribution of $\upsilon_n^{-1}(\hat{\theta}_n-\theta)$ to obtain better approximations of the distribution of $\hat{\theta}_n$. Recently, Hall (1987) proposes a continuity correction procedure for obtaining smooth bootstrap distributions of $\upsilon_n^{-1}(\hat{\theta}_n-\theta)$.

All the above studies are based on Edgeworth expansion approach. Moreover, most of the results were obtained for the case when $\hat{\theta}_n$ is the arithmetic mean of n independent observations on a particular random variable. The present approach focuses on bootstrap distributions of $\hat{\beta}$; in particular, on sample moments of bootstrap estimates of β and on some of the moments of $\hat{\beta}$.

It will be shown below that, for a fixed design matrix, X, the moments of $\hat{\beta}$ are uniquely determined by a set of moments of the underlying disturbances, ε. Similarly, it can be shown that the moments of the bootstrap estimates of β are uniquely determined by a set of moments of the residuals, $(\hat{\varepsilon}_1,\hat{\varepsilon}_2,\ldots,\hat{\varepsilon}_n)$, that are used for bootstrapping. Consequently, when the moments of $\hat{\varepsilon}$ correspond exactly with the moments of ε, one can conclude that the bootstrap distribution of β corresponds exactly with the distribution of $\hat{\beta}$.

Notation and assumptions will be established in Section 3.2. Extended properties of the selection matrix, $S_{(j)}$, which are relevant to the discussions of this chapter will also be presented in this section. In Section 3.3, higher moments of the bootstrap estimates of β will be examined analytically. Owing to mathematical complexity, only the third and fourth moments will be considered. In statistical practice, sample moments of order higher than the fourth are rarely required and in many cases are subjected to a large margin of error (see, e.g. Kendall and Stuart, 1977, p. 58). Consequently, these higher moments are of little practical value. Further, the influence of a particular sample moment of $\hat{\beta}$ on the distribution of $\hat{\beta}$ declines as its order increases. For comparison purposes, a Monte Carlo simulation study will be conducted in Section 3.4. In this section, the first ten sample moments of several regression residuals will be compared to the corresponding exact moments of the underlying disturbances. The underlying disturbances are assumed to be normally distributed. Finally, a brief discussion of the results obtained in this chapter and some concluding remarks will be presented in Section 3.5.

The objective of this chapter is to examine the bootstrap estimates of β beyond the first two moments. One of the main contributions here is to show that the sample moments of the bootstrap estimates of β depend on the sample moments of the regression residuals that are being used for bootstrapping. Thus, one only needs to examine the sample moments of the regression residuals when the objective is to examine the bootstrap distribution of $\hat{\beta}$. This leads to much simplification in general and a reduction in computational costs for a Monte Carlo study.

3.2 Preliminaries and Notation

The focus of this chapter is on the closeness of the distribution of the bootstrap estimates of β to the exact distribution of $\hat{\beta}$. A review of the following statistical theorems may be helpful. The notation is mostly that of Kendall and Stuart (1977, vol.1), [hereinafter, K&S].

The Inversion Theorem: The characteristic function of a random variable uniquely determines its distribution function

Proof: See K&S p. 97.

The First Limit Theorem: If a sequence of distribution functions $\{F_n\}$ tends to a continuous distribution function F, then the corresponding sequence of characteristic functions of $\{F_n\}$ tend to the characteristic function of F uniformly in any finite t-interval.

Proof: See K&S p. 107.

Converse of the First Limit Theorem: If the corresponding sequence of characteristic functions of $\{F_n\}$ tends to the characteristic function of F uniformly in any finite t-interval, then the sequence of distribution functions $\{F_n\}$ tends to F.

Proof: See K&S p. 108.

The Second Limit Theorem: Let $\{\mu_j(n)\}$ and μ_j be the j'th moments of the sequence $\{F_n\}$ and F respectively, and j be non-negative. Then, if $\{F_n\}$ converges to F, all the moments $\{\mu_j(n)\}$ converge to μ_j respectively.

Proof: See K&S p. 118.

Converse of the Second Limit Theorem: Let $\{\mu_j(n)\}$ and μ_j be the j'th moments of the sequence $\{F_n\}$ and F respectively, and j be non-negative. Also, let the moments $\{\mu_j(n)\}$ exists. If the sequence $\{\mu_j(n)\}$ converges to μ_j for all j, as n goes to infinity, then the sequence $\{F_n\}$ converges to F, provided F is uniquely determined by its moments.

Proof: See K&S p. 119.

The following lemmas may also be stated:

Lemma 3.1: Let r be some function of n and J, where J is the number of bootstrap replications. Also, let $\{G_r\}$ converges to G, the corresponding sequence of characteristic functions of $\{G_r\}$ converges to the characteristic function of G. The converse is also true.

Proof:

The proof is apparent from the First Limit Theorem and its converse. *Q.E.D.*

Lemma 3.2: Let r be as defined in Lemma 3.1. Also, let the moments $\{\mu_j(r)\}$ corresponding to $\{G_r\}$ exist. If the sequence $\{\mu_j(r)\}$ converges to μ_j for all j, as both n and J go to infinity, then the sequence $\{G_r\}$ converges to G, provided G is uniquely determined by its moments. The converse is also true.

Proof:

The proof is straightforward from the Second Limit Theorem and its converse. *Q.E.D.*

Although the condition in Lemma 3.1 is less restrictive than Lemma 3.2, the latter has more practical value. This is because sample moments are easier to compute than characteristic functions. With Lemma 3.2, one needs only to examine how closely the sample moments of the bootstrap estimates of β mimic the exact moments of $\hat{\beta}$. Nevertheless, Lemma 3.2 is useful if the distribution function of $\hat{\beta}$ is uniquely determined by its moments. This happens if and only if all absolute moments of G exist. This condition can also be stated as:

Lemma 3.3: Let G be the distribution of $\hat{\beta}$ and μ_j be the j'th moment of G. Also, let the characteristic function of G be written as

$$c(t) = \sum_{j=0}^{\infty} \left\{(it)^j \mu_j / j!\right\}$$

where $i^2=-1$. Then if the series on the right hand side of the above equation converges for $t \neq 0$, a set of moments will determine the distribution function G uniquely.

Proof: See K&S p. 113-114.

In addition to the above lemmas, the following lemmas concerning the extended properties of the selection matrix $S_{(j)}$ are also needed for subsequent discussions in this chapter.

Lemma 3.4: Let $S_{(j)}$ be an nxn selection matrix, and let $S_{(j)r.}$ be its r'th row $(r=1,2,\ldots,n)$, comprising elements $S_{(j)r1}, S_{(j)r2}, \ldots, S_{(j)rn}$. In addition, let υ be an nx1 vector of real arguments υ_1, υ_2, $\ldots, \upsilon_n$ and let

$$\bar{\upsilon} = n^{-1}\left[\upsilon_1 + \upsilon_2 + \ldots \upsilon_n\right].$$

Let n be finite. Then,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[S_{(j)r.}\upsilon\right]\right\} \xrightarrow[J]{a.s.} \bar{\upsilon}.$$

Proof:

Note that for each $j=1,2,\ldots,J$, an element is being drawn at random with replacement from υ, a sample of size n. The relative frequency of each element of υ in each bootstrap sample is (n^{-1}). For J samples, the relative frequency of each element being selected will be $(n^{-1}J)$. Thus,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[S_{(j)r.}\upsilon\right]\right\} \xrightarrow[J]{a.s.} \frac{1}{J}\left\{\sum_{i=1}^{n}(\upsilon_{i})\right\}(n^{-1}J) = \bar{\upsilon}. \quad Q.E.D.$$

Lemma 3.5: Given the notation of Lemma 3.4 and n finite,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon)\right]\right\} \xrightarrow[J]{a.s.} \left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_{i}^{2}\right]\right\} \quad \text{for } q=r$$

and

$$\xrightarrow[J]{a.s.} \left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_{i}\right]\right\}^{2} \quad \text{for } q\neq r.$$

Proof:

When q=r,

$$\left\{(S_{(j)r.}\upsilon)(S_{(j)r.}\upsilon)\right\} = S_{(j)r.}\left[\upsilon_{i}^{2}\right]$$

in which $\left[\upsilon_{i}^{2}\right]$ denotes an nx1 column vector whose i'th observation is the square of the i'th observation of υ. Given Lemma 3.4, the remainder of the proof for the first statement is straightforward.

For the case when q≠r,

$$(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon) = \upsilon_a\upsilon_b$$

for some $a,b=1,2,\ldots,n$. Note that the matrix

$$\upsilon\upsilon^T = \left[\ \upsilon_a\upsilon_b\ \right],\quad a,b=1,2,\ldots,n,$$

has n^2 elements. The relative frequency of each $\upsilon_a\upsilon_b$ in each bootstrap sample is n^{-2}. For J samples, the relative frequency of each element is $(n^{-2}J)$. Thus,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon)\right]\right\} \xrightarrow[J]{a.s.} \frac{1}{J}\left\{\sum_{a,b}^{n}(\upsilon_a\upsilon_b)\right\}(n^{-2}J) = \bar{\upsilon}^2. \qquad Q.E.D.$$

Lemma 3.6: Using the notation of Lemma 3.4, for n finite,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[(S_{(j)p.}\upsilon)(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon)\right]\right\} \xrightarrow[J]{a.s.}$$

(i) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^3\right]\right\}$ when p=q=r,

(ii) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i\right]\right\}\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^2\right]\right\}$ when any pair of p, q and r are the same,

(iii) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i\right]\right\}^3$ when p≠q≠r.

Proof:

In the first case where p=q=r,

$$\left[(S_{(j)r.}\upsilon)(S_{(j)r.}\upsilon)(S_{(j)r.}\upsilon)\right] = S_{(j)r.}\left[\upsilon_i^3\right]$$

in which $\left[\upsilon_i^3\right]$ denotes an nx1 column vector whose i'th observation is the third power of the i'th observation of υ. The remainder of the proof for the first statement is then straightforward upon application of Lemma 3.4.

For the case when any two of the vectors $S_{(j)i.}$ $(i=p,q,r)$ are identical,

$$\left[(S_{(j)p.}\upsilon)(S_{(j)r.}\upsilon)(S_{(j)r.}\upsilon)\right] = \left[(S_{(j)p.}\upsilon)(S_{(j)r.}\upsilon)\right],$$

in which $p\neq r$. From this point, the remainder of the proof for the second statement is straightforward upon application of Lemma 3.5.

Finally, for the case when none of the vectors $S_{(j)i.}$ $(i=p,q,r)$ are the same,

$$\left[(S_{(j)p.}\upsilon)(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon)\right] = \left[\upsilon_a\upsilon_b\upsilon_c\right],$$

for some $a,b,c=1,2,\ldots,n$. Note that the three-dimensional array $\left[\upsilon_a\upsilon_b\upsilon_c\right]$ is of dimension nxnxn and, in each bootstrap sample the relative frequency of $\left[\upsilon_a\upsilon_b\upsilon_c\right]$ is n^{-3}. For J samples, the relative frequency of each element is $(n^{-3}J)$. Thus, for n finite,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[(S_{(j)p.}\upsilon)(S_{(j)q.}\upsilon)(S_{(j)r.}\upsilon)\right]\right\} \xrightarrow[J]{a.s.} \frac{1}{J}\left\{\sum_{a,b,c}^{n}\left[\upsilon_a\upsilon_b\upsilon_c\right]\right\}(n^{-3}J) = \bar{\upsilon}^3.$$

Q.E.D.

Lemma 3.7: Let n be finite. Then,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\left[(S_{(j)h_1.}\upsilon)(S_{(j)h_2.}\upsilon)(S_{(j)h_3.}\upsilon)(S_{(j)h_4.}\upsilon)\right]\right\} \xrightarrow[J]{a.s.}$$

(i) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^4\right]\right\}$ when $h_1=h_2=h_3=h_4$,

(ii) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^3\right]\right\}\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i\right]\right\}$ when any three of the h_i's are identical,

(iii) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^2\right]\right\}^2$ when any pair of the h_i's are identical and the remaining pair are also identical but not the same as the first pair.

(iv) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^2\right]\right\}\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i\right]\right\}^2$ when only two of the h_i's are identical,

(v) $\left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i\right]\right\}^4$ when none of the h_i's is the same as another.

Proof:

The proof is similar to that of Lemma 3.6. *Q.E.D.*

Lemma 3.8: Let

$$\left\{\prod_{r=1}^{R}\left[S_{(j)h_r.}\upsilon\right]\right\} = \left[(S_{(j)h_1.}\upsilon)(S_{(j)h_2.}\upsilon)\ldots(S_{(j)h_R.}\upsilon)\right]$$

and , let μ_t be defined by

$$\mu_t = \left\{\sum_{i=1}^{n}\left[n^{-1}\upsilon_i^t\right]\right\}.$$

Then, for n finite,

$$\frac{1}{J}\left\{\sum_{j=1}^{J}\prod_{r=1}^{R}\left[S_{(j)r.}\upsilon\right]\right\} \xrightarrow[J]{a.s.}$$

(i) μ_R when $h_1 = h_2 = \ldots h_R$,

(ii) $\mu_{(R-1)}\mu_1$ when (R-1) of the h_i's are identical,

(iii) $\mu_{(R-2)}\mu_1^2$ when (R-2) of the h_i's are identical,

(iv) $\mu_{(R-2)}\mu_2$ when (R-2) of the h_i's are identical and the remaining pair are also identical but not the same as the other (R-2),

(v) $\mu_{(R-3)}\mu_1^3$ when (R-3) of the h_i's are identical,

(vi) $\mu_{(R-3)}\mu_2\mu_1$ when (R-3) of the h_i's are identical, and two of the remaining three are similar but not the same as the other (R-3),

(vii) $\mu_{(R-3)}\mu_3$ when (R-3) of the h_i's are identical, and the remaining three are similar to other but not the same as the other (R-3). The list continues until

(viii) $\mu_2\mu_1^{(R-2)}$ when any two of the h_i's are identical,

(ix) μ_1^R when none of the h_i's are similar.

Proof:

The proof is similar to, and can be deduced from, the proof of Lemma 3.6. *Q.E.D.*

For a real vector υ, Lemma 3.4 relates the mean of the r'th observation in a bootstrap sample, $S_{(j)}\upsilon$ ($j=1,\ldots,J$), to the sample mean of υ. It must be mentioned that this lemma is a slight variant of Theorem 2.1. Lemma 3.5 relates the second moments of observations in a bootstrap sample to the first and second sample moments of υ. The results of Lemma 3.5 agree with those of Theorem 2.4. Similarly, Lemma 3.6 relates the third moments of observations in a bootstrap sample to the first three sample moments of υ. Lemma 3.7 extends the results to the fourth moments of observations in a bootstrap sample, while Lemma 3.8 generalizes to the m'th moments.

Lemmas 3.4 through to 3.7 are needed for analyzing the higher moments of the bootstrap estimates of β. It must be mentioned that these lemmas will be applied to the regression residuals, $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_n)$, that are used for bootstrapping. Thus, in the application of the above lemmas, one only needs to replace υ by $\tilde{\varepsilon}$. This i possible because $\hat{\beta}$ is a linear function of ε. As a result, $\beta^{*}_{(j)}$ is a linear function of $\tilde{\varepsilon}$.

In the subsequent section, a comparison will be made between the relation of $\hat{\beta}$ with ε and the relation of $\beta^{*}_{(j)}$ and $\tilde{\varepsilon}$. The two relations share common similarities but they are not exactly the same. The focus of the following section will be to investigate how these similarities and differences affect the bootstrap distribution of $\hat{\beta}$.

3.3 Higher Moments of the Bootstrap Estimates

In Chapter 2, it is shown that, when $\hat{\varepsilon}$ is the OLS residuals, $\left[n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}\right]$ underestimates σ^2. This causes the bootstrap estimate of $D(\hat{\beta})$ to be biased downward, and is true when, in fact, the bootstrap estimates of β are obtained by bootstrapping $\hat{\varepsilon}$. This deficiency can be corrected either by inflating $\hat{\varepsilon}$ by a factor depending on the number of coefficients to be estimated or, by using BLUS residuals. One need not go further when the sole interest is to obtain an unbiased estimate of $D(\hat{\beta})$. On the other hand, a confidence interval of β is often required for testing among competing (nested or non-nested) hypotheses and in regression diagnostics. From the Converse of the Second Limit Theorem, it is clear that this confidence interval is an unique function of the moments of $\hat{\beta}$. However, the importance of a sample moment of $\hat{\beta}$ to this declines as its order increases.

The following lemmas and theorems investigate the relation between the sample moments of the bootstrap estimates of β and the moments of $\hat{\beta}$. It is assumed throughout the investigation that the design matrix, X, is fixed. Consequently, $\hat{\beta}$ can also be expressed as

$$\hat{\beta} = \beta + A\varepsilon$$

in which $A = \left[a_{hi}\right] = \left[(X^TX)^{-1}X^T\right]$ is a real Kxn matrix and ε is a nx1 vector of true disturbances. It is convenient to derive the moments of $\hat{\beta}$ about β, which happens also to be its mean. Let $\hat{t}=(\hat{\beta}-\beta)$. Then, $E(\hat{t}_h^m)$ gives the m'th central moment of $\hat{\beta}_h (h=1,\dots,K)$.

Since the proofs are rather straightforward, the following lemma will be stated without proof.

Lemma 3.9: Let $E(\varepsilon_i)=0 \ \forall\ i$. Then, for n finite,

(i) $$E\left[\hat{t}_b\hat{t}_c\hat{t}_d\right] = \left\{\sum_{i=1}^{n}\left[a_{bi}a_{ci}a_{di}\right]\right\}\mu_3,$$

(ii) $$E\left[\hat{t}_b\hat{t}_c\hat{t}_d\hat{t}_e\right] = \left\{\sum_{i=1}^{n}\left\{\left[a_{bi}a_{ci}a_{di}a_{ei}\right]\right\}\left[\mu_4-3\mu_2^2\right]\right\}$$

$$+ \left[\sum_{i=1}^{n}\left(a_{bi}a_{ci}\right)\left\{\sum_{i=1}^{n}\left(a_{di}a_{ei}\right)\left(\mu_2^2\right)\right\}\right]$$

$$+ \left[\sum_{i=1}^{n}\left(a_{bi}a_{di}\right)\left\{\sum_{i=1}^{n}\left(a_{ci}a_{ei}\right)\left(\mu_2^2\right)\right\}\right]$$

$$+ \left[\sum_{i=1}^{n}\left(a_{bi}a_{ei}\right)\left\{\sum_{i=1}^{n}\left(a_{ci}a_{di}\right)\left(\mu_2^2\right)\right\}\right].$$

The above lemma gives the moments of $\hat{\beta}$ while the following lemma gives the samples moments of the bootstrap estimates of β. For the following lemma, let $\tilde{\varepsilon}_i\,(i=1,\ldots,n)$ be the residuals used for bootstrapping and the corresponding bootstrap estimates of β be

$$\beta^{*}_{(j)} = \tilde{\beta} + AS_{(j)}\tilde{\varepsilon}_i$$

for the j'th $(j=1,\ldots,J)$ replication. The sample mean of the ε_i's which is given by

$$\tilde{\mu}_1 = \left\{\sum_{i=1}^{n}\left(n^{-1}\tilde{\varepsilon}_i\right)\right\}$$

is assumed to be zero. Also, let

$$\text{(i)}\quad \tilde{\mu}_2 = \left\{\sum_{i=1}^{n}\left(n^{-1}\tilde{\varepsilon}_i^2\right)\right\},$$

$$\text{(ii)}\quad \tilde{\mu}_3 = \left\{\sum_{i=1}^{n}\left(n^{-1}\tilde{\varepsilon}_i^3\right)\right\},$$

$$\text{(iii)}\quad \tilde{\mu}_4 = \left\{\sum_{i=1}^{n}\left(n^{-1}\tilde{\varepsilon}_i^4\right)\right\}.$$

Lemma 3.10: Let the above notation be used and let n be finite. Then,

(i) $$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[t_{(j)b}t_{(j)c}t_{(j)d}\right]\right\} \xrightarrow[J]{a.s.} \left\{\sum_{i=1}^{n}\left[a_{bi}a_{ci}a_{di}\right]\right\}\tilde{\mu}_3,$$

(ii) $$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[t_{(j)b}t_{(j)c}t_{(j)d}t_{(j)e}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[\sum_{i=1}^{n}\left[a_{bi}a_{ci}a_{di}a_{ei}\right]\right]\left[\tilde{\mu}_4-3\tilde{\mu}_2^2\right]\right\}$$

$$+\left\{\left[\sum_{i=1}^{n}\left[a_{bi}a_{ci}\right]\right]\left[\sum_{i=1}^{n}\left[a_{di}a_{ei}\right]\left[\tilde{\mu}_2^2\right]\right]\right\}$$

$$+\left\{\left[\sum_{i=1}^{n}\left[a_{bi}a_{di}\right]\right]\left[\sum_{i=1}^{n}\left[a_{ci}a_{ei}\right]\left[\tilde{\mu}_2^2\right]\right]\right\}$$

$$+\left\{\left[\sum_{i=1}^{n}\left[a_{bi}a_{ei}\right]\right]\left[\sum_{i=1}^{n}\left[a_{ci}a_{di}\right]\left[\tilde{\mu}_2^2\right]\right]\right\}$$

in which $t_{(j)k} = \left[\beta^*_{(j)k}-\beta_k\right]$ $\forall$ $k=1,\ldots,K$.

Proof:

Note that the L.H.S. of (i) and (ii) can also be written as

$$\left[\sum_{g,h,i}^{n}\left[a_{bg}a_{ch}a_{di}\right]\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)g.}\tilde{\varepsilon}S_{(j)h.}\tilde{\varepsilon}S_{(j)i.}\tilde{\varepsilon}\right]\right\}\right]$$

and

$$\left[\sum_{f,g,h,i}^{n}\left[a_{bf}a_{cg}a_{dh}a_{ei}\right]\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)f.}\tilde{\varepsilon}S_{(j)g.}\tilde{\varepsilon}S_{(j)h.}\tilde{\varepsilon}S_{(j)i.}\tilde{\varepsilon}\right]\right\}\right],$$

respectively.

Upon applications of Lemmas 3.6 and 3.7 and replacing υ with $\tilde{\varepsilon}$, the remainder of the proof is then straightforward. *Q.E.D.*

Lemma 3.9 states that the third and fourth central moments of $\hat{\beta}$ are linear functions of the moments of ε. This result can easily be extended to higher moments of $\hat{\beta}$. On the other hand, Lemma 3.10 states that the observed third and fourth sample moments of the bootstrap estimates of β are linear functions of the sample moments of $\tilde{\varepsilon}$, as defined above. This result can also easily be extended to higher sample moments o´ $\beta^{*}_{(j)}$ $(j=1,\ldots,J)$ with the application of Lemma 3.8.

The above results show that a set of moments of ε uniquely determines the moments of $\hat{\beta}$. These results also show that a set of sample moments of $\tilde{\varepsilon}$ uniquely determines the sample moments of the bootstrap estimates of β. Thus, the results of Lemmas 3.2, 3.3, 3.9 and 3.10 suggest that; (i) the moments of ε uniquely determine the distribution of $\hat{\beta}$; and, (ii) the sample moments of $\tilde{\varepsilon}$ uniquely determine the sampling distribution of the bootstrap estimates of β.

With Lemmas 3.9 and 3.10, it is now possible to investigate how the use of uifferen.. residuals for bootstrapping affect the third and fourth sample moments of the bootstrap estimates of β. Only the OLS residuals, inflated OLS residuals and BLUS residuals will be considered here. To facilitate the derivations of the main theorems, the following three lemmas are needed.

Lemma 3.11: Let $\tilde{\varepsilon}_i$ $(i=1,\ldots,n)$ be replaced by the OLS residuals and let n be finite. Then,

(i) $$nE(\tilde{\mu}_3) = \left\{\sum_{i,j}^{n}\left[m_{ij}^3\right]\right\}\mu_3.$$

$$\text{(ii)}\quad nE(\tilde{\mu}_4) = \left\{\sum_{i,j}^{n}\left[m_{ij}^4\right]\right\}\left[\mu_4-3\mu_2^2\right] + \left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\}\left[3\mu_2^2\right].$$

Proof:

Note that the proof for part (i) is straightforward from least-squares theory. For part (ii), the proof is rather straightforward with the help of the following relations;

$$\text{(i)}\quad tr(M) = \left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\} + \left\{\sum_{\substack{i,j\\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\}.$$

$$\text{(ii)}\quad \left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\} = \left\{\sum_{i,j}^{n}\left[m_{ij}^4\right]\right\} + \left\{\sum_{\substack{h,i,j\\ j\neq h}}^{n}\left[m_{hi}^2-m_{ij}^2\right]\right\}. \qquad \textit{Q.E.D.}$$

Lemma 3.12: Let $\tilde{\varepsilon}_i=\left[n(n-K)^{-1}\right]^{1/2}\hat{\varepsilon}_i$ $\forall$ i, in which $\hat{\varepsilon}_i(i=1,\dots,n)$ are the OLS residuals. Also, let n be finite. Then,

$$\text{(i)}\quad nE(\acute{\mu}_3) = \left[n(n-K)^{-1}\right]^{3/2}\left\{\sum_{i,j}^{n}\left[m_{ij}^3\right]\right\}\mu_3.$$

$$\text{(ii)}\quad nE(\tilde{\mu}_4) = \left[n(n-K)^{-1}\right]^{2}\left[\left\{\sum_{i,j}^{n}\left[m_{ij}^4\right]\right\}\left[\mu_4-3\mu_2^2\right] + \left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\}\left[3\mu_2^2\right]\right].$$

Q.E.D.

Lemma 3.13: Let L=(n-K) and $\tilde{\varepsilon}=B\hat{\varepsilon}$ in which $\hat{\varepsilon}=(\hat{\varepsilon}_1,\hat{\varepsilon}_2,\dots,\hat{\varepsilon}_n)^T$ is the nx1 vector of OLS residuals and such that BM=B, $BB^T=I_L$ and $B^TB=M$. Note that $\tilde{\varepsilon}$ is the Lx1 vector of BLUS residuals. For n finite,

$$\text{(i)}\quad E(\tilde{\mu}_3) = \left\{\frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{n}\left[b_{ij}^3\right]\right\}\mu_3.$$

(ii) $E(\tilde{\mu}_4) = \left\{\frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{n}\left[b_{ij}^4\right]\right\}\left[\mu_4 - 3\mu_2^2\right] + 3\mu_2^2.$

Proof:

Note that the proof for part (i) is straightforward from the definition of $\tilde{\varepsilon}$. For part (ii), one needs to note that;

(i) $tr\left[BB^TBB^T\right] = L.$

(ii) $\left\{\frac{1}{L}\sum_{i=1}^{L}\sum_{j,k}^{n}\left[b_{ij}^2 b_{ik}^2\right]\right\} = 1.$

The remainder of the proof is then straightforward. *Q.E.D.*

Theorem 3.1: When the error distribution is symmetric, the third sample moment of the bootstrap estimates of β will be an unbiased estimate of the third moment of $\hat{\beta}$. This is true regardless of whether the OLS residuals, inflated OLS residuals or BLUS residuals are used for bootstrapping.

Proof:

Apply Lemmas 3.9, 3.10, 3.11, 3.12 and 3.13. *Q.E.D.*

An important corollary can be derived from Theorem 3.1 for the case when the error distribution is normal. This is Corollary 3.1 below.

Corollary 3.1: When the errors are normally distributed, bootstrapping based upon OLS residuals, inflated OLS residuals or BLUS residuals will lead to unbiased estimates of the third moment of $\hat{\beta}$.

Theorem 3.2: Let the error distribution be nonsymmetric. Then, bootstrapping based upon OLS residuals, inflated OLS residuals and BLUS residuals will lead to biased estimates of the third moment of $\hat{\beta}$. The directions of bias depending on the quantities

$$\text{(i)} \quad \left[n^{-1}\left\{\sum_{i,j}^{n}\left[m_{ij}^{3}\right]\right\} - 1\right],$$

$$\text{(ii)} \quad \left[n^{1/2}(n-K)^{-3/2}\left\{\sum_{i,j}^{n}\left[m_{ij}^{3}\right]\right\} - 1\right]$$

and

$$\text{(iii)} \quad \left[\left\{\frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{n}\left[b_{ij}^{3}\right]\right\} - 1\right], \text{ respectively.}$$

Proof:

Same as Theorem 3.1. *Q.E.D.*

In most applications, the three quantities in Theorem 3.2 can be computed directly. It must be mentioned that the first quantity will always be negative and that the second would most often be negative. Thus, bootstrapping based upon OLS or inflated OLS residuals when the disturbances come from a nonsymmetric population would lead to a downward bias in the estimation of the third moment of $\hat{\beta}$. The first result is straightforward. On the other hand, it is not clear when the second quantity of Theorem 3.2 will be negative. The following example shows for the case when the design matrix is random. Since the design matrix is assumed fixed for bootstrapping, the result of the following example would not be exact for the above bootstrap problem. Nevertheless, it does give a general idea about the underlying problem.

Example 3.1: Let

$$m_{ij} = \left[\delta_{ij}-(n^{-1}K)\right] \qquad (i,j=1,\dots,n)$$

in which δ_{ij} is the Kronecker delta. Then, it can easily be show that $\forall$ i $(i=1,\dots,n)$,

$$\left\{\sum_{j=1}^{n}(m_{ij}^3)\right\} = \left[n^{-2}(n^2+3K^2-3nK-K^3)\right].$$

When K=1, it follows that

$$n^{-1}\left\{\sum_{i,j}^{n}(m_{ij}^3)\right\} = \left[n^{-2}(n-1)(n-2)\right] < 1,$$

and

$$n^{-1}\left[n(n-1)^{-1}\right]^{3/2}\left\{\sum_{i,j}^{n}(m_{ij}^3)\right\} = \left[n(n-1)^{-1}\right]^{1/2}\left\{n^{-1}(n-2)\right\} < 1.$$

For finite n, it can be shown that the following equation

$$n^{-2}(n^2+3K^2-3nK-K^3) = 1$$

has three roots; one real and two imaginary roots, and the real root is K=0. Thus, for n finite and K>0,

$$n^{-1}\left\{\sum_{i,j}^{n}(m_{ij}^3)\right\} < 1.$$

On the other hand, no general result can be obtained for

$$n^{-1}\left[n(n-1)^{-1}\right]^{3/2}\left\{\sum_{i,j}^{n}(m_{ij}^3)\right\} = \left[n^{-2}(n-1)^{-3/2}(n^2+3K^2-3nK-K^3)\right].$$

For any known integer K, the right hand side will be less than 1 when $n>\eta(K)$, where $\eta(K)$ is the largest real root of the solution to

$$n^{-1/2}(n-1)^{-3/2}(n^2+3K^2-3nK-K^3) = 1.$$

For example, $\eta(1)=4/3$, $\eta(2)\approx 3$ and $\eta(3)\approx 5.2$. Note that the ratio $\{K/\eta(K)\}$ is quite large. When K=1, 2 and 3, this ratio is 3/4, 2/3 and 3/(5.2), respectively. Thus, when K is small relative to n, one can be sure that

$$n^{-1}\left[n(n-1)^{-1}\right]^{3/2}\left\{\sum_{i,j}^{n}(m_{ij}^{3})\right\} < 1.$$

When the design matrix is random, $E(m_{ij})=[\delta_{ij}-(n^{-1}K)]$. Thus, in this case, bootstrapping based upon OLS residuals will always lead to a downward biased estimation of the third moment of $\hat{\beta}$. When the design matrix is fixed, bootstrapping based upon OLS residuals will also lead to a biased estimation of the third moment of $\hat{\beta}$. However, its direction of bias depends on elements of the design matrix and the amount of percentage bias can be determined from the data. On the average, one would expect this bias to be a downward bias.

On the other hand, when the design matrix is random, no general conclusion can be made with respect to the third moment of $\hat{\beta}$, for the case when inflated OLS residuals are used for bootstrapping. The direction of bias in the third sample moment of bootstrap estimates of β depends on both n and K. However, when K is relatively small as compared to n, this bias will also be a downward bias. The implication for the case when the design matrix is fixed is that, on the average, one would also expect a downward bias in the third sample moment of bootstrap estimates of β.

Attention is now focussed on the fourth moment of $\hat{\beta}$. Lemmas 3.14 and 3.15 below investigate the fourth sample moments of the OLS and inflated OLS residuals, respectively. The results are then summarized in Theorems 3.3 and 3.4.

Lemma 3.14: Let ε be a normal variate, such that $\mu_4=3\mu_2^2$. Then, for any $n>K$, the fourth sample moment of the OLS residuals will underestimate μ_4. The degree of underestimation will depend only on M.

Proof:

Let $\tilde{\varepsilon}=\hat{\varepsilon}$. Then, upon application of Lemma 3.11 part (ii),

$$nE(\tilde{\mu}_4) = \left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\}\left[3\mu_2^2\right].$$

However,

$$\left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\} = (n-K) - \left\{\sum_{\substack{i,j \\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\}.$$

The proof becomes rather straightforward since

$$K + \left\{\sum_{\substack{i,j \\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\} > 0.$$

Q.E.D.

Lemma 3.15: Let the error disturbances be normal. Then, for a given $n>K$, the fourth sample moment of the inflated OLS residuals will underestimate μ_4, provided that

$$n(n-K)^{-1}\left\{\sum_{\substack{i,j \\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\} < K.$$

Proof:

Let $\tilde{\varepsilon}=\left[n(n-K)^{-1}\right]^{1/2}\hat{\varepsilon}$. Then, upon application of Lemma 3.12 part (ii),

$$nE(\tilde{\mu}_4) = \left[n(n-K)^{-1}\right]^2\left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\}\mu_4.$$

Since

$$\left\{\sum_{i=1}^{n}\left[m_{ii}^2\right]\right\} = (n-K) - \left\{\sum_{i,j}^{n}\left[m_{ij}^2\right]\right\},$$

the expectation of $\tilde{\mu}_4$ can be written as

$$E(\tilde{\mu}_4) = n(n-K)^{-2}\left[(n-K)-\left\{\sum_{\substack{i,j\\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\}\right]\mu_4.$$

The remainder of the proof is then straightforward. *Q.E.D.*

Example 3.2: Let the notation be that of Lemma 3.15. Also, let $M=\left\{I_n-n^{-1}\left[E(n,n)\right]\right\}$. Then,

$$\left\{\sum_{\substack{i,j\\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\} = n^{-1}(n-1).$$

Upon substitution of this relation into $E(\mu\ )$ of Lemma 3.15, one obtains

$$E(\tilde{\mu}_4) = n(n-K)^{-2}\left[(n-K) - n^{-1}(n-1)\right]\mu_4.$$

When K=1, it can be shown that $E(\tilde{\mu}_4)=\mu_4$. In cases when K>1, $E(\tilde{\mu}_4)\gtreqless\mu_4$ when $n\gtreqless(n-1)^{-1}(K^2-1)$. Since it can be shown that $\left[(K-1)^{-1}(K^2-1)-1\right]=K$, it follows that $E(\tilde{\mu}_4)>\mu_4$. The reason being that n is normally required to be greater than K in regression models.

Upon application of Lemma 3.15, one can show through Example 3.2 that when the design matrix is random, bootstrapping based upon inflated OLS residuals will lead to biased estimates of the fourth moment of ε.

This bias will always be upward, irrespective of the type of error distribution. The result of Example 3.2 implies that, on the average, bootstrapping based upon inflated OLS residuals will lead to an upward bias in the estimates of the fourth moment of ε, in the case when the design matrix is fixed.

Lemma 3.15 states that for a given design matrix, the fourth sample moment of inflated OLS residuals will underestimate (or overestimate) the fourth moment of ε, depending on K and the off-diagonal elements of the projection matrix, M. This is true regardless of the type of error distribution.

Even in the case of normal errors, both OLS and inflated OLS residuals will perform poorly as estimates of ε. However, inflated OLS residuals are slightly better than OLS residuals. One reason is that the former's second sample moment is an unbiased estimate of the second moment of ε, whereas the latter's second sample moment is biased downward (c.f., Chapter 2). Secondly, the former's third and fourth sample moments are closer estimates of the corresponding exact moments of ε, when compared to the latter's third and fourth sample moments, respectively.

Lemmas 3.14 and 3.15 are useful for proving Theorem 3.3 below. No separate lemma is required for the BLUS residuals since it is straightforward from Lemma 3.13, that the fourth sample moment of the BLUS residuals is an unbiased estimate of μ_4, provided ε is a normal variate.

Theorem 3.3 below investigates how well the bootstrap method estimates the fourth moment of $\hat{\beta}$ for the three cases when OLS, inflated OLS and BLUS residuals are used for bootstrapping. It assumes that the disturbance terms are normally distributed.

Theorem 3.3: Let ε be a normal variate and let $\tilde{\varepsilon}$ be used for bootstrapping. Also, let n be finite. Then,

(1) When $\tilde{\varepsilon}=\hat{\varepsilon}$, bootstrapping will lead to an underestimation of the fourth moment of $\hat{\beta}$. For any given n>K, this underestimation increases as K increases.

(2) When $\tilde{\varepsilon}=\left[n(n-K)^{-1}\right]^{1/2}\hat{\varepsilon}$, bootstrapping will, on the average, lead to biased estimates of the fourth moments of $\hat{\beta}$. For given n, K and M, the direction of bias depends on the sign of the following quantity,

$$\left[K-n(n-K)^{-1}\left\{\sum_{\substack{i,j \\ j\neq i}}^{n}\left(m_{ij}^{2}\right)\right\}\right].$$

(3) When $\tilde{\varepsilon}=B\hat{\varepsilon}$, bootstrapping will always lead to unbiased estimates of the fourth moment of $\hat{\beta}$.

Proof:

Upon application of Lemma 3.10, it can be observed that fourth sample moments of bootstrap estimates of β depend only on the second and fourth sample moments of $\tilde{\varepsilon}$. It can easily be shown that the second sample moment of OLS residuals underestimates μ_2, while the second sample moments of both inflated OLS and BLUS residuals are unbiased estimates of μ_2.

The remainder of the proof is then straightforward, upon application of Lemmas 3.13, 3.14 and 3.15. *Q.E.D.*

Theorem 3.3 illustrates only for the case when the random disturbances are normal. It states that when these disturbances are normal, bootstrapping will lead to a downward bias in the estimation of the fourth moment of $\hat{\beta}$, provided that OLS residuals are used for bootstrapping. Bootstrapping based upon inflated OLS residuals will also lead to biased estimates of the fourth moment of $\hat{\beta}$. However, the bias can be in either direction, depending on the sample size, the column rank and elements of X. On the other hand, bootstrapping based upon BLUS residuals will always lead to unbiased estimates of the fourth moment of $\hat{\beta}$.

When the disturbances are non-normal, the results will be different. The following lemma and theorem examine the case when ε is either platykurtic or leptokurtic. Note that ε is platykurtic when $\mu_4<3\mu_2^2$ and is leptokurtic when $\mu_4>3\mu_2^2$. For the following discussion, let $\mu_4=\tau\mu_2^2$, in which $(\tau-3)$ is a measure of kurtosis.

Lemma 3.16: Let $\phi=\left[E(\tilde{\mu}_4)-\mu_4\right]\mu_4^{-1}$. Then,

(i) When $\tilde{\varepsilon}=\hat{\varepsilon}$, $\phi\gtreqless 0$ provided that $\tau\gtreqless\hat{\tau}$ and

$$\hat{\tau} = \frac{3\left\{\sum_{h,i,j}^{n}\left[m_{hi}^2 m_{hj}^2\right]\right\}}{K+\left\{\sum_{\substack{i,j\\ j\neq i}}^{n}\left[m_{ij}^2\right]\right\}+\left\{\sum_{\substack{h,i,j\\ j\neq i}}^{n}\left[m_{hi}^2 m_{hj}^2\right]\right\}}.$$

(ii) When $\tilde{\varepsilon}=\left[n(n-1)^{-1}\right]^{1/2}\hat{\varepsilon}$, $\phi \gtreqless 0$ provided that $\tau \gtreqless \hat{\hat{\tau}}$ and

$$\hat{\hat{\tau}} = \frac{3\left\{\sum_{h,i,j}^{n}\left[m_{hi}^{2}m_{hj}^{2}\right]\right\}}{\left\{\sum_{\substack{i,j \\ j\neq i}}^{n}\left[m_{ri}^{2}\right]\right\}+\left\{\sum_{\substack{h,i,j \\ j\neq i}}^{n}\left[m_{hi}^{2}m_{hj}^{2}\right]\right\}n^{-1}(n-K)K}.$$

(iii) When $\tilde{\varepsilon}=B\hat{\varepsilon}$, $\phi \gtreqless 0$ provided that $\tau \gtreqless 3$.

Also, for all three cases above,

$$\partial\phi \;{:}\tau>0.$$

Proof:

With the substitution of $\mu_4=\tau\mu_2^2$, the proof is rather straightforward upon application of Lemmas 3.11, 3.12 and 3.13. *Q.E.D.*

Theorem 3.4: Let the notation be that of Lemma 3.16. For finite n;

(i) Bootstrapping based upon OLS residuals will lead to biased estimates of the fourth moment of $\hat{\beta}$. The direction of bias will be downward when $\tau<\hat{\tau}$, but it will be uncertain when $\tau\geq\hat{\tau}$.

(ii) Bootstrapping based upon inflated OLS residuals may lead to unbiased estimates of the fourth moment of $\hat{\beta}$. This happens when $\tau=\hat{\hat{\tau}}$. Otherwise, these estimates will be biased downward (upward) when τ is less (greater) than $\hat{\hat{\tau}}$.

(iii) Bootstrapping based upon BLUS residuals will lead to unbiased estimates of the fourth moment of $\hat{\beta}$, provided that the error distribution is normal. When this distribution is platykurtic (leptokurtic), these estimates will be biased downward (upward).

Proof:

The proof is straightforward upon application of Lemmas 3.9 and 3.16. *Q.E.D.*

The above results have serious implications for the bootstrap confidence intervals, which will be examined in Chapter 4. It is usual in the applications of linear regression models to assume that the error distribution is normal. When this assumption is valid, bootstrapping based upon BLUS residuals will yield unbiased estimates for the first four moments of $\hat{\beta}$. On the other hand, the use of OLS residuals for bootstrapping should be avoided, especially when the sample size is small. However, it must be mentioned that BLUS residuals are quite cumbersome to obtain. In most applications, one may use inflated OLS residuals for bootstrapping. On the other hand, the resulting bootstrap estimates should be used with caution, especially when the number of coefficients is large as compared to the sample size.

To complete the investigation of the higher moments of $\hat{\beta}$, a Monte Carlo simulation study is conducted below to examine the first ten sample moments of several regression residuals. Only the case when the error distribution is normal will be considered. Consequently, only the even moments will be reported. This is because when the error distribution is normal, all odd moments will be unbiased, irrespective of the type of regression residuals.

3.4 A Monte Carlo Simulation Study of the Sample Moments of Some Selected Residuals

It has been illustrated in the preceding section that one can examine the sample moments of the underlying residuals for variations (like bias and MSE) in the sample moments of bootstrap estimates of β. This reduces the amount of computation considerably. The reason is that bootstrap estimates of β requires additional computations which have to be repeated J times, J can be as large as 10,000. On the other hand, the residuals need only be computed once.

In Chapter 2, it is shown that bootstrapping based upon either inflated OLS or BLUS residuals (under some general conditions) leads to unbiased estimate of $D(\hat{\beta})$. When ε is a normal variate, it is also shown in the preceding section, that bootstrapping based upon either inflated OLS or BLUS residuals leads to unbiased estimate of the third moment of $\hat{\beta}$. However, when ε is not normally distributed, both procedures will yield biased estimates. Bootstrapping based upon BLUS residuals will also yield unbiased estimates of the fourth moment of $\hat{\beta}$, provided that ε is normal.

It must be mentioned that in most regression models, BLUS residuals may not have zero sample means. Since one of the requirments for bootstrapping in regression models is that the underlying residuals have zero sample mean, these BLUS residuals must first be centered, by substraction of the sample mea·. The resulting BLUS residuals will be known as centered BLUS residuals. For reporting the simulation results, the original BLUS residuals will be represented by BLUS1 while BLUS2 represents the centered BLUS residuals. Likewise, OLS residuals will be represented by OLS1 while OLS2 represents inflated OLS residuals.

The use of centered BLUS residuals for bootstrapping will lead to downward biases in all estimates of moments of $\hat{\beta}$, except the first moment. When the error distribution is normal, all estimates of the odd moments of $\hat{\beta}$ will still be unbiased. The expectations of these estimates will all be zero. Let $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ be the OLS and centered BLUS residuals, respectively. When centered BLUS residuals are multiplied by the factors $\left[n(n-1)^{-1}\right]^{1/2}$ and $\left[n(n-K)^{-1}(\hat{\varepsilon}^T\hat{\varepsilon})(\tilde{\varepsilon}^T\tilde{\varepsilon})^{-1}\right]^{1/2}$, the transformed residuals will be denoted BLUS3 and BLUS4, respectively. It can easily be shown that the second sample moment of BLUS4 will be exactly the second sample moment of either BLUS1 or OLS2. However, the same cannot be said for BLUS3. Second sample moments of both BLUS3 and BLUS4 are unbiased estimates of the second moment of ε. On the other hand, fourth sample moments of both BLUS3 and BLUS4 will be biased, even for the normal case. Looking at the brighter side, one of these biases may be less than when inflated OLS residuals are used.

The objective of this section is to examine sample moments of the various residuals beyond the fourth. It must be recognized that Monte Carlo pseudo-random samples generated by computer subroutines are only approximate random sample, drawn from the true population. The accuracy of each sample is difficult to assess. [See e.g., Bronshtein and Semendyayev (1985, p.938).]

In addition to the above residuals, the transformation suggested by Stine (1985) will also be applied to the OLS residuals. The transformed residuals are represented by STINE. Stine suggests dividing each OLS residuals by the square-root of the corresponding main diagonal elements of the projection matrix, M. That is, when $\hat{\varepsilon}_i$ is the i'th observation of the OLS residuals and m_{ii} is the i'th element along the main diagonal of M, the i'th observation of the rescaled residuals is $r_i=(m_{ii})^{-1/2}\hat{\varepsilon}_i$.

The simulation results are all based upon the simple LR model of Section 2.2 above. When K=2, the intercept and slope of the regression line are assigned the values 1.0 and 2.0, respectively. Both the disturbances and values for the independent variable are generated by IMSL's GGNPM subroutine. The disturbances are assumed to come from a normal population having zero mean and unit variance. The observations of the independent variable are assumed to come from a normal population with mean zero and variance 4.0. For each n, these observations are drawn once and then fixed throughout the experiment.

Two separate experiments are conducted (for n=10,20), to determine the effects of sample size on the results. For each experiment, 500 trials are conducted. In each trial, the first 10 sample moments of the pseudo-normal errors and the regression residuals are computed. These sample moments are then compared with the theoretical moments of a N(0,1) distribution. Sample moments of the regression residuals are also compared with sample moments of the pseudo-normal errors. Both the bias and MSE are used as yardsticks for comparison among the residuals.

The simulation results are reported only for the case when the error distribution is normal. In this case, all odd moments will be zero. Consequently, it would not be of interest to report the results for odd sample moments. Hence, only even sample moments are reported. These results are tabulated in Tables 3.1 and 3.2.

From Tables 3.1 and 3.2, it can be observed that for all cases and for all even moments, the absolute bias of STINE is the largest. This is followed closely by OLS1. Thus, it is again demonstrated that for small samples ($n \le 20$), both OLS residuals and residuals obtained via Stine's transformations are unsuitable for bootstrapping.

Of the BLUS residuals, BLUS2 has the largest absolute bias for all cases and for all even moments. These biases are also larger than those of OLS2. In practical applications to real world data, the sample mean of BLUS1 is rarely zero. When the sample mean is non-zero, bootstrapping based upon BLUS1 and BLUS2 will yield the same results, provided that sample moments of bootstrap estimates of β are computed about the sample mean of these estimates. Consequently, both BLUS1 and BLUS2 should also be avoided when the sample is small.

The residuals that remain to be compared are OLS2, BLUS3 and BLUS4. When $n=10$, the results of both tables indicate that OLS2 and BLUS3 are very similar. However, this may be true only for this example. When $n=20$, the two sets of results become slightly different. However, it is not absolutely clear which of the two residuals is better suited for bootstrapping. Results in Table 3.1 indicate that OLS2 is better than BLUS3 while the reverse is implied by the results of Table 3.2.

Between OLS2 and BLUS4, the results in Table 3.1 clearly indicate that OLS2 is better suited for bootstrapping. On the other hand, Table 3.2 indicates the contrary. Thus, there is uncertainty over whether

Table 3.1: Bias and MSE of Sample Moments of Residuals when Compared to the Error Population Moments (K=2)[1]

a) n=10

Sample Moments	OLS1	OLS2	STINE	BLUS1	BLUS2	BLUS3	BLUS4
$\tilde{\mu}_2$	-0.25	-0.06	-0.39	-0.06	-0.17	-0.06	-0.06
	(0.42)	(0.43)	(0.48)	(0.43)	(0.44)	(0.46)	(0.43)
$\tilde{\mu}_4$	-1.40	-0.50	-1.90	-1.52	-1.08	-0.49	-0.55
	(1.97)	(2.22)	(2.13)	(2.29)	(2.10)	(2.41)	(2.24)
$\tilde{\mu}_6$	-9.86	-4.96	-11.98	-5.18	-8.27	-4.96	-5.48
	(12.17)	(14.80)	(12.75)	(15.83)	(13.25)	(16.22)	(14.80)
$\tilde{\mu}_8$	-84.10	-54.05	-94.50	-55.44	-75.07	-53.93	-58.26
	(95.1)	(121.2)	(97.4)	(130.9)	(102.9)	(131.7)	(115.5)
$\tilde{\mu}_{10}$	-845.5	-641.4	-901.8	-648.6	-787.8	-638.6	-676.5
	(902)	(1156)	(914)	(1219)	(952)	(1222)	(1035)

b) n=20

Sample Moments	OLS1	OLS2	STINE	BLUS1	BLUS2	BLUS3	BLUS4
$\tilde{\mu}_2$	-0.13	-0.04	-0.21	-0.04	-0.09	-0.03	-0.04
	(0.32)	(0.32)	(0.34)	(0.32)	(0.32)	(0.33)	(0.32)
$\tilde{\mu}_4$	-0.85	-0.35	-1.19	-0.36	-0.66	-0.42	-0.40
	(1.61)	(1.72)	(1.66)	(1.77)	(1.67)	(1.76)	(1.72)
$\tilde{\mu}_6$	-6.98	-4.00	-8.71	-3.91	-5.94	-4.52	-4.41
	(10.54)	(11.56)	(10.76)	(12.67)	(11.14)	(11.96)	(11.72)
$\tilde{\mu}_8$	-67.34	-47.59	-77.40	-45.11	-60.64	-51.60	-50.36
	(85.36)	(93.06)	(86.89)	(108.2)	(90.26)	(97.39)	(95.35)
$\tilde{\mu}_{10}$	-740.4	-598.4	-804.2	-561.2	-692.2	-630.6	-616.6
	(833)	(880)	(849)	(1044)	(868)	(929)	(901)

1. Root MSE's are reported in parenthesis.

Table 3.2: Bias and MSE of Sample Moments of Residuals when Compared to the Error Sample Moments (K=2)[1]

a) n=10

Sample Moments	OLS1	OLS2	STINE	BLUS1	BLUS2	BLUS3	BLUS4
$\tilde{\mu}_2$	-0.20 (0.28)	-0.01 (0.22)	-0.34 (0.40)	-0.01 (0.22)	-0.12 (0.28)	-0.01 (0.27)	-0.01 (0.22)
$\tilde{\mu}_4$	-0.78 (1.32)	0.13 (1.32)	-1.28 (1.74)	0.10 (1.52)	-0.46 (1.43)	0.13 (1.66)	0.07 (1.41)
$\tilde{\mu}_6$	-3.44 (7.95)	1.47 (10.67)	-5.56 (9.27)	1.24 (12.76)	-1.85 (9.51)	1.46 (12.95)	0.94 (11.09)
$\tilde{\mu}_8$	-16.76 (53.0)	13.30 (94.7)	-27.13 (55.3)	11.91 (110.9)	-7.72 (70.1)	13.41 (109.5)	9.09 (80.1)
$\tilde{\mu}_{10}$	-87.2 (377)	117.0 (899)	-143.5 (350)	109.7 (1001)	-29.5 (546)	119.7 (1006)	81.9 (746)

b) n=20

Sample Moments	OLS1	OLS2	STINE	BLUS1	BLUS2	BLUS3	BLUS4
$\tilde{\mu}_2$	-0.09 (0.13)	0.01 (0.10)	-0.17 (0.20)	0.01 (0.10)	-0.05 (0.13)	0.01 (0.13)	0.01 (0.10)
$\tilde{\mu}_4$	-0.41 (0.72)	-0.09 (0.67)	-0.75 (0.99)	-0.08 (0.79)	-0.22 (0.79)	0.07 (0.83)	0.04 (0.75)
$\tilde{\mu}_6$	-2.15 (5.00)	0.83 (5.56)	-3.88 (6.19)	-0.90 (7.41)	-1.11 (6.09)	0.59 (6.91)	0.42 (6.54)
$\tilde{\mu}_8$	-12.56 (38.5)	7.18 (49.3)	-22.60 (43.7)	9.66 (72.1)	-5.87 (50.5)	5.53 (61.1)	4.42 (57.8)
$\tilde{\mu}_{10}$	78.4 (311)	63.6 (453)	-142.2 (329)	100.8 (714)	-30.2 (437)	53.5 (565)	45.4 (522)

1. Root MSE's are reported in parenthesis.

OLS2 or BLUS4 is better suited for bootstrapping. One possible reason for this uncertainty is that K=2, which is small. Consequently, another experiment is conducted with K=5, representing the case when K is moderate to large. It must be noted that the case with n=10 and k=5 may not be very useful in actual applications of the bootstrap. However, the purpose of this experiment is to highlight the difference in using OLS2 and BLUS4 for bootstrapping, and to determine which of the two is better suited for bootstrapping when $K=\frac{1}{2}n$ and n is any positive even integer. The design of this experiment is the same as for the previous experiment, except that only OLS2 and BLUS4 are compared and only the case when n=10 is considered. The results are reported in Table 3.3.

It can be observed from Table 3.3 that BLUS4 is more suitable for

Table 3.3: Bias and MSE of Sample Moments of OLS2 and BLUS4[1]

a) When Compared to the Error Population Moments

	$\tilde{\mu}_4$	$\tilde{\mu}_6$	$\tilde{\mu}_8$	$\tilde{\mu}_{10}$
OLS2	0.54	6.05	62.08	638.1
	(4.40)	(43.2)	(510)	(6669)
BLUS4	-0.19	-3.15	-41.00	-539.6
	(3.42)	(23.6)	(187)	(1619)

b) When Compared to the Error Sample Moments

	$\tilde{\mu}_4$	$\tilde{\mu}_6$	$\tilde{\mu}_8$	$\tilde{\mu}_{10}$
OLS2	0.75	8.58	89.87	954.0
	(3.02)	(31.3)	(380)	(5057)
BLUS4	0.10	0.89	9.95	106.3
	(2.41)	(18.2)	(154)	(1361)

1. Root MSE's are reported in parenthesis. [n=10, K=5].

bootstrapping relative to OLS2 when K is moderate large. This conclusion is restricted to normal errors and further work is needed before this conclusion can be generalized to a broad class of error distributions. This is left for future research.

The emphasis of the above analysis has been on bias. The reason for this emphasis is that the bias of sample moments of bootstrap estimates of β affects empirical significance levels and the bounds of bootstrap confidence intervals. In considering confidence intervals, an appropriate criterion for judging their reliability is the MSE. Judging from the results of Table 3.3, of OLS2 and BLUS4, BLUS4 is still more suitable for bootstrapping because the MSE's of sample moments of BLUS4 are considerably smaller than those of OLS2, regardless of whether these sample moments are compared to population or sample moments.

When K=2, the results of both Tables 3.1 and 3.2 suggest that OLS2 and BLUS4 are very similar in terms of MSE. Thus, in most practical applications, it may be more economical to use OLS2 for bootstrapping when K is small relative to n. It is convenient too, since BLUS4 is quite cumbersome to obtain; moreover, the gain in efficiency from using BLUS4 may only be negligible.

3.5 Summary

The focus of this chapter has been on higher moments of bootstrap estimates of β. Applying least-squares theory, it can easily be shown that moments of $\hat{\beta}$ are linear functionals of the underlying error moments. Using techniques developed in this chapter, sample moments of bootstrap estimates of β are shown to be linear functionals of sample moments of the underlying residuals. The two procedures above lead to quite similar results.

The most important lemma in this chapter is Lemma 3.10. It forms the basis for all the main theorems. For simplicity and convenience, this lemma is applied to the third and fourth sample moments only. However, an extension to higher sample moments can be obtained by applying Lemma 3.8. The first and second moments of the bootstrap estimates of β have been investigated in Chapter 2. For most purposes, moments higher than the fourth order have little practical value. In the literature, several methods have been suggested for obtaining bootstrap estimates of β. Associated with these methods is an array of transformations applied to OLS residual, prior to bootstrapping. Of these, OLS, inflated OLS, BLUS and Stine residuals have been studied in this chapter.

When the error distribution is symmetric, both analytical and simulation results show that the third sample moments of the bootstrap estimates of β will be unbiased estimates of the third moment of $\hat{\beta}$. This statement remains valid regardless of the type of residuals used for bootstrapping. By the same token, it can easily be shown that all odd sample moments of these bootstrap estimates will be zero, provided that sample means of the underlying residuals are zero. Consequently, the odd moments have been omitted in Tables 3.1. 3.2 and 3.3.

When the error distribution is nonsymmetric, it is shown analytically that all third sample moments of the bootstrap estimates of β will be biased The direction and proportion of this bias can be obtained directly from the data. No simulation results are undertaken for this case.

The fourth sample moments of bootstrap estimates of β depend on both second and fourth sample moments of the residuals. Let the second sample moment be an unbiased estimate of μ_2, μ_2 being the second moment of underlying population disturbances. Then, fourth sample moments of bootstrap estimates of β will be unbiased, provided the fourth sample moments of the residuals are unbiased estimates of μ_4. This property can only be found in BLUS residuals when the error distribution is normal. When this distribution is not normal, all the fourth sample moments will be biased.

Although both second and fourth sample moments of BLUS residuals are unbiased estimates of μ_2 and μ_4, respectively, the sample means of these residuals are rarely zero. Since one of the requirements for bootstrapping in regression models is that the underlying residuals have zero sample mean, BLUS residuals have to be centered. This transformation leads to further complications. Consequently, BLUS residuals should be avoided whenever K is small (relative to n). When K is large, relative to n, a slight variant of the BLUS residuals may be used for bootstrapping. This is suggested by the results of Table 3.3 when the error distribution is normal.

Results of both Tables 3.1 and 3.2 above suggest that, when K is small (relative to n), inflated OLS residuals and the better variant of BLUS residuals are very similar in terms of both bias and MSE. In most applications, BLUS residuals are quite cumbersome to obtain, whereas, inflated OLS residuals can easily be computed. Thus, it may be more economical to use inflated OLS residuals for bootstrapping. In many cases, the gain in efficiency from using BLUS residuals for bootstrapping may be very small.

Both OLS residuals and the residuals obtained by Stine's procedure are once again, found to be unsuitable for bootstrapping. Stine's procedure is based upon an idea borrowed from Hinkley's (1977) weighted-jackknife methodology. Although suitable for the weighted-jackknife, this transformation does not do well i bootstrapping. On the other hand, OLS residuals can still be used for bootstrapping, provided that both n is large and K small.

The above results have important implications for the bootstrap confidence intervals, which is the focus of Chapter 4 below. The above simulation results are restricted to the case when the error distribution is normal. Further work needs to be done before one can generalize these results to a broad class of error distributions. It will be of interest to incorporate this type of work in future research.

CHAPTER FOUR

BOOTSTRAP CONFIDENCE INTERVAL OF β

4.1 Introduction

The sample moments of the bootstrap estimates of β have been examined in Chapters 2 and 3. As a sequel to these earlier two chapters, the focus of this chapter is on the bootstrap confidence intervals (BCI's) of β. Three different ways of constructing BCI's will be investigated. These are based upon OLS, inflated OLS and BLUS residuals.

Wu (1986) proposed several methods for obtaining BCI's of β and found that empirical coverages of these BCI's are generally lower than nominal, which is disappointing in view of the second-order asymptotics on the bootstrap. Singh (1986) suggests that second-order asymptotics may not be sufficient and third-order asymptotics are necessary for comparing BCI's. However, these third-order asymptotics are rather rare, at least for the present moment.

Thus, it would be of interest to incorporate some third-order asymptotics on the bootstrap in the present context. Also, it would be interesting to compare the properties of different BCI's when OLS, inflated OLS and BLUS residuals are used for bootstrapping. This makes use of the results obtained earlier in Chapters 2 and 3. The present approach is also different from the approaches of existing works in that both Edgeworth expansions and sample moments of bootstrap estimates of β are used to compare bootstrap confidence intervals. None of the existing literature uses the sample moments of bootstrap estimates of β

(or any other statistic).

One method of constructing BCI's is to sort the bootstrap estimates in ascending order and then record the appropriate percentiles of the sorted bootstrap sample. Let a BCI of this type be known as a naive BCI. For a naive BCI to qualify as an exact confidence interval of β, all sample moments of the bootstrap estimates of β must be exactly the same as the corresponding moments of $\hat{\beta}$. This, of course, requires the assumption that the probability density function of $\hat{\beta}$ be completely determined by its moments.

In Chapter 3, it is shown that for a fixed design matrix X, the moments of $\hat{\beta}$ are uniquely determined by a set of moments of the underlying disturbances, ε. Also, the sample moments of the bootstrap estimates of β are uniquely determined by a set of moments of the underlying residuals, $\tilde{\varepsilon}$. These residuals can be the OLS residuals, inflated OLS residuals or BLUS residuals. Thus, when the sample moments of $\tilde{\varepsilon}$ correspond exactly with the moments of ε, the naive BCI qualifies as an exact confidence interval of β.

The bootstrap method is normally used for small to moderate samples. Moreover, for most economic data, a large sample size is rather rare. Also, moments of ε higher than the first order are generally unknown, especially in empirical applications to real world data. Although unbiased estimates of these moments of ε can easily be obtained from sample moments of the OLS residuals, it is difficult to obtain an $\tilde{\varepsilon}$ whose sample moments correspond exactly with the moments of ε. Hence, in practice, the naive BCI seldom qualifies as an exact confidence interval of β.

For a simple class of parametric problems with multinormal variates, Efron's (1985, p. 53) Theorem 2 shows that the BCI is accurate to $O_p(n^{-1})$. This is subject to the assumption that samples of these random variables can be obtained for bootstrapping. When ε is a normal variate and a sample of ε can be obtained for bootstrapping, it is shown below that regression analysis belongs to this simple class of parametric problems.

Unfortunately, in regression analysis, ε is generally not observable and has to be replaced by $\tilde{\varepsilon}$, $\tilde{\varepsilon}$ being a suitable estimate of ε. When all moments of $\tilde{\varepsilon}$ correspond exactly with the moments of ε, a naive BCI based upon $\tilde{\varepsilon}$ will qualify as an exact confidence interval of β. When the exact moments of ε are unknown and when sample moments of ε can be computed, a naive BCI based upon a sample of ε of size n is accurate to $O_p(n^{-1})$. Further, when sample moments of $\tilde{\varepsilon}$ are unbiased estimates of the means of sample moments of ε, a naive BCI based upon a sample of $\tilde{\varepsilon}$ of size n will also be accurate to $O_p(n^{-1})$.

When the above conditions are not fulfilled, a naive BCI based upon $\tilde{\varepsilon}$ will no longer be accurate to $O_p(n^{-1})$. In such case, alternative methods are needed for constructing better BCI's of $\hat{\beta}$. One such method is proposed in this chapter. The proposed method is based upon the theoretical results of Babu and Singh, [hereinafter, B&S]. Other related papers include Singh (1981) and Beran (1982).

It is shown in Chapter 3, that sample moments of BLUS residuals and inflated OLS residuals do not always coincide exactly with the sample moments of ε. It is also shown that not all of these sample moments are unbiased estimates of the means of sample moments of ε. Thus, a naive

BCI of $\hat{\beta}$ based upon either BLUS or inflated OLS residuals is unlikely to be accurate to $O_p(n^{-1})$. Hence, there is a good reason to adopt the proposed method.

The organization of this chapter is as follows. In Section 4.2, the use of Edgeworth expansions is discussed and notation is established along with some preliminaries. In Section 4.3, the focus is on the case when a sample of ε can be obtained for bootstrapping. This serves both as a bench-mark for further discussions and as a focal point from which further insights can be gained. The case when a sample of ε is replaced with $\tilde{\varepsilon}$ is discussed in Section 4.4. Only the cases when the true disturbances are replaced with the OLS, inflated OLS and BLUS residuals are investigated. In Section 4.5, the robustness of BCI's against alternative distributions is examined. The definition of robustness is that of Hampel (1971). Lastly, a summary of the main results is presented in Section 4.6.

4.2 Preliminaries and Notation

The following theorem establishes the notion of robustness, which is introduced by Hampel (1971), for a general test statistic. However, the notation is mainly that of Huber (1981). No originality is claimed for this theorem, which somewhat brings together the various concepts discussed in Huber (1981, pp. 27-42). This theorem is useful for studying the robustness of bootstrap confidence intervals.

Theorem 4.1 (Hampel): Let $x_1,\ldots,x_n$ be i.i.d. with common distribution F, and let $T_n=T_n(x_1,\ldots,x_n)$ be a sequence of estimates or test statistics with values in $\mathbb{R}^k$. This sequence is robust at $F=F_0$ when the map

$$F \longrightarrow L_F(T_n)$$

is equicontinuous at F_0. L is a linear functional on $\mathbb{M}$, the space of all probability measures on $(\Omega,\mathbb{B})$, where $\mathbb{B}$ is the Borel σ-algebra and Ω is a topological space whose topology is metrizable by some metrix d, such that Ω is complete and separable. That is, $\forall\ \varepsilon>0$, $\exists\ \delta>0$ and an n_0 such that, $\forall\ F$ and $\forall\ n\geq n_0$,

$$d_L(F_0,F) \leq \delta \Rightarrow d_P\left[L_{F_0}(T_n),L_F(T_n)\right] \leq \varepsilon.$$

For any two distribution functions F and G, $d_L(F,G)$ and $d_P\left[L_F(T_n),L_G(T_n)\right]$ denote the Levy and Parohorov distances between F and G, respectively.

Consequently; (1) when T is weakly continuous at F, $\{T_n\}$ is consistent at F, in the sense that $T_n \longrightarrow T(F)$ *i.p.* and *a.s.*; (2) when $\{T_n\}$ is consistent in a neighborhood of F_0, T is continuous at F_0 iff $\{T_n\}$ is robust at F_0.

Proof:

Part (1) follows from the Glivenko-Cantelli theorem and the Kolmogrov distance that, *i.p.* and *a.s.*,

$$d_L(F,F_n) \leq d_K(F,F_n) \longrightarrow 0$$

in which $d_K(F,F_n)=\sup|F(x)-F_n(x)|$.

For part (2), when T is continuous at F_0,

$$d_P\left[L_{F_0}(T_n),L_F(T_n)\right] \leq d_P\left[\delta_{T(F_0)},L_{F_0}(T_n)\right] + d_P\left[\delta_{T(F_0)},L_F(T_n)\right]$$

in which $\delta_{T(F_0)}$ is the degenerate law concentrated at $T(F_0)$. The remaining of the the proof is then rather straightforward upon application of Lemma 3.6, Theorems 3.7 and 3.8 of Huber (1981, pp. 27-29). (C.f. Huber, 1981 pp. 40-42). Q.E.D.

One other useful tool for studying confidence intervals is the traditional Edgeworth expansion. Let $x_1, x_2, \ldots, x_n$ be an i.i.d. sequence of random variables with absolute continous cummulative density function $F(x)$ and let $G_n(x)$ be the cumulative distribution function of the normalized sum $(x_1+\ldots+x_n-n\mu)/(n^{1/2}\sigma)$, in which μ and σ^2 are the mean and variance of F. Also, let $\Phi(x)$ and $\phi(x)$ be the cumulative distribution and density functions of the standa.:d normal distribution, respectively. In terms of central moments of ·, the Edgeworth expansion for $G_n(x)$ can be written as (see eg., Patel and Read p. 157)

$$G_n(x) = \Phi(x)-\phi(x)\left[\frac{\mu_3(x^2-1)}{\sigma^3 3!\sqrt{n}} + n^{-1}\left\{\left(\frac{\mu_4}{\sigma^4}-3\right)\left(\frac{x^3-3x}{4!}\right) + \frac{10\ \mu_3^2(x^5-10x^3+15x)}{\sigma^6 6!}\right\}\right.$$

$$+ n^{-3/2}\left\{\left(\frac{\mu_5}{\sigma^5}-10\frac{\mu_3}{\sigma^3}\right)\left(\frac{x^4-6x^2+3}{5!}\right) + \frac{\mu_3}{\sigma^3}\left(\frac{\mu_4}{\sigma^4}-3\right)\left(\frac{35(x^6-15x^4+45x^2-15)}{7!}\right)\right.$$

$$\left.\left. + \left(\frac{280\mu_3^3(x^8-28x^6+210x^4-420x^2+105)}{\sigma^9 9!}\right)\right\}+0(n^{-5/2})\right].$$

To obtain the cumulative density function of $\{(x_1-\mu)/\sigma\}$, one only needs to put n=1 in the above expansion.

This type of expansion gives a somewhat poor approximation in the tails, which is precisely in the region of most interest. Saddlepoint techniques (see e.g. Daniels, 1954) and Hampel's (1973) procedure are two of the existing methods that can be applied to obtain very accurate approximations.

Although inadequate for some other purposes, the Edgeworth expansion is essential for the main discussions of this chapter. Other methods may be more accurate but these methods normally require numerical computation and do not have explicit expressions. For this latter reason, Saddlepoint techniques and other numerical procedures are not considered here.

The use of Edgeworth expansions as approximations to the finite sample distributions of econometric estimators has been considered by Sargan (1976), Phillips (1978) and Sargan and Tse (1979). A general theorem on the validity of Edgeworth expansions, with respect to its applications to finite sample distributions of econometric estimators, can be found in Sargan and Satchell (1986).

4.3 Bootstrap Confidence Intervals of β When ε is Observable

When ε is observable and when a sample of size n of ε can be obtained for bootstrapping, the naive BCI of β is similar to those BCI's considered by Efron (1985, 1987). Improvements can be made to this BCI by adopting the method of either Diciccio and Tibshirani (1987) or Loh (1987). The first method consists of the composition of a variance-stabilizing transformation and a skewness-reducing transformation. On the other hand, the second method involves computer simulation and density estimation.

The present approach is based upon a similar idea of Babu and Singh (1983). The following discussion is on bootstrap estimates of linear regression coefficients, $\beta_k (k=1,\ldots,K)$. For simplicity, the subscript k will be omitted. Let BCI0 be the naive BCI, based upon the percentiles of the bootstrap distribution of $\hat{\beta}$. Also let BCI1 and BCI2 be the BCI's based upon the percentiles of ordered samples of bootstrap estimates of $\sigma_{\beta}^{-1}(\hat{\beta}-\beta)$ and $s_{\beta}^{-1}(\hat{\beta}-\beta)$, respectively. The variance of $\hat{\beta}$ is σ_{β}^{2} and s_{β}^{2} is an unbiased estimate of σ_{β}^{2}. The three BCI's are then compared when OLS, inflated OLS and BLUS residuals are used for bootstrapping. This will be done in the subsequent section.

The case when ε can be observed will be used as the bench-mark for comparison. In this section, Edgeworth expansions are derived for bootstrap distributions associated with BCI0, BCI1 and BCI2, for the case when a sample of ε is used for bootstrapping.

Let ε be obtained from a normal population, and let $t_1=\sigma_{\beta}^{-1}(\hat{\beta}-\beta)$ and $t_2=s_{\beta}^{-1}(\hat{\beta}-\beta)$. From least-squares theory, it is known that $\hat{\beta}$ follows a normal distribution. Subsequently, it can easily be shown that t_1 and t_2 follow standardized normal distribution and Student-t distribution with (n-K) degrees of freedom, respectively.

One problem pertaining to distributions of bootstrap estimates of regression coefficients is the lack of a coherent finite-sample theory. In most cases, one has to rely upon existing theorems for a general class of statistics or for a statistic which is not related to regression coefficients. These theorems many sometimes be inappropriate, especially for the case when a sample of ε is not available for bootstrapping.

For the case of sample means, Beran (1982) suggests that the bootstrap distribution of $\sqrt{n}\, s_{\theta}^{-1}(\hat{\theta}-\theta)$ has a faster rate of convergence to its exact distribution that that of the bootstrap distribution of $\sqrt{n}(\hat{\theta}-\theta)$. The sample and population means are denoted by $\hat{\theta}$ and θ, respectively, and s_{θ}^{2} is a consistent estimate of the variance of $\hat{\theta}$. Beran also conjectured that the rate of convergence of $\sqrt{n} s_{\theta}^{-1}(\hat{\theta}-\theta)$ is (n^{-1}). For the same statistic, Babu and Singh (1983) show that the bootstrap approximation is asymptotically more accurate for $s_{\theta}^{-1}(\hat{\beta}-\beta)$ than for $(\hat{\theta}-\theta)$. For a general class of statistics, the histogram of the bootstrap values of t_2 approximates the Student-t distribution with (n-K) degrees of freedom with a remainder of $O(n^{-1})$. [C.f. Efron (1979), Singh (1981) and Bickel and Freedman (1981).]

The lack of a coherent finite-sample theory for the bootstrap confidence intervals of regression coefficients leads to speculations that these confidence intervals also admit errors of $O(n^{-1})$. The following results show that some of these speculations can be erroneous.

For convenience, the subscript k is again omitted here. Let $\beta^{*}_{(j)}(j=1,\dots,J)$ be the bootstrap estimates of β and let

$$\bar{\beta}^{*} \equiv \frac{1}{J}\left\{\sum_{j=1}^{J}\left[\beta^{*}_{(j)}\right]\right\}.$$

Further, let

$$t^{*}_{1(j)} = \sigma_{\beta}^{*-1}(\beta^{*}_{(j)}-\bar{\beta}^{*})$$

and

$$t^{*}_{2(j)} = s_{\beta}^{*-1}(\beta^{*}_{(j)}-\bar{\beta}^{*}),$$

in which $j=1,\dots,J$,

$$\sigma_{\beta}^{*2} = (J-1)^{-1}\left\{\sum_{j=1}^{J}\left[\beta^{*}_{(j)}-\bar{\beta}^{*}\right]^{2}\right\}$$

and s_{β}^{*2} is an unbiased estimate of $\lim_{J\to\infty} \sigma_{\beta}^{*2}$.

The following lemma compares the Edgeworth expansions of t_1 and t_1^* for the case when a sample of ε can be obtained for bootstrapping and when ε comes from a normal population. Let F_1 and $\hat{F}_1$ denote the distributional laws of t_1 and t_1^*, respectively. Also, let $\Phi(x)$ and $\phi(x)$ be the cumulative distribution and density of the standard normal variate, respectively.

Lemma 4.1: Suppose that t_1 admits the following Edgeworth expansion

$$p\left\{t_1 \le x\right\} = \Phi(x) + \left\{\sum_{i=1}^{k-1} n^{-i/2} p_i(F_1, x)\right\} + O(n^{-k/2})$$

where the p_i's are certain polynomials in x and $\phi(x)$, whose coefficients depend upon the first few central moments of F_1. Also, assume that t_1^* admits an Edgeworth expansion

$$p\left\{t_1^* \le x \mid \hat{F}_1\right\} = \Phi(x) + \left\{\sum_{i=1}^{k-1} n^{-i/2} p_i(\hat{F}_1, x)\right\} + O(n^{-k/2}).$$

In the case when ε comes from a normal population and when a sample of size n can be obtained for bootstrapping,

$$(1) \quad p\left\{t_1^* \le x \mid \hat{F}_1\right\} = p\left\{t_1 \le x\right\} + O(n^{-1/2}).$$

The above is true when n is finite. As $n\to\infty$, (1) becomes

$$(2) \quad p\left\{t_1^* \le x\right\} = p\left\{t_1 \le x\right\} + O(n^{-2}).$$

Proof:

When ε is normal, all odd central moments of $\hat{\beta}$ will be zero and $\mu_4 = 3\sigma^4$. In this case, t_1 follows a standard normal distribution. Hence,

$$p\left\{t_1 \le x\right\} = \Phi(x).$$

However, t_1^* does not always follow a standard normal distribution, and

$$p\left\{t_1^* \le x \,|\, \hat{F}_1\right\} = \Phi(x) - \phi(x)\left[n^{-1/2}\left(\tilde{\mu}_3\tilde{\mu}_2^{-3/2}\right)\left(\frac{x^2-1}{3!}\right) + n^{-1}\left\{\left(\tilde{\mu}_4\tilde{\mu}_2^{-2}-3\right)\right.\right.$$

$$\left.\left.\left(\frac{x^3-3x}{4!}\right) + 10\tilde{\mu}_3^2\tilde{\mu}_2^3\left(\frac{x^5-10x^3+15x}{6!}\right)\right\}\right] + O(n^{-3/2}),$$

in which $\tilde{\mu}_i$ $(i=2,3,4)$ are the sample moments of ε. Thus, it can be shown that t_1^* follows a standard normal distribution, provided that $\tilde{\mu}_{2r+1}=0$ and $\tilde{\mu}_{2r} = (2^r r!)^{-1}\left\{\tilde{\mu}_2^r\left[(2r)!\right]\right\}$, in which $r=1,2,\ \ .$

In the leading term of $O(n^{-1/2})$, $\left(\tilde{\mu}_3\tilde{\mu}_2^{-3/2}\right)$ is $O(1)$, because both $\tilde{\mu}_2$ and $\tilde{\mu}_3$ are $O(1)$. Consequently, the term $\left\{n^{-1/2}\left(\tilde{\mu}_3\tilde{\mu}_2^{-3/2}\right)\left(\frac{x^2-1}{3!}\right)\right\}$ is $O(n^{-1/2})$, and

$$p\left\{t_1^* \le x \,|\, \hat{F}_1\right\} = \Phi(x) + O(n^{-1/2}).$$

The remainder of the proof for part (1) is then straightforward.

For the proof of part (2), slight variants of Theorems 3.2 and 3.3 above are required. However, these variants can easily be obtained.

One needs to show that $E(\tilde{\mu}_3)=0$ and that both $E(\tilde{\mu}_2)$ and $E(\tilde{\mu}_4)$ admit errors $O(n^{-1})$. The proof then becomes straightforward from the Edgeworth expansion above. *Q. E. D.*

The results of Lemma 4.1 are not restricted to the case when the error distribution is normal. Both results also apply to the case when the error distribution is symmetric but not normal. However, it must be noted that for the latter case,

$$p\left\{t_1 \le x\right\} = \Phi(x) + O(n^{-2}).$$

This assumes that $\left[\mu_4\mu_2^{-2}-3\right]$ is $O(n^{-1})$.

The following lemma compares the Edgeworth expansions of t_1 and t_1^* for the case when the error distribution is skewed and when a sample of ε is used for bootstrapping.

Lemma 4.2: Let ε comes from a skewed distribution. Then,

(1) $p\left\{t_1^* \le x \mid \hat{F}_1\right\} = p\left\{t_1 \le x\right\} + O(n^{-1/2})$.

(2) $p\left\{t_1^* \le x\right\} = p\left\{t_1 \le x\right\} + O(n^{-1})$.

Proof:

The proof for part (1) is straightforward from a similar proof in Lemma 4.1 above. For part (2), one needs to know that all $E(\tilde{\mu}_2)$, $E(\tilde{\mu}_3)$ and $E(\tilde{\mu}_4)$ admit errors $O(n^{-1})$. The remainder of the proof is then also straightforward from Lemma 4.1. *Q.E.D.*

In must be mentioned that the results of Lemma 4.6 agree with the earlier results of Efron (1979, 1985), Singh (1981), and Abramovitch and Singh (1985). Thus, those earlier results apply to the more general case when the error distribution is unknown. When the error distribution is symmetric, bootstrap approximation of t_1 is more accurate than what have been suggested by earlier results.

The following theorem concerns the empirical coverages of BCI's based upon t_1^*. Let

$$\int_{-\infty}^{t_1(a)} \phi(x)dx = a, \qquad a=\alpha, (1-\alpha).$$

Also, let $t_1^*(a)$ be the (100a)'th percentile of the bootstrap distribution of t_1^*.

Theorem 4.2: Let the error distribution be symmetric and let $x=\sigma_\beta^{-1}(\hat{\beta}-\beta)$. The lower and upper bounds of x are $t_1(\alpha)$ and $t_1(1-\alpha)$, respectively, such that $p\left[t_1(\alpha)\le x\le t_1(1-\alpha)\right] = 1-2\alpha$.

Then,

$$p\left\{t_1^*(\alpha)\le x\le t_1^*(1-\alpha)\right\} < 1-2\alpha.$$

The difference is $O(n^{-2})$.

Proof:

From the proof Lemma 4.1, it can be shown that

$$p\left\{t_1^*(\alpha)\le t_1(\alpha)\right\} < \alpha$$

and $$p\left\{t_1^*(1-\alpha)\le t_1(1-\alpha)\right\} < (1-\alpha)$$

Consequently,

$$p\left\{x<t_1^*(\alpha)\right\} > \alpha$$

and $$p\left\{x<t_1^*(1-\alpha)\right\} < (1-\alpha)\,.$$

Upon application of Lemma 4.1, the remainder of the proof is then straightforward. *Q.E.D.*

Theorem 4.3: Let the notation be that of Theorem 4.2 and let the error distribution be nonsymmetric. Then,

$$p\left\{t_1^*(\alpha)\le x\le t_1^*(1-\alpha)\right\} < 1-2\alpha$$

and the difference will be $O(n^{-1})$.

Proof:

The proof is similar to that of Theorem 4.2. It is rather straightforward upon application of Lemma 4.2. *Q.E.D.*

Both Theorems 4.2 and 4.3 above suggest that the empirical coverage of BCI1 will be lower than nominal when observations of the true disturbances are used for bootstrapping. However, this coverage will be slightly higher when the error distribution is symmetric as compared to the case when error distribution is nonsymmetric.

The following lemmas compare the Edgeworth expansions of t_2 and t_2^*. As in the earlier case, it will be assumed that a sample of ε can be obtained for bootstrapping. An unbiased estimate of σ^2 can be obtained from $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$. Consequently, t_2 follows a Student-t distribution with (n-1) degrees of freedom, provided that ε comes from a normal population. In this case, the distribution of t_2 is not affected by K. Further, let F_2 and $\hat{F}_2$ denote the distributional laws of t_2 and t_2^*, respectively.

Lemma 4.3: Let the error distribution be normal and let

$$p\left\{t_2 \le x\right\} = \Phi(x) + \left\{\sum_{i=1}^{k-1} n^{-i/2} P_i(F_2, x)\right\} + O(n^{-k/2}),$$

Further, let

$$p\left\{t_2^* \le x\right\} = \Phi(x) + \left\{\sum_{i=1}^{k-1} n^{-i/2} P_i(\hat{F}_2, x)\right\} + O(n^{-k/2}).$$

Then,

(1) $p\left\{t_2^* \le x \mid \hat{F}_2\right\} = p\left\{t_2 \le x\right\} + O(n^{-1/2})$.

(2) $p\left\{t_2^* \le x\right\} = p\left\{t_2 \le x\right\} + O(n^{-2})$.

Proof:

First note that t_2 is a Student-t distribution with (n-1) degrees of freedom. Since the distribution of t_2 is symmetric, all its odd moments equal zero. Hence,

$$p\left\{t_2 \le x\right\} = \Phi(x) + n^{-1}\left\{\left[\mu_4 - 3\right]\left[\frac{x^3 - 3x}{24}\right]\right\} + 0\left[n^{-3/2}\right]$$

in which $\mu_4 = E\left[s_\beta^{-4}(\hat{\beta} - \beta)^4\right]$. The results are not affected by the assumption that $\tilde{\mu}_2 = \mu_2 = 1$. Consequently, it is possible to write

$$P\left\{t_2^{\bullet}\le x|\hat{F}_2\right\} = \Phi(x)-\phi(x)\left[n^{-1/2}\left(\tilde{\mu}_3\right)\left(\frac{x^2-1}{6}\right) + n^{-1}\left\{\left(\tilde{\mu}_4-3\right)\right.\right.$$

$$\left.\left.\left(\frac{x^3-3x}{24}\right) + \tilde{\mu}_3^2\left(\frac{x^5-10x^3+15x}{72}\right)\right\}\right] + O(n^{-3/2}).$$

Thus,

$$P\left\{t_2^{\bullet}\le x|\hat{F}_2\right\} = P\left\{t_2\le x\right\} + \phi(x)\left[n^{-1/2}\left(\tilde{\mu}_3\right)\left(\frac{x^2-1}{6}\right) + n^{-1}\left\{\left(\tilde{\mu}_4-\tilde{\mu}_4\right)\right.\right.$$

$$\left.\left.\left(\frac{x^3-3x}{24}\right) + \tilde{\mu}_3^2\left(\frac{x^5-10x^3+15x}{72}\right)\right\}\right] + O(n^{-3/2})$$

$$= P\left\{t_2\le x\right\} + O(n^{-1/2}).$$

This assumes that both $\tilde{\mu}_3$ and $\tilde{\mu}_4$ are $O(1)$.

Secondly, it can be shown that $E(\tilde{\mu}_3)=0$ and $E(\tilde{\mu}_4)=\mu_4+O(n^{-1})$. Hence,

$$P\left\{t_2^{\bullet}\le x\right\} = P\left\{t_2\le x\right\} + \phi(x)\left[n^{-1/2}\left\{\left[E(\tilde{\mu}_4)-\mu_4\right]\left(\frac{x^3-3x}{24}\right)\right\}\right] + O(n^{-2})$$

$$= P\left\{t_2\le x\right\} + O(n^{-2}). \quad Q.E.D.$$

The problem of deriving Edgeworth expansions for Student's t-statistic is quite an old concept. It dates back to the work of Pearson and Adyanthaya (1929) and Bartlett (1935). Recent examples include Cressie (1980), Hall (1983) and Abramovitch and Singh (1985).

The Edgeworth expansions in Lemma 4.3 are similar to an Edgeworth expansion obtained by Hall (1983) for the sample mean of a random variable. Similar expansions are also obtained for a general statistic by Abramovitch and Singh (1985), whose motivation is to obtain a modified t-statistic which can be closely approximated by the standard normal distribution. In contrast, the motivation in Lemma 4.3 is to compare the bootstrap distribution of t_2 with its exact distribution.

For a general statistic, the bootstrap distribution of t_2 approximates its exact distribution with an error term of order $O(n^{-1})$. This is stated in Efron (1979), Singh (1981) and Babu and Singh (1983). However, in the case of $\hat{\beta}$, Lemma 4.3 demonstrates that this statement understates the accuracy of a bootstrap distribution of t_2 when the error distribution is normal. The results of Lemma 4.3 are not restricted to the case when the error distribution is normal. Both results also apply to the case when the error distribution is symmetric but not normal.

Lemma 4.4 below compares the Edgeworth expansions of t_2 and t_2^* for the case when the error distribution is nonsymmetric and when a sample of ε can be obtained for bootstrapping.

Lemma 4.4: Let the error distribution be nonsymmetric. Then,

$$(1) \quad p\left\{t_2^* \le x \,|\, \hat{F}_2\right\} = p\left\{t_2 \le x\right\} + O(n^{-1/2}).$$

$$(2) \quad p\left\{t_2^* \le x\right\} = p\left\{t_2 \le x\right\} + O(n^{-1}).$$

Proof:

The results are similar to those of Lemma 4.2. The proof for part (1) can be obtained in a straightforward fashion from a similar proof in Lemma 4.3. For part (2), the same arguments for Lemma 4.2 apply and the proof is also straightforward from Lemma 4.3. *Q.E.D.*

The following theorems concern the empirical coverages of BCI's based upon t_2^*. Let $\tau(x)$ be the Student's density function and let

$$\int_{-\infty}^{t_2(a)} \tau(x)dx = a, \qquad a=\alpha, (1-\alpha).$$

Further, let $t_2^*(a)$ be the (100a)'th percentile of ordered samples of t_2^*.

Theorem 4.4: Let the error distribution be symmetric and let $x=s_\beta^{-1}(\hat{\beta}-\beta)$. The lower and upper bounds of x are $t_2(\alpha)$ and $t_2(1-\alpha)$, respectively, such that $p\left\{t_2(\alpha)\le x\le t_2(1-\alpha)\right\}\le 1-2\alpha$.

Then,

$$p\left\{t_2^*(\alpha)\le x\le t_2^*(1-\alpha)\right\}\le 1-2\alpha.$$

The difference is $O(n^{-2})$.

Proof:

Upon application of Lemma 4.3, the proof is similar to that of an earlier theorem, Theorem 4.2. *Q.E.D.*

Theorem 4.5: Let the notation be that of Theorem 4.4 and let the error distribution be nonsymmetric. Then,

$$p\left\{t_2^*(\alpha)\le x\le t_2^*(1-\alpha)\right\} < 1-2\alpha$$

and the difference will be $O(n^{-1})$.

Proof:

The proof is similar to that of Theorem 4.2. One needs only to apply Lemma 4.4 to complete the proof. *Q.E.D.*

The results of Theorems 4.4 and 4.5 are similar to those of two earlier theorems, Theorems 4.2 and 4.3, respectively. One interesting observation is that the rates of convergence of the bootstrap distributions of t_1 and t_2 to their respective exact distributions are of the same order. This result is not affected by the underlying distribution of ε.

For bootstrap estimates of regression coefficients, the bootstrap distributions of $(\hat{\beta}-\beta)$ and $\sigma_{\beta}^{-1}(\hat{\beta}-\beta)$ are similar, except for a scaling factor. Thus, the rates of convergence of these two bootstrap distributions to their respective exact distributions are identical. Consequently, it can be stated that, for the case of linear regression coefficients, the rates of convergence of the bootstrap distributions of $\sigma_{\beta}^{-1}(\hat{\beta}-\beta)$ and $(\hat{\beta}-\beta)$ to their respective exact distributions are of the same order.

In most applications, σ_{β}^{2} is generally unknown and has to be replaced with s_{β}^{2}. For these cases, the student-t distribution is the appropriate distribution for constructing confidence intervals, especially when the sample size is small. Using the same analogy, it can be stated that BCI2 is the appropriate BCI for obtaining confidence intervals. It must be mentioned that both BCI0 and BCI1 will give the same confidence intervals. This happens in the case of linear regression coefficients, provided that ε is observable and that ε is scaled by the factor $\left[n(n-1)^{-1}\right]^{1/2}$ prior to bootstrapping.

The subsequent section investigates the three cases when the OLS, inflated OLS and BLUS residuals are used for bootstrapping. The results are not substantially different from those obtained in this section.

4.4 Bootstrap Confidence Intervals of β When ε is Not Observable

In applications of Efron's bootstrap to the estimation of confidence intervals for the unknown regression coefficients using real world data, ε is not observable in practically all cases. This poses some problems for the unsuspecting investigator. This investigator is mostly likely to use the OLS residuals since they are relatively easy to compute.

The most serious problem arises when the naive BCI is used. In this case, the estimated confidence intervals is much shorter than the exact confidence interval of β, and the coverage of this BCI is very poor, especially when the number of regression coefficients is large (relative to n). Besides the OLS residuals, two other types of regression residuals, the inflated OLS and BLUS residuals, can also be used for bootstrapping.

In this section, the effects of the choice of residuals on the accuracies of BCI0, BCI1 and BCI2 are investigated. The choice of residuals is limited to the OLS, inflated OLS and BLUS residuals. In the case when inflated OLS residuals are used for bootstrapping, the result of Theorem 2.12 guarantees that both BCI0 and BCI1 will yield the same results, provided that J goes to infinity. Similarly, the result of Theorem 2.15 guarantees that, when BLUS residuals are used for bootstrapping and when J goes to infinity, both BCI0 and BCI1 will yield the same results. The difference between BCI0 and BCI1 lies in the difference between the dispersion of the bootstrap estimates of β and the least squares estimate of $D(\hat{\beta})$. Both Theorems 2.12 and 2.15

guarantee that this difference goes to zero as J goes to infinity. Thus, using the result of Theorem 2.10, it can be shown that when OLS residuals are used for bootstrapping, BCI0 and BCI1 differs by scale factor $\left[n(n-K)^{-1}\right]$. Specifically, BCI0 will be shorter than BCI1 and this difference can be corrected by multiplying both lower and upper bounds of BCI0 by the factor $\left[n(n-K)^{-1}\right]$.

The following is also in order. Namely, the same BCI1 is obtained when either OLS or inflated OLS residuals are used for bootstrapping. The same is also true for BCI2. Consequently, one only needs to investigate the properties of BCI1 and BCI2 for the two cases when either inflated OLS or BLUS residuals are used for bootstrapping.

The notation used for the following lemmas is the same as that used for Lemmas 4.1 through to 4.4. Lemmas 4.5 through to 4.7 below concern the case when inflated OLS residuals are used for bootstrapping. The case when BLUS residuals are used for bootstrapping is illustrated in Lemmas 4.8 through to 4.10.

For the following three lemmas, it is assumed that inflated OLS residuals were used for bootstrapping and that J goes to infinity.

Lemma 4.5: Let the error distribution be normal. Then,

(1) $p\left\{t_1^* \le x \middle| \hat{F}_1\right\} = p\left\{t_1 \le x\right\} + O(n^{-1/2})$.

(2) $p\left\{t_1^* \le x\right\} = p\left\{t_1 \le x\right\} + O(kn^{-2})$.

Proof:

The proof for part (1) is similar to that of Lemma 4.1. In fact the two results are identical. When ε is normal, it is easy to establish that

$$p\left\{t_1 \le x\right\} = \Phi(x).$$

Let $\tilde{\mu}_i$ (i=2,3,4) be sample moments of the inflated OLS residuals. Then, it can easily be shown that

$$p\left\{t_1^{*} \le x \,|\, \hat{F}_1\right\} = \Phi(x) - \phi(x)\left[n^{-1/2}\left(\tilde{\mu}_3\tilde{\mu}_2^{-3/2}\right)\left(\frac{x^2!}{3!}\right) + n^{-1}\left\{\left(\tilde{\mu}_4\tilde{\mu}_2^{-2}-3\right)\right.\right.$$

$$\left.\left.\left(\frac{x^3-3x}{4!}\right) + 10\tilde{\mu}_3^{2}\tilde{\mu}_2^{3}\left(\frac{x^5-10x^3+15x}{6!}\right)\right\}\right] + O(n^{-3/2}).$$

Note that $\tilde{\mu}_i$ (i=2,3,4) are O(1). Consequently,

$$p\left\{t_1^{*} \le x \,|\, \hat{F}_1\right\} = \Phi(x) + O(n^{-1/2}).$$

For the proof of part (2), one requires the applications of Theorems 2.12, 3.2 and 3.3. It remains to show that both $\tilde{\mu}_2$ and $\tilde{\mu}_3$ are unbiased estimates of μ_2 and μ_3, respectively, and that $E(\tilde{\mu}_4)$ admits errors $O(Kn^{-1})$. The proof then becomes rather straightforward from the above Edgeworth expansion. Q.E.D.

The results of Lemma 4.5 are not restricted to the case when the error distribution is normal. Both results can also be applied to any case, whenever the error distribution is symmetric.

Lemma 4.6: Let ε comes from a nonsymmetric distribution. Then,

(1) $p\left\{t_1^{*} \le x \,|\, \hat{F}_1\right\} = p\left\{t_1 \le x\right\} + O(n^{-1/2}).$

(2) $p\left\{t_1^{*}\le x\right\} = p\left\{t_1\le x\right\} + O\left[n^{-1}\sqrt{K}\right].$

Proof:

The proof of part (1) is straightforward from a similar proof in Lemma 4.5. For part (2), one needs to know that $E(\tilde{\mu}_3)$ and $E(\tilde{\mu}_4)$ admit errors of $O\left[n^{-1/2}\sqrt{K}\right]$ and $O(Kn^{-1})$, respectively, while $\tilde{\mu}_2$ is unbiased. The remainder of the proof is then also straightforward from Lemma 4.5. Q.E.D.

Lemmas 4.5 and 4.6 compare the bootstrap distribution of t_1 with its exact distribution. The results are similar to those of Lemmas 4.1 and 4.2, respectively. When K=1, the two sets of results are identical. As the number of regression coefficients increases, bootstrap approximations of t_1 based upon inflated OLS residuals become less accurate.

The following theorem gives the accuracies of BCI's based upon t_1^{*}, when inflated OLS residuals are used for bootstrapping. The notation is that of Theorem 4.2.

Theorem 4.6: (1) Let the error distribution be symmetric. Then,

$$p\left\{t_1^{*}(\alpha)\le x\le t_1^{*}(1-\alpha)\right\}<1-2\alpha$$

and the difference will be $O(Kn^{-2})$.

(2) The difference will be $O\left[n^{-1}\sqrt{K}\right]$ when the error distribution is nonsymmetric.

Proof: With reference to the proof of Theorem 4.2, the proof is rather straightforward upon applications of Lemmas 4.5 and 4.6. Q.E.D.

Similar extension can also be made to the bootstrap distribution of t_2 in the case when inflated OLS residuals are used for bootstrapping. Since the proofs of the following lemmas can easily be adapted from the proofs of Lemmas 4.3 through to 4.6, Lemma 4.7 below will be stated without proof.

Lemma 4.7: (1) Let the error distribution be continuous having finite moments up to the fifth moment. Then,

$$p\left\{t_2^* \le x \mid \hat{F}_2\right\} = p\left\{t_2 \le x\right\} + O(n^{-1/2}).$$

(2) $p\left\{t_2^* \le x\right\} = p\left\{t_2 \le x\right\} + O(Kn^{-2})$ in the case when the error distribution is symmetric.

(3) Let the error distribution be nonsymmetric. Then,

$$p\left\{t_2^* \le \quad\right\} = p\left\{t_2 \le x\right\} + O\left[n^{-1}\sqrt{K}\right].$$

Theorem 4.7: Let the notation be that of Lemma 4.7 and Theorem 4.4. Then,

$$p\left\{t_2^*(\alpha) \le x \le t_2^*(1-\alpha)\right\} < 1-2\alpha.$$

The difference will be $O(Kn^{-2})$ when the error distribution is symmetric. Otherwise, it will be $O\left[n^{-1}\sqrt{K}\right]$.

Proof:

Upon application of Lemma 4.7, the proof is similar to that of Theorem 4.2. Q.E.D.

It is observed again in Theorems 4.6 and 4.7, that the rates of convergence of the bootstrap distributions of t_1 and t_2 to their respective exact distributions are of the same order. Both theorems suggest that the empirical coverages of BCI1 and BCI2 will be lower than nominal, when inflated OLS residuals are used for bootstrapping. These coverages are also lower, when compared to the case when observations of the true disturbances are used for bootstrapping. However, these coverages are slightly higher when the error distribution is symmetric, as compared to the case when the error distribution is nonsymmetric.

The results of Theorems 4.6 and 4.7 suggest that, when K is small and when the error distribution is symmetric, bootstrap approximations of both t_1 and t_2 are more accurate than what have been suggested by the results of Efron (1979, 1985) and others. The latter results suggest that the bootstrap approximations admit errors $O(n^{-1})$. On the other hand, when K is large relative to n and when the error distribution is nonsymmetric, bootstrap approximation of both t_1 and t_2 are poorer than expected. In this latter case, the bootstrap approximations may admit errors $O(n^{-1/2})$, rendering the BCI's to be of no practical value.

One other type of regression residuals that can be used for bootstrapping is the class of BLUS residuals. When bootstrap approximations based upon inflated OLS residuals failed to admit errors up to $O(n^{-1})$, bootstrap approximations based upon BLUS residuals are feasible alternatives. This is illustrated by the following lemmas and theorems.

To fix ideas, let $L=(n-K)$, and let $\tilde{\varepsilon}=B\hat{e}$, in which $\hat{\varepsilon}=(\hat{\varepsilon}_1,\hat{\varepsilon}_2,\ldots,\hat{\varepsilon}_n)^T$ is the nx1 vector of OLS residuals and such that $BX=0$, $BB^T=I_L$ and $B^TB=M$. It is further assumed that $\tilde{\varepsilon}$, the Lx1 vector of BLUS residuals, is used for bootstrapping and that J goes to infinity. The remaining notation is mainly that of Lemma 4.5.

Lemma 4.8: Let the errors be normally distributed, and let $\tilde{\varepsilon}$ be centered prior to bootstrapping, such that the sample mean of $\tilde{\varepsilon}$ is zero. Further, let the centered BLUS residuals be multiplied by the factor $\left\{n(n-1)^{-1}\right\}^{1/2}$. Then,

$$(1) \quad p\left\{t_1^{*}\le x \mid \hat{F}_1\right\} = p\left\{t_1 \le x\right\} + O(n^{-1/2})$$

$$(2) \quad p\left\{t_2^{*}\le x \mid \hat{F}_2\right\} = p\left\{t_2 \le x\right\} + O(n^{-1/2})$$

$$(3) \quad p\left\{t_1^{*}\le x\right\} = p\left\{t_1 \le x\right\} + O\left[(nL)^{-1}\right]$$

$$(4) \quad p\left\{t_2^{*}\le x\right\} = p\left\{t_2 \le x\right\} + O\left[(nL)^{-1}\right]$$

Proof:

The proofs of parts (1) and (2) are similar to those of Lemmas 4.1 and 4.3 above, respectively. For the proof of parts (3) and (4), one requires the application of Lemmas 3.13 and Theorems 2.14, 3.2 and 3.4.

Note that while the fourth sample moment of $B\hat{e}$ is an unbiased estimate of μ_4, the fourth sample moment of the tranformed BLUS residuals admits an error $O(L^{-1})$. On the other hand, the second and third sample moments of the transformed BLUS residuals are unbiased. The remainder of the proof is then straightforward from the proofs of Lemmas 4.1 and 4.3. Q.E.D.

Lemma 4.9: Let the notation be that of Lemma 4.8, and let the error distribution be symmetric. Then,

(1) $p\left\{t_1^{\bullet} \le x\right\} = p\left\{t_1 \le x\right\} + O(\eta^{-1})$

(2) $p\left\{t_2^{\bullet} \le x\right\} = p\left\{t_2 \le x\right\} + O(\eta^{-1})$

in which $n<\eta<n^3$, the value of η depending on the type of error distribution. Specifically, $\eta=Ln^2$ for a certain leptokurtic error distribution on the one hand, while on the other hand, η approaches n when the error distribution is either extremely platykurtic or extremely leptokurtic.

Proof:

The proofs are similar to those of Lemmas 4.1, 4.3 and 4.8, except for the application of Lemma 3.16. Upon application of Lemma 3.16, it can easily be seen that when the error distribution is leptokurtic, the fourth sample moment of $B\hat{e}$ overestimates μ_4. On the other hand, the fourth sample moment of the transformed BLUS residuals can be an unbiased, over or underestimate of μ_4, depending on the leptokurtosis. For small leptokurtosis, it is likely that $E(\tilde{\mu}_4)<\mu_4$. In the case when the leptokurtosis is very large, one would expect that $E(\tilde{\mu}_4)>\mu_4$. Thus, there exists a leptokurtic error distribution such that $E(\tilde{\mu}_4)=\mu_4$. When this happens, it can be shown that $E(\tilde{\mu}_5)=0$ and that $E(\tilde{\mu}_6)$ admits an error $O(L^{-1})$. Consequently, for this particular case, $\eta=Ln^2$.

When the error distribution is platykurtic, the fourth sample moment of the transformed BLUS residuals underestimates μ_4. However, this underestimation can at most be O(1). This only happens when the platykurtosis of the error distribution is extremely large. In other words, the error distribution should be extremely flat in order for the above to be observed. In this case, η approaches n.

When the leptokurtosis of the error distribution is extremely large or when the error distribution is extremely peaked, one would observe an overestimation of μ_4 by the fourth sample moment of the transformed BLUS residuals. This overestimation can at most be O(1). Consequently, the least value which η can attain is n. Q.E.D.

Lemma 4.8 states that when the errors are normally distributed, the rates of convergence of bootstrap distributions of t_1 and t_2 to their exact distributions are the same. Both bootstrap approximations of t_1 and t_2 admit errors $O\left[(nL)^{-1}\right]$. When K is large relative to n, these approximations are better than those based upon inflated OLS residuals, which admit errors $O\left[Kn^{-2}\right]$. However, when K is small, the gain (in terms of accuracy) in using BLUS residuals for bootstrapping may be negligible.

In the case when the error distribution is leptokurtic, there exists a leptokurtic distribution such that the bootstrap approximations can be extremely accurate. This is one of the statements of Lemma 4.13. When the leptokurtosis of an error distribution is small to moderate, it is better to use BLUS residuals (in comparison to inflated OLS residuals) for bootstrapping, especially when K is large relative to n. On the other hand, when the leptokurtosis (or platykurtosis) is extremely large, one is better off using inflated OLS residuals.

The results of Lemmas 4.8 and 4.9 have implications for the empirical converages of BCI's. The following theorem examines the accuracies of BCI1 and BCI2, when transformed BLUS residuals are used for bootstrapping. The notation is mainly that of Lemma 4.8 and

Theorems 4.2 and 4.4. It is assumed that J goes to infinity.

Theorem 4.8: Let the error distribution be symmetric, such that moments up to the sixth order are finite. Then,

(1) $p\left\{t_1^{*}(\alpha)\le x\le t_1^{*}(1-\alpha)\right\} \gtrless 1-2\alpha.$

(2) $p\left\{t_2^{*}(\alpha)\le x\le t_2^{*}(1-\alpha)\right\} \gtrless 1-2\alpha.$

The differences will be $O(\eta^{-1})$, $n<\eta\le n^3$, depending on the kurtosis of the error distribution.

Further, let $\mu_4=\tau\mu_2^2$ and let there exists a τ^* such that $E(\tilde{\mu}_4)=\mu_4$. Then, the left hand sides of (1) and (2) are less than $(1-2\alpha)$ when $\tau\le\tau^*$. Otherwise, the contrary holds. When $\tau=\tau^*$, $\eta=n^3$; whereas, when $\tau=3$, $\eta=nL$. In the extreme case when $\tau\to\infty$ (or $-\infty$), $\eta\to n$.

Proof:

The proof is rather straightforward upon applications of Lemmas 4.8 and 4.9. Q.E.D.

The results are different for the case when the error distribution is nonsymmetric. This is evident in the following lemma. The notation is that of Lemmas 4.1 and 4.8, and it is assumed that J goes to infinity.

Lemma 4.10: Let the error distribution be skewed. Then,

(1) $p\left\{t_1^{*}\le x\,|\,\hat{F}_1\right\} = p\left\{t_1\le x\right\} + O(n^{-1/2})$

(2) $p\left\{t_2^{*}\le x\,|\,\hat{F}_2\right\} = p\left\{t_2\le x\right\} + O(n^{-1/2})$

(3) $p\left\{t_1^{*}\le x\right\} = p\left\{t_1\le x\right\} + O\left[(nL)^{-1/2}\right]$

(4) $p\left\{t_2^{*}\le x\right\} = p\left\{t_2\le x\right\} + O\left[(nL)^{-1/2}\right]$

Proof:

The proof is similar to that of Lemma 4.8. Upon application of Lemma 3.13 and Theorem 2.14, one needs to note that $E(\tilde{\mu}_3\tilde{\mu}_2^{-3/2})$ admits an error $O(L^{-1/2})$. The rest of the proof is then rather straightforward from the Edgeworth expansions of t_1^{*} and t_2^{*}. Q.E.D.

Theorem 4.9: In the case when the error distribution is nonsymmetric,

(1) $p\left\{t_1^{*}(\alpha)\le x\le t_1^{*}(1-\alpha)\right\} < 1-2\alpha.$

(2) $p\left\{t_2^{*}(\alpha)\le x\le t_2^{*}(1-\alpha)\right\} < 1-2\alpha.$

The differences will be $O(\eta^{-1})$, η=nL. The notation is mainly that of Lemma 4.8, Theorems 4.2 and 4.4.

Proof:

The proof is rather straightforward upon application of Lemma 4.10. Q.E.D.

It is apparent from Theorem 4.9 that when the error distribution is nonsymmetric, it would be advantageous to use BLUS residuals for bootstrapping, especially when K is large relative to n. When K is large, BCI's based upon inflated OLS residuals are shorter than the exact confidence interval and the difference is $O(n^{-1/2})$. In comparison, BCI's based upon (transformed) BLUS residuals admit errors $O(\eta^{-1/2})$, η=nL. This is slightly less than the $O(n^{-1})$ accuracy, which is suggested by earlier studies. However, it is still an improvement since conventional approximations admit errors $O(n^{-1/2})$.

4.5 A Note on the Robustness of Bootstrap Confidence Intervals of β

The bootstrap method has generally been thought of as a distribution-free method when it "does not rely for its validity or its utility on any assumptions about the form of distribution that is taken to have generated the sample values on the basis of which references about the population distribution are to be made." (Maritz, 1981 p.1). Based upon this broad definition, the bootstrap method is distribution-free since its validity does not depend on the underlying distribution. This is the view of existing results in the literature.

In this section, Hampel's definition of robustness (Theorem 4.1 above) is used to study the robustness of BCI's against alternative distributions. The results of Theorems 4.2 through to 4.5 suggest that when a sample of size n of the true errors are used for bootstrapping, both BCI1 and BCI2 are robust to alternative distributions up to $O(n^{-1})$. With proper scaling of the observations on the true errors, BCI0 is also robust up to $O(n^{-1})$. Otherwise, it is robust up to $O(n^{-1/2})$.

When inflated OLS residuals (or OLS residuals) are used for bootstrapping, the results of Theorems 4.6 and 4.7 suggest that both BCI1 and BCI2 are robust to alternative distributions up to $O\left(n^{-1}\sqrt{K}\right)$. BCI0 based upon OLS residuals admits an error $O(n^{-1/2})$ while that based upon inflated OLS residuals admits an error of at most $O\left(n^{-1}\sqrt{K}\right)$.

On the other hand, when properly transformed BLUS residuals are used for bootstrapping, the results of Theorems 4.8 and 4.9 suggest that BCI0, BCI1 and BCI2 are all robust up to $O(\eta^{-1/2})$ η=nL. Thus, the conventional view that the bootstrap method is robust against alternative distributions up to $O(n^{-1})$ is somewhat erroneous.

4.6 Summary

Edgeworth expansions are used to compare the bootstrap distributions of t_1 and t_2 to their exact distributions. The speeds of convergence of both bootstrap distributions to their exact distributions are shown to be the same for all types of error distributions.

In the case when a sample of size n of the true errors is used for obtaining the bootstrap estimates of the regression coefficients, it is shown that both bootstrap distributions of t_1 and t_2 admit errors of at most $O(n^{-1})$. This is consistent with the results of Efron (1979,1985), Singh (1981) and Abramovitch and Singh (1985). When the error distribution is symmetric, the bootstrap approximations are more accurate than the $O(n^{-1})$ accuracy suggested above.

Three types of bootstrap confidence intervals are al... studied. These are BCI0, BCI1, and BCI2. BCI0 is the 'naive' BCI obtained by ordering the bootstrap estimates of β. BCI1 and BCI2 are based upon t_1 and t_2, respectively. When the sample size is small to moderate, BCI2 is the appropriate interval. Nevertheless, BCI2 is still shorter than the exact confidence interval and the difference is at most $O(n^{-1})$. When the error distribution is symmetric, this difference will be $O(n^{-2})$. These results are valid for the case when the true errors can be observed.

When the true errors are nonobservable and when regression residuals are used for bootstrapping, the bootstrap estimates may yield inferior results, especially when the number of regression coefficients is large relative to n. When OLS or inflated OLS residuals are used for bootstrapping, the bootstrap approximations of t_1 and t_2 admit errors $O\left[n^{-1}\sqrt{K}\right]$. In comparison, these approximations admit errors $O(\eta^{-1/2})$, $\eta=nL$, when BLUS residuals are used for bootstrapping.

It has been shown that existing bootstrap results for a general statistic cannot be applied to the bootstrap estimates of regression coefficients. One reason is that the true errors are nonobservable. When K is small and n large, the discrepancy between existing results and the results obtained here is negligible. On the other hand, this discrepancy can be significant when K is large relative to n.

For the case when OLS or inflated OLS residuals are used for bootstrapping, the empirical coverages of BCI1 and BCI2 are lower than the nominal coverage. The difference is $O(Kn^{-2})$ when the error distribution is symmetric and $O\left[n^{-1}\sqrt{K}\right]$ otherwise. In the case when OLS residuals are used, BCI0 should be avoided because it is extremely short, especially when n is small and K large.

In comparison, when BLUS residuals are used for bootstrapping the above difference will be $O(\eta^{-1})$, $n<\eta\le n^3$, depending on the kurtosis, when the error distribution is symmetric. When the error distribution is normal, $\eta=n(n-K)$. For nonsymmetric errors, bootstrap approximations admit errors $O(\eta^{-1/2})$, $\eta=n(n-K)$.

Thus, when K is large relative to n, BLUS residuals should be used for bootstrapping. This is consistent with the results of Chapter 3. On the other hand, BLUS residuals are cumbersome to compute. Also, the difference between $n^{-1}\sqrt{K}$ and $\left\{n(n-K)\right\}^{-1/2}$ is negligible when K is small. Hence, when K is small, it is recommended that inflated OLS residuals be used for bootstrapping.

The bootstrap method has generally been thought of as a distribution-free statistical method, at least up to $O(n^{-1})$. Again, this generalization cannot be applied to the case of regression coefficients. Hampel's definition of robustness is used to study the robustness of BCI's against alternative distributions. It is found that both BCI1 and BCI2 are robust to alternative distributions up to $O(n^{-1})$, provided that the true errors are observable. Otherwise, they are robust up to $O\left[n^{-1}\sqrt{K}\right]$ and $O(\eta^{-1/2})$, $\eta=n(n-K)$, when inflated OLS and BLUS residuals are used for bootstrapping, respectively.

CHAPTER FIVE

BOOTSTRAPPING IN A MULTIPLICATIVE MODEL

5.1 Introduction

Double-logarithmic models are commonly used in economics, especially as production, utility and demand functions. The simplest form is the Cobb-Douglas model, which is a basic tool in economics, especially in production and consumer theories. [See e.g., Bodkin and Klein (1967).] Examples of higher order double-logarithmic models are the transcendental logarithmic (translog) and generalized Cobb-Douglas models. [See e.g., Guilkey, Lovell and Sickles (1983).] These types of models are used in production [see e.g., Fuss, McFadden and Mundlak (1978), and MaCurdy and Pencavel (1986)], utility [see e.g., Jackson (1984)] and demand functions [see e.g., Pollak and Wales (1980), and Lurano, Pierse and Richard (1986).]

These models are useful for situations in which factor shares are required to be constant. This type of situation is likely to prevail when both product and factor markets are assumed to be perfectly competitive. Moreover, by assuming a multiplicative lognormal disturbance term, an apparently nonlinear model may be transformed into a linear one and estimated by the least-squares procedure. Although there are several other nonlinear functions which are better suited to production functions, the multiplicative double-log model is the least complicated in terms of the estimation of its coefficients and their standard errors, and the construction of confidence intervals for these coefficients.

An important difficulty in a log linear model is the estimation of the constant term and its standard error. Let β be the coefficient to be estimated, and let $\hat{\beta}$ be an unbiased estimate of β. A nonlinear function $\hat{\beta}$, say $g(\hat{\beta})$, is generally a biased estimate of $g(\beta)$. [See e.g., Goldberger (1968).] In this chapter, attention is restricted to the case when $g(.)$ is an exponential function. One reason is that this functional form is frequently used in the econometric literature. Another reason is that an unbiased estimate of $\exp(\beta)$ is readily available, and this can be obtained by applying Finney's (1951) procedure.

Based upon the assumption that a double-log model has a multiplicative error term, the model can be linearized using logarithms. Although the model can be linearized by taking any form of logarithmic transformation, the natural logarithm is often used because of the common practice in assuming that the errors are identically and independently distributed with a lognormal probability distribution, having unit mean and finite variance. Let υ_t be one such error and σ_υ^2 be its finite variance. Then, it is well known that $E(\log \upsilon_t)=-\frac{1}{2}\log(1+\sigma_\upsilon^2)$. Consequently, when β_1 is the intercept term of the linearized model and when $\hat{\beta}_1$ is its OLS estimate, there will be a downward bias in $\hat{\beta}_1$. However, this bias can easily be corrected.

Let B be the constant term of a double-log model. A 'naive estimate', defined as $\exp(\hat{\beta}_1)$, will also be a biased estimate of B. Unlike the bias in $\hat{\beta}_1$, which is easy to correct, it is somewhat difficult to correct for the bias in $\exp(\hat{\beta}_1)$. Bradu and Mundlak (1970) [hereinafter, B-M] suggested a bias-correction procedure based upon a g-function, introduced by Finney (1951), which requires extensive tables and may yield unacceptable negative values for some values of its arguments (see Teeken and Koerts, 1972, p.804). Recently, Srivastava and Singh (1989) [hereinafter, S-S] proposed a bias-correction procedure which is simple to use and does not require the use of Finney's g-function. It is found in the Monte Carlo simulation study below, using B-M's original example of Israeli agriculture, that the B-M and S-S estimates are almost the same under the assumption of multiplicative lognormal errors. This is because the B-M estimate is a uniformly minimum variance unbiased estimate (UMVUE) of B and the S-S estimate has negligibly small bias (less than 0.05%). Both estimates were found to be better than the naive estimate, in terms of both bias and dispersion. However, for other examples, the S-S estimate can be biased and its bias can be large enough to be of some concern, especially when the error variance is large. In these situations, Efron's bootstrap can be successfully applied to obtain an unbiased estimate of B, using the S-S estimate. Alternatively, by applying a bias-correction formula given below, the S-S estimate can be modified to be almost UMVU. This latter method gives almost exact results, whereas, the bootstrap modification of the S-S estimate is only accurate to $O(n^{-1})$. When the bias is smaller than $O(n^{-1})$, the bootstrap modification of the S-S estimate may not outperform the original S-S estimate. In the Monte Carlo study below, it is found that both the B-M and S-S estimates have almost similar dispersions; these are both significantly smaller than the dispersion of the naive estimate.

In order to construct a confidence interval for the constant term, the distribution of its estimate needs to be known. Even under the assumption of lognormality, these distributions are difficult to obtain and are not available in the literature. For both the B-M and S-S estimates, unbiased estimates of their standard errors can easily be computed. One of the alternatives, then, is to apply asymptotic theory and use the normal or t-tables. The jackknife method has been considered by Chaubey and Singh (1988). Other methods have also been considered by Dhrymes (1962), Goldberger (1968), Heien (1968), Teeken and Koerts (1972) and Evans and Shaban (1976). Phillips (1984, 1985) shows that exact confidence intervals for finite samples can be obtained, provided that the distributions of the estimates are known. However, Phillips' method which is based upon extended rational approximants is rather cumbersome and requires the use of symbolic operators. These symbolic operators are computationally intensive and may sometimes yield estimates which admit errors of order $O(n^{-1/2})$. Moreover, when the distributions of the estimates are unknown, Phillips' approach is not applicable. On the other hand, the bootstrap, jackknife and other resampling techniques [see e.g., Efron (1979) and Wu (1986)], being distribution-free methods, can still be used. Only the bootstrap method will be considered here.

It is shown in Chapter 4, that naive bootstrap confidence intervals (BCI's), obtained by ordering the bootstrap estimates of the regression coefficients, are inferior compared to other BCI's. Moreover, it is demonstrated below that for the constant term of a multiplicative double-log model, the naive BCI is biased and should be avoided. One reason is that when either the B-M or S-S estimate is used to construct this BCI, the mean of the bootstrap estimates will be a biased estimate of the constant term. Secondly, the variance of the bootstrap estimates will be larger than the variance of either the B-M or S-S estimate, depending on which of the two estimates is subjected to bootstrapping. In what follows, two alternative BCI's are proposed and evaluated by Monte Carlo simulations. The BCI's are examined by comparing their empirical coverages with a pre-determined nominal value. The empirical coverage of a confidence interval is defined as the observed frequency that it contains the true value of the unknown parameter.

Section 5.2 introduces the model and establishes notation. Least-squares estimation of the model is discussed in Section 5.3. In Section 5.4, Monte Carlo simulation results for the various estimates are compared. In Section 5.5, bootstrap theory is developed for the B-M and S-S estimates, followed by theoretical developments of two alternative BCI's for both the B-M and S-S estimates of the constant term.

5.2 The Model

Let the model to be considered be

$$Y_t = B \left\{ \prod_{i=2}^{K} \left[Z_{ti}^{\beta_i} \right] \right\} \upsilon_t \qquad (t=1,\ldots,n) \tag{5.2.1}$$

in which Y_t is the t'th observation on the dependent variable, Z_{ti} is the corresponding observation on the i'th $(i=2,3,\ldots,K)$ non-stochastic independent variable, B and β_i $(i=2,3,\ldots,K)$ are the unknown coefficients to be estimated from the data, and υ_t is the t'th observation of the disturbance term. The disturbance term, υ_t, is required to satisfy the following assumptions:

(A.5.1) $\upsilon_t > 0$ and $E(\upsilon_t) = 1 \quad \forall\ t=1,\ldots,n.$

(A.5.2) $\text{Var}(\upsilon_t) = \sigma_\upsilon^2 \quad \forall\ t=1,\ldots,n.$

(A.5.3) $\text{Cov}(\upsilon_s,\upsilon_t) = 0 \quad \forall\ s\neq t.$

Although there are other distributions which satisfy all the above assumptions, it is customary (especially, among applied econometricians) to assume that the disturbances are lognormally distributed. Asssumption **A.5.1** is essential for most economic data to ensure that, for example, prices or quantities demanded (or produced) are non-negative.

When the disturbances are lognormal, the above model can be transformed by taking natural logarithms on both sides of (5.2.1) into

$$y_t = \beta_0 + \sum_{i=2}^{K}\left[\beta_i x_{ti}\right] + \varepsilon_t \quad (t=1,\ldots,n) \tag{5.2.2}$$

in which $y_t=\log(Y_t)$, $x_{ti}=\log(Z_{ti})$, $i=2,\ldots,K$, $\beta_0=\log(B)$ and $\varepsilon_t=\log(\upsilon_t)$. Note that (5.2.2) can also be written compactly as

$$y = X\beta + \varepsilon \tag{5.2.3}$$

in which $x_{t1}=1$ $(t=1,\ldots,n)$. The design matrix X is presumed to satisfy assumptions **(A.2.1)** and **(A.2.3)**, as stated in Chapter 2. Under the lognormality assumption of the disturbance term in (5.2.1 satisfies (see e.g., B-M),

(i) $E(\varepsilon_t) = -\frac{1}{2}\sigma^2,$

(ii) $\text{Var}(\varepsilon_t) = \sigma^2 \quad \forall\ t,$

(iii) $\text{Cov}(\varepsilon_s,\varepsilon_t) = 0 \quad \forall\ s\neq t,$

in which $\sigma^2=\log(1+\sigma_\upsilon^2)$. Let $\beta_1=\log(B)-\frac{1}{2}\sigma^2$ and $\omega_t=\varepsilon_t+\frac{1}{2}\sigma^2$. Then, the regression model

$$y = X\beta + \omega \tag{5.2.4}$$

becomes the linear regression model of Chapter 2, satisfying assumptions **(A.2.1)**, **(A.2.2)** and **(A.2.3)**. Later, the bootstrap theorems of Chapters 2 and 3 above are applied to (5.2.4).

5.3 Estimation of the Model

The OLS estimate of β in (5.2.4) is given by

$$\hat{\beta} = (X^TX)^{-1}X^Ty. \tag{5.3.1}$$

Note that $\hat{\beta}$ is BLUE but an estimate of B given by $\exp(\hat{\beta}_1)$ would be biased. Since $\beta_1 = \log(B) - \frac{1}{2}\sigma^2$, an estimate of B given by $\hat{B}_1 = \exp(\hat{\beta} + \frac{1}{2}\sigma^2)$ has B as its median and is a 'median-unbiased' estimate of B. [See e.g.,Goldberger (1968) and B-M.] An unbiased estimate of σ^2 is given by

$$s^2 = \left\{(\hat{W}^T\hat{W})/(n-k)\right\} \tag{5.3.2}$$

in which $\hat{W}$=My.

However, the expected value of $\hat{B}_1$ is not B. When the errors are lognormal, B-M showed that Finney's (1951) g-function can be used to obtain an unbiased estimate of B. This approach had also been adopted by Heien (1968) and Goldberger (1968).

Using Finney's g-function, an unbiased estimate of B can be obtained as

$$\hat{B}_2 = \left\{\exp(\hat{\beta}_1)\right\}\left\{g_m\left[\tfrac{1}{2}m^{-1}(m+1)(1-h)s^2\right]\right\}$$

in which

$$g_m(\tau) = \sum_{j=0}^{\infty} \frac{m^j(m+2j)\{m/(m+1)\}^j\{\tau^j/(j!)\}}{m(m+2)\ldots(m+2j)},$$

$$\tau = \left[\tfrac{1}{2}m^{-1}(m+1)(1-h)s^2\right],$$

$$m = (n-K),$$

$$h = a^T(X^TX)^{-1}a, \quad \text{and} \quad a^T = (1,0,\ldots,0).$$

B-M mistook $\hat{\beta}_1$ as an unbiased estimator of log(B) and, consequently, erroneously gave

$$\hat{B}_3 = \left\{\exp(\hat{\beta}_1)\right\}\left\{g_m\left[-\tfrac{1}{2}m^{-1}(m+1)hs^2\right]\right\}.$$

as an unbiased estimate of B. Note that $\hat{B}_2$ is the UMVUE of B and that $\hat{B}_3$ is biased downward.

The variance of B_2 is

$$\text{Var}(\hat{B}_2) = B^2\left[2\phi^{-1}\Phi(\sigma^2)-1\right] \tag{5.3.3}$$

in which

$$\Phi(\sigma^2) = \int_0^1 \left[v^r(1-v)^r\exp\left\{2\sigma^2\left[(v-\tfrac{1}{2})+h(1-v)\right]\right\}\right]dv,$$

$$\phi = \left[\left\{\Gamma\left[\tfrac{1}{2}(m+1)\right]^2\right\}\left[\Gamma(m+1)\right]^{-1}\right],$$

$$r = \left[\tfrac{1}{2}(m-3)\right].$$

B-M (p. 204) suggest estimating $\text{Var}(\hat{B}_2)$ with

$$\hat{Var}(\hat{B}_2) = \hat{B}_2^2 A_1^2$$

in which

$$A_1^2 = \left[1 - \left\{g_m\left[m^{-1}(m+1)(1-2h)s^2\right]\right\}\left\{g_m\left[\frac{1}{2}m^{-1}(m+1)(1-h)s^2\right]\right\}^{-2}\right].$$

The g-function is somewhat difficult to use and it may sometimes yield unacceptable negative values, especially for $\hat{B}_3$ when s^2 is large. To overcome this difficulty, several alternative (but biased) estimates have been proposed. A promising estimate is recently proposed by S-S and is given by

$$\hat{B}_4 = \exp\left[\hat{\beta}_1 + \frac{1}{2}(1-h)s^2\right].$$

It is shown by S-S that

$$E(\hat{B}_4) = \left\{\exp\left[\beta_1 + \frac{1}{2}h\sigma^2\right]\right\}\left[1 - m^{-1}(1-h)\sigma^2\right]^{-m/2} \tag{5.3.4}$$

and

$$\begin{aligned} MSE(\hat{B}_4) = \Bigg[\left[\exp(2\beta_1)\right]\Bigg\{\left[\exp(2h\sigma^2)\right]\left[1 - 2m^{-1}(1-h)\sigma^2\right]^{-m/2} \\ - 2\left\{\exp\left[\frac{1}{2}(1+h)\sigma^2\right]\right\}\left[1-m^{-1}(1-h)\sigma^2\right]^{-m/2} + \exp(\sigma^2)\Bigg\}\Bigg]. \end{aligned} \tag{5.3.5}$$

Since $MSE(\hat{B}_4)$ depends *inter alia* on β_1 and σ^2, which are unknown, S-S suggest

$$\widehat{MSE}(\hat{B}_4) = \left[\left\{\exp(2\hat{\beta}_1)\right\}\left\{\left[1-2m^{-1}(1-h)s^2\right]^{-m/2}\right.\right.$$

$$\left.\left. - 2\left\{\exp\left[\tfrac{1}{2}(1-3h)s^2\right]\right\}\left[1-m^{-1}(1-h)s^2\right]^{-m/2} + \exp\left[(1-2h)s^2\right]\right\}\right]$$

be used as an unbiased estimate of $MSE(\hat{B}_4)$.

From equation (5.3.3), the bias of $\hat{B}_4$ is given by

$$\text{Bias}(\hat{B}_4) = E(\hat{B}_4) - B = B\left\{1-\left[\exp(\tfrac{1}{2}h\sigma^2)\right]^{-1}\left[1-m^{-1}(1-h)s^2\right]^{-m/2}\right\}.$$

An estimate of this bias can be obtained by replacing B and σ^2 with $\hat{B}_4$ and s^2, respectively. This estimate is

$$\widehat{\text{Bias}}(\hat{B}_4) = (\hat{B}_4)\left\{1-\left[\exp(\tfrac{1}{2}h\sigma^2)\right]^{-1}\left[1-m^{-1}(1-h)s^2\right]^{-m/2}\right\}.$$

Note that $\widehat{\text{Bias}}(\hat{B}_4)$ is a biased estimate of $\widehat{\text{Bias}}(\hat{B}_4)$ because both $\hat{B}_4$ and $\exp(\frac{1}{2}hs^2)$ are biased estimates of $\hat{B}$ and $\exp(\frac{1}{2}h\sigma^2)$, respectively.

The variance of $\hat{B}_4$ is often needed for the construction of confidence intervals. By definition, this variance is given by

$$\text{Var}(\hat{B}_4) = MSE(\hat{B}_4) - \left[\text{Bias}(\hat{B}_4)\right]^2.$$

and an estimate of it can be obtained as

$$\widehat{\text{Var}}(\hat{B}_4) = \widehat{MSE}(\hat{B}_4) - \left[\widehat{\text{Bias}}(\hat{B}_4)\right]^2.$$

Wheareas $\widehat{MSE}(\hat{B}_4)$ is an unbiased estimate of $MSE(\hat{B}_4)$, $\widehat{\text{Bias}}(\hat{B}_4)$ is a biased estimate of $\text{Bias}(\hat{B}_4)$. Consequently, $\widehat{\text{Var}}(\hat{B}_4)$ would be a biased estimate of $\text{Var}(\hat{B}_4)$. This estimate can be improved by replacing $\widehat{\text{Bias}}(\hat{B}_4)$ by a better estimate of $\text{Bias}(\hat{B}_4)$. One such estimate of $\text{Bias}(\hat{B}_4)$ can be obtained by replacing B and $\exp(\frac{1}{2}h\sigma^2)$ with unbiased estimates.

When the naive estimate

$$\hat{B}_5 = \exp(\hat{\beta}_1)$$

was used to estimate B in their example, B-M estimated that the bias is nearly 14.91% of the true value. However, their estimate was based upon the erroneous assumption that $\hat{B}_3$ is an unbiased estimate of B. In the Monte Carlo simulation study below, when $\hat{B}_1$, $\hat{B}_3$ and $\hat{B}_5$ are used as estimates of B, their 'actual' biases were found to be approximately 14.64%, -7.05% and 5.17%, respectively. These values were obtained from Table 5.1 and adjusted for simulation errors. Note that the discrepancy between B-M's estimate of the bias of $\hat{B}_5$ and its 'actual' bias arises, because $\hat{B}_3$ was used in B-M's calculation instead of $\hat{B}_2$ and that $\hat{B}_3$ is biased downward. Thus, the actual bias in using $\hat{B}_5$ is less than B-M's estimate. The simulation results suggested that $\hat{B}_3$ should not be used in the first place.

When both β_1 and σ^2 are unknown but are replaced, respectively, by $\hat{\beta}_1$ and s^2 in (5.3.5), the resulting estimate is obtained:

$$M\tilde{S}E(\hat{B}_4) = \left[\left(\exp(2\hat{\beta}_1)\right)\left\{\left(\exp(2hs^2)\right)\left(1-2m^{-1}(1-h)s^2\right)^{-m/2} - 2\left\{\exp\left(\frac{1}{2}(1+h)s^2\right)\right\}\left(1-m^{-1}(1-h)s^2\right)^{-m/2} + \exp(s^2)\right\}\right].$$

This estimate will overestimate $MSE(\hat{B}_4)$. Monte Carlo simulation results below indicate that the mean of $M\tilde{S}E(\hat{B}_4)$ is almost twice $MSE(\hat{B}_4)$. Let

$$V\tilde{a}r(\hat{B}_2) = \hat{B}_2^2\left(2\phi^{-1}\Phi(s^2)-1\right)$$

in which the functions ϕ and Φ are as defined in (5.3.3).

Since $\mathrm{Var}(\hat{B}_2)$ is similar to $\mathrm{MSE}(\hat{B}_4)$, to the extent that both are functions of B and $\exp(\sigma^2)$, one would also expect that $\tilde{\mathrm{Var}}(\hat{B}_2)$ will also overestimate $\mathrm{Var}(\hat{B}_2)$. However, due to difficulties involving numerical integration, the estimate $\tilde{\mathrm{Var}}(\hat{B}_2)$ is not included in the simulation study below. This has implications for the bootstrap estimates of B, when either the B-M or S-S estimate is subjected to bootstrapping. These implications will be discussed in Section 5.5 below.

5.4 Small Sample Properties of the Estimates: Some Monte Carlo Results

All the simulations below are based upon the data given in B-M's example of Israeli agriculture. The sample size, n, is 10 and the number of exogeneous variables (including the intercept), K, is 4. A total of 800 trials is conducted and the parameter values are set at $B=36.237$, $\beta_2=2.840$, $\beta_3=-0.355$ and $\sigma^2=0.1384$. All of the above values correspond to the OLS estimates of the linearized model, using B-M's example.

Monte Carlo simulation results are reported in Tables 5.1 and 5.2. In Table 5.1, the means, standard errors and percentiles of the five estimates of B are tabulated. All results are adjusted for simulation error. To study the effects of a scale change in σ^2, the variance of ε was multiplied by a factor of five and the results are reported in Table 5.1(b). Finally, Table 5.2 reports the biases and confidence intervals of estimates of the variance of $\hat{B}_2$ and MSE of $\hat{B}_4$.

Pseudo-normal errors are first generated by IMSL's GGNML subroutine with an initial SEED value of 9999. These are then transformed to conform with (A.5.1), (A.5.2) and (A.5.3). In other words, let $(u_1, u_2, \ldots, u_n)$ be the N(0,1) errors generated by GGNML. The variables ε and V are then generated as

(i) $\varepsilon_t = u_t\sigma - \frac{1}{2}\sigma^2$,

(ii) $\upsilon_t = \exp(\varepsilon_t)$,

in which $t=1,2,\ldots,n$.

Given υ_t, the t'th observation on the dependent variable is reconstructed as

$$Y_t = B\left\{\prod_{i=2}^{K}\left[Z_{ti}^{\beta_i}\right]\right\}\upsilon_t, \quad (t=1,2,\ldots,n). \tag{5.4.1}$$

Taking natural logarithms on both sides of (5.4.1), and writing $\log(Y_t)=y_t$, $\log(Z_{t2})=x_{t2}$, $\log(Z_{t3})=x_{t3}$ and $\beta_1=\log(B)-\frac{1}{2}\sigma^2$,

$$y_i = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \omega_i \tag{5.4.2}$$

Equation (5.4.2) may then be written as

$$y = X\beta + \omega$$

in which X is a nx3 matrix comprising of unity in the first column and x_{ti} $(t=1,2,\ldots,n)$ in the i'th column for i=2 and 3. The least-squares estimates of B and σ^2 can then be obtained in the usual way.

Since exact confidence intervals for the various estimates are difficult to obtain, calibrated confidence bounds based upon the percentile method are obtained by Monte Carlo simulations. These are used as proxy measures of the corresponding exact confidence bounds. The calibrated confidence bounds are obtained for 800 trials. The corresponding values of the various estimates of B are reported in Table 5.1. The following remarks are in order. First, all distributions are heavily skewed to the right. Secondly, as the nominal bias of an estimate of B increases, the distribution of that estimate shifts away from the origin and its dispersion increases. Thirdly, as the nominal bias of an estimate increases, both its lower and upper confidence bounds shift to the right and away from the origin.

The results in Table 5.1 also indicate that among the biased estimates, the S-S estimate ($\hat{B}_4$) has the least bias and its standard error is almost identical to that of the UMVUE of B. They suggest that the S-S estimate is the best among the four alternative estimates considered. When the variance of ε is multiplied by a factor of 5, it is observed from Table 5.1(b) that, the biases of all estimates increase by more than five-fold, except for the UMVUE of B. The S-S estimate is still the best among the four alternative estimates, but its bias increased considerably from 0.02 to 0.512.

To complete this Monte Carlo simulation study, Table 5.2 reports the biases and confidence intervals of the estimates for the exact variance and MSE's of the B-M and S-S estimates. In this table, two other variables were also included. These are

(i) $\hat{A}_1 = \left[\widehat{\mathrm{Var}}(\hat{B}_2)\right]\left[\hat{B}_2\right]^{-2}$,

Table 5.1: Means, Standard Errors and Percentiles of Several Estimates of B[1]

a) $\sigma^2 = 0.138$

Estimate	Bias	Std. Error	Percentiles 2.5	5.0	10.0	90.0	95.0	97.5
$\hat{B}_1$	5.304	22.91	13.85	15.68	19.00	67.54	80.39	102.57
$\hat{B}_2$	0.000	19.28	11.82	14.19	16.71	60.16	71.19	83.75
$\hat{B}_3$	-2.555	17.78	11.01	13.01	15.45	55.86	66.20	76.39
$\hat{B}_4$	0.020[2]	19.30	11.83	14.22	16.73	60.17	71.35	83.91
$\hat{B}_5$	1.873	20.80	12.82	14.77	17.75	63.73	75.81	91.01

b) $\sigma^2 = 0.690$

Estimate	Bias	Std. Error	Percentiles 2.5	5.0	10.0	90.0	95.0	97.5
$\hat{B}_1$	38.95	136.6	5.77	7.23	10.13	158.9	221.6	409.3
$\hat{B}_2$	0.00	50.4	3.80	4.66	5.96	77.3	114.3	154.0
$\hat{B}_3$	-11.18	33.3	2.81	3.43	4.36	53.9	80.4	112.3
$\hat{B}_4$	0.54[3]	51.5	3.80	4.71	6.12	79.1	115.0	157.9
$\hat{B}_5$	13.81	77.1	4.63	4.69	7.69	106.1	155.8	233.7

1. All results are based upon 800 trials and are adjusted for simulation error. The exact value of B is 36.237.

2. Actual bias of $\hat{B}_4$ is 0.018 when $\sigma^2 = 0.138$.

3. Actual bias of $\hat{B}_4$ is 0.430 when $\sigma^2 = 0.690$.

Table 5.2: Biases and Percentiles of the Estimates of $\mathrm{Var}(\hat{B}_2)$ and $\mathrm{MSE}(\hat{B}_4)$[1]

	Bias	Percentiles					
Estimate	(% Bias)	2.5	5.0	10.0	90.0	95.0	97.5
$\widehat{\mathrm{Var}}(\hat{B}_2)$	44.78 (12.6%)	22.75	30.68	47.17	1840.1	1270.2	849.9
$\widehat{\mathrm{MSE}}(\hat{B}_4)$	19.63 (5.5%)	22.12	29.82	45.94	1701.1	1181.7	806.5
$\widetilde{\mathrm{MSE}}(\hat{B}_4)$	443.30 (124%)	28.94	40.76	64.25	5053.4	2700 4	1607.2
$\hat{A}_1$	0.0068 (2.91%)	0.061	0.073	0.102	0.511	0.449	0.401
$\hat{A}_2$	0.0007 (0.31%)	0.060	0.072	0.100	0.463	0.411	0.371

1. All results are based upon 800 trials and all unknown true values are replaced by the corresponding Monte Carlo estimates. Exact values of B and σ^2 are 36.237 and 0.138, respectively. On the other hand, the Monte Carlo estimates of $\mathrm{Var}(\hat{B}_2)$, $\mathrm{MSE}(\hat{B}_4)$, A_1 and A_2 are 355.22, 356.80, 0.2388 and 0.2267, respectively.

(ii) $\hat{A}_2 = \left\{\hat{MSE}(\hat{B}_4) - \left[\hat{Bias}(\hat{B}_4)\right]^2\right\}\left[\hat{B}_4\right]^{-2}$.

The corresponding true values of $\hat{A}_1$ and $\hat{A}_2$ are

$$A_1 = \left[Var(\hat{B}_2)\right]\left[\hat{B}_2\right]^{-2}$$

and

$$A_2 = \left\{MSE((\hat{B}_4) - \left[\hat{Bias}(\hat{B}_4)\right]^2\right\}\left[\hat{B}_4\right]^{-2},$$

respectively.

The results in Table 5.2 show that the bias in $\hat{MSE}(\hat{B}_4)$ is less than the bias in $\hat{Var}(\hat{B}_2)$. In particular, $\hat{MSE}(\hat{B}_4)$ has a bias of 5.5% as compared to a bias of 12.6% in $\hat{Var}(\hat{B}_4)$. When $\hat{B}_2$ is used as an estimate of B, both $\hat{B}_2$ and $\hat{Var}(\hat{B}_2)$ are needed for constructing confidence intervals of B. On the other hand, when $\hat{B}_2$ is replaced by $\hat{B}_4$ as an estimate of B, both $\hat{B}_4$ and $\hat{MSE}(\hat{B}_4)$ will be needed for constructing these intervals. Consequently, the results seem to suggest that confidence intervals based upon $\hat{B}_4$ may have better empirical coverages than those intervals based upon $\hat{B}_2$.

It will shown in Section 5.5, that the MSE of bootstrap estimates of B based upon $\hat{B}_4$ converges almost surely to $\tilde{MSE}(\hat{B}_4)$, and that the variance of bootstrap estimates of B based upon $\hat{B}_2$ converges almost surely to $\tilde{Var}(\hat{B}_2)$, as J goes to infinity. The results in Table 5.2 indicate that the mean of $\tilde{MSE}(\hat{B}_4)$ is more than twice the actual MSE of $\hat{B}_4$. Although no result is obtained for $\tilde{Var}(\hat{B}_2)$, because of difficulties relating to numerical integration, one would expect the mean of $\tilde{Var}(\hat{B}_4)$ to be about twice the actual variance of $\hat{B}_2$. This is because $Var(\hat{B}_2)$ is very similar to $MSE(\hat{B}_4)$, to the extent that both are functions of B and $\exp(\sigma^2)$. Consequently, the simulation results suggest that naive bootstrap confidence intervals should not be used for constructing confidence intervals of B. This is irrespective of whether the bootstrap estimates of B are based upon $\hat{B}_2$ or $\hat{B}_4$. The resulting confidence intervals will be too wide and in most cases, the length of these intervals will be at least twice the length of the exact confidence interval of B. Consequently, the empirical coverages of naive bootstrap confidence intervals will be larger than their nominal coverages.

Better bootstrap confidence intervals of B can be constructed by obtaining the bootstrap distribution of $\hat{A}_1$, when $\hat{B}_2$ is used as an estimate of B. When $\hat{B}_4$ is used as an estimate of B, these intervals can be constructed by obtaining the bootstrap distribution of $\hat{A}_2$. From Table 5.2, $\hat{A}_1$ has a bias of 2.9% whereas $\hat{A}_2$ has a bias of only 0.3%. Thus, a confidence interval of B obtained by using the bootstrap distribution of $\hat{A}_2$ would be more reliable than a similar confidence interval based upon $\hat{A}_1$.

5.5 Bootstrap Distributions of $\hat{B}_2$ and $\hat{B}_4$

In Section 5.2, the n components of the disturbance vector $\upsilon^T=(\upsilon_1,\ldots,\upsilon_n)$ are assumed to be i.i.d. with unknown distribution function, F, having mean unity and variance σ_υ^2. For any random sample of υ, the sample mean and variance are not necessarily unity and σ_υ^2, respectively. However, these sample measures are unbiased estimates of their respective population quantities.

The OLS residuals of (5.2.4), $\hat{\omega}=M\varepsilon$ has sample mean zero and $n^{-1}E(\hat{\omega}^T\hat{\omega})=\sigma^2\left[n^{-1}(n-K)\right]$. Let $t=1,2,\dots,n$ and let $\hat{\upsilon}_t=\exp(\hat{\varepsilon}_t)$, in which

$$\hat{\varepsilon}_t = \hat{\omega}_t\left[(n-K)^{-1}n\right]^{1/2} - \tfrac{1}{2}s^2 \tag{5.5.1}$$

and s^2 is as defined in equation (5.3.2). Then, it can be shown that $\hat{\upsilon}_t$ is always positive, provided that $|\hat{\varepsilon}_t|<\infty$. Also, it can be shown that $E\left\{n^{-1}\sum_{t=1}^{n}\left[\hat{\upsilon}_t\right]\right\} = 1$ and $E\left\{n^{-1}\sum_{t=1}^{n}\left[\hat{\upsilon}_t-1\right]^2\right\} = \sigma^2_\upsilon$. The objective of the above transformations is to obtain a sample of random values that closely resembles a sample of lognormal errors with mean $(-\frac{1}{2}\sigma^2)$ and variance σ^2. This is possible, provided that the underlying distribution function of the true errors is in fact lognormal.

Given $\hat{\beta}$ and $\hat{\varepsilon}$ of (5.3.1) and (5.5.1), respectively, the bootstrap responses can be obtained by first bootstrapping $\hat{\varepsilon}$ and then reconstructing (5.2.3) as

$$y^*_{(j)} = X\tilde{\beta} + \varepsilon^*_{(j)} \tag{5.5.2}$$

in which $\varepsilon^*_{(j)} = S_{(j)}\hat{\varepsilon}$ and $\tilde{\beta}^T=\left[(\hat{\beta}_1+\frac{1}{2}s^2),\hat{\beta}_2,\dots,\hat{\beta}_k\right]$, for each bootstrap replication $(j=1,\dots,J)$. The estimate $\tilde{\beta}$ is used instead of β, because β is unknown and $\tilde{\beta}$ is an unbiased estimate of β.

The OLSE's of $\tilde{\beta}$ and s^2 in (5.5.2) are the bootstrap estimates $\beta^*_{(j)}$ and s^{*2}, respectively. These bootstrap quantities are obtained as

$$\beta^*_{(j)} = (X^TX)^{-1}X^Ty^*_{(j)} \tag{5.5.3}$$

and

$$s^{*2}_{(j)} = \left\{(n-K)^{-1}\left[\omega^{*T}_{(j)}\omega^{*}_{(j)}\right]\right\},$$

in which $\omega^{*}_{(j)}=M\varepsilon^{*}_{(j)}$. Note that $s^{*2}_{(j)}$ is a bootstrap estimate of σ^2 and, besides it, another bootstrap estimate of σ^2 is

$$s^{*2}_{2(j)} = \left\{(n-1)^{-1}\left[(\varepsilon^{*}_{(j)} - \bar{\varepsilon}^{*}_{(j)})^{T}(\varepsilon^{*}_{(j)} - \bar{\varepsilon}^{*}_{(j)})\right]\right\}$$

where

$$\bar{\varepsilon}^{*}_{(j)} = n^{-1}\left\{\sum_{t=1}^{n}\left[\varepsilon^{*}_{(j)t}\right]\right\}.$$

Theorem 5.1 and 5.2 below show that both $s^{*2}_{(j)}$ and $s^{*2}_{2(j)}$ are unbiased estimates of σ^2. Thus, either $s^{*2}_{(j)}$ or $s^{*2}_{2(j)}$ can be used without affecting the bootstrap results. The following lemma states another useful property of the selection matrix, $S_{(j)}$. Lemma 5.1 below will be used to prove Theorem 5.1.

Lemma 5.1: Let $S_{(j)}$ be a nxn selection matrix and let $A=\left[a_{hi}\right]$ be a real nxn matrix of finite elements. Then, for n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}AS_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{n^{-1}\sum_{h=1}^{n}\left[a_{hh}\right]\right\}I_n + \left\{n^{-1}\sum_{i\neq h}^{n}\sum^{n}\left[a_{hi}\right]\right\}\left[E(n,n)\right].$$

Proof:

Let $$S^{T}_{(j)} = \left[S^{T}_{(j).1},\ldots,S^{T}_{(j).n}\right].$$ Then,

$$S^{T}_{(j)}AS_{(j)} = \sum_{h=1}^{n}\sum_{i=1}^{n}\left\{S^{T}_{(j).h}a_{hi}S_{(j)i.}\right\},$$

and $$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}AS_{(j)}\right]\right\} = \left[\sum_{i=1}^{n}\sum_{h=1}^{n}a_{hi}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j).h}S_{(j)i.}\right]\right\}\right].$$

From Lemma 2.2

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j).h}S_{(j)i.}\right]\right\} \xrightarrow[J]{a.s.} \left[n^{-1}I_{n}\right]$$

when i=h. Also, it is straightforward from the proof of Lemma 2.4 that when i is different from h,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j).h}S_{(j)i.}\right]\right\} \xrightarrow[J]{a.s.} \left\{n^{-2}\left[E(n,n)\right]\right\}.$$

The remainder of the proof is then straightforward. *Q.E.D.*

Theorem 5.1: Let n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[s^{*2}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} s^{2}.$$

Proof:

Note that $s^{*2}_{(j)}$ can be written as

$$s^{*2}_{(j)} = \left\{(n-K)^{-1}\left[\varepsilon^{*T}_{(j)}M\varepsilon^{*}_{(j)}\right]\right\} = \left\{(n-K)^{-1}\left[\hat{\varepsilon}^{T}S^{T}_{(j)}MS_{(j)}\hat{\varepsilon}\right]\right\}.$$

Thus,

$$(n-K)\left\{\frac{1}{J}\sum_{j=1}^{J}\left[s^{*2}_{(j)}\right]\right\} = \hat{\varepsilon}^{T}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}MS_{(j)}\right]\right\}\hat{\varepsilon}.$$

Upon application of Lemma 5.1,

$$\left[(n-K)\left\{\frac{1}{J}\sum_{j=1}^{J}\left[s^{*2}_{(j)}\right]\right\}\right] \xrightarrow[J]{a.s.} \left[\left\{n^{-1}\left[\hat{\varepsilon}^{T}\hat{\varepsilon}\right] - \bar{\hat{\varepsilon}}^{2}\right\}\left[tr(M)\right] + \left[\bar{\hat{\varepsilon}}^{2}\right]\left\{\sum_{h,i}^{n}\left[m_{hi}\right]\right\}\right].$$

Note that $\bar{\varepsilon}^2 = -\frac{1}{2}s^2$ and $\left[\hat{\varepsilon}^T\hat{\varepsilon}\right] = \left\{\left[n(n-K)^{-1}\right]\left[\hat{w}^T\hat{w}\right] - \left[\frac{1}{2}ns^2\right]\right\}$. Further note that the linearized model always has an intercept term. From least-squares theory, it is known that when the regression model has an intercept, the mean of the OLS residuals will be zero. This implies that

$$\left\{\sum_{i=1}^{n}\left[\hat{w}_i\right]\right\} = 0 \quad \text{or} \quad \left\{\sum_{i}^{n}\sum_{j}^{n}\left[m_{ij}w_j\right]\right\} = 0. \quad \text{Since } E(w_j) = -\frac{1}{2}\sigma^2,$$

$$E\left\{\sum_{i=1}^{n}\left[\hat{w}_i\right]\right\} = E\left\{\sum_{i}^{n}\sum_{j}^{n}\left[m_{ij}w_j\right]\right\} = \left[-\frac{1}{2}\sigma^2\right]\left\{\sum_{i}^{n}\sum_{j}^{n}\left[m_{ij}\right]\right\} = 0.$$

This requires that $\left\{\sum_{i}^{n}\sum_{j}^{n}\left[m_{ij}\right]\right\} = 0.$

The remainder of the proof is straightforward upon substitution of the appropriate terms and noting that $tr(M)=(n-K)$ and $\left\{(n-K)^{-1}\left[\hat{w}^T\hat{w}\right]\right\}=s^2$.

Q.E.D.

Theorem 5.2: For n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[s^{*2}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} s^2.$$

Proof:

Note that this theorem is a slight variant of Theorem 2.3 above. Further, note that

$$s^{*2}_{2(j)} = \left\{(n-1)^{-1}\left[\varepsilon^{*T}_{(j)}\varepsilon^{*}_{(j)} - n\bar{\varepsilon}^{*2}_{(j)}\right]\right\}$$

in which $\bar{\varepsilon}^{*}_{(j)} \equiv \left\{n^{-1}\sum_{i=1}^{n}\left[\varepsilon^{*}_{(j)}\right]\right\}$.

Upon application of Lemma 2.2,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\varepsilon^{*T}_{(j)}\varepsilon^{*}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \hat{\varepsilon}^{T}\hat{\varepsilon} = n\left[s^{2} - \frac{1}{2}s^{2}\right].$$

Also, it is straightforward to show that

$$\bar{\varepsilon}^{*2}_{(j)} = \left\{n^{-2}\left[\varepsilon^{*T}_{(j)}\varepsilon^{*}_{(j)}\right]\right\} + n^{-2}\left\{\sum_{i\neq h}^{n}\sum^{n}\left[S_{(j)h.}\hat{\varepsilon}\hat{\varepsilon}^{T}S^{T}_{(j).i}\right]\right\}$$

and, from Theorem 2.4,

$$\left\{\frac{1}{J}\sum_{i\neq h}^{n}\sum^{n}\left[S_{(j)h.}\hat{\varepsilon}\hat{\varepsilon}^{T}S^{T}_{(j).i}\right]\right\} \xrightarrow[J]{a.s.} \left[-\frac{1}{2}s^{2}\right].$$

Upon substitution of the proper terms,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\bar{\varepsilon}^{*2}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} n^{-2}\left\{n\left[s^{2} - \frac{1}{2}s^{2}\right]\right\} + n^{-2}\left\{n\left[n^{-1}\right]\left[-\frac{1}{2}s^{2}\right]\right\}.$$

The R.H.S. simplifies to $\left[n^{-1}s^{2}-\frac{1}{2}s^{2}\right]$ and the remainder of the proof is then straightforward. *Q.E.D.*

The next step is to replace $\hat{\beta}$ and s^2 by $\beta^{*}_{(j)}$ and $s^{*2}_{(j)}$ (or $s^{*2}_{2(j)}$), respectively, in order to obtain the bootstrap analogues of the B-M and S-S estimates. These bootstrap estimates are

$$B^{*}_{2(j)} = \left\{\exp\left[\beta^{*}_{(j)}\right]\right\}\left\{g_{m}\left[\frac{1}{2}m^{-1}(m+1)(1-h)s^{*2}_{(j)}\right]\right\}$$

and

$$B^{*}_{4(j)} = \exp\left\{\beta^{*}_{1(j)} + \left[\frac{1}{2}(1-h)s^{*2}_{(j)}\right]\right\}.$$

Let the sample variance of $B^{*}_{2(j)}$ and sample MSE of $B^{*}_{4(j)}$ be defined as

$$\hat{\mathrm{Var}}\left[B_2^*\right] = \left\{(J-1)^{-1}\sum_{j=1}^{J}\left[B_{2(j)}^* - \bar{B}_{2(j)}^*\right]^2\right\}$$

in which $\bar{B}_2^* \equiv \left\{\frac{1}{J}\sum_{j=1}^{J}\left[B_{2(j)}^*\right]\right\}$, and

$$\hat{\mathrm{MSE}}\left[B_4^*\right] = \left\{\frac{1}{J}\sum_{j=1}^{J}\left[B_{4(j)}^* - \exp(\hat{\beta}_1)\right]^2\right\}.$$

Also, let

$$\tilde{\mathrm{Var}}(B_{2(j)}^*) = \left[B_{2(j)}^*\right]^2 A_{1(j)}^{*2}$$

and

$$\tilde{\mathrm{MSE}}(B_{4(j)}^*) = \left[\left\{\exp(2\beta_{1(j)}^*)\right\}\left\{1-\left[2m^{-1}(1-h)s_{(j)}^{*2}\right]\right\}^{-m/2}\right.$$

$$\left. - 2\left\{\exp\left[\frac{1}{2}(1-3h)s_{(j)}^{*2}\right]\right\}\left[1-m^{-1}(1-h)s_{(j)}^{*2}\right]^{-m/2} + \exp\left\{(1-2h)s_{(j)}^{*2}\right\}\right],$$

in which $A_{1(j)}^{*2}$ is obtained by replacing s^2 with $s_{(j)}^{*2}$ in A_1^2.

Theorem 5.3 below concerns the mean of the bootstrap estimates of β. It is used to prove Theorems 5.4 and 5.5.

Theorem 5.3: Let $\tilde{\beta}_1 = \left[\hat{\beta}+\frac{1}{2}s^2\right]$ and let n be finite. Then,

(i) $\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta_{1(j)}^*\right]\right\} \xrightarrow[J]{a.s.} \hat{\beta}_1$

(ii) $\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta_{1(j)}^* + \frac{1}{2}s^2\right]\right\} \xrightarrow[J]{a.s.} \hat{\beta}_1 + \frac{1}{2}s^2$

and

$$\text{(iii)}\quad E\left[\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{1(j)}+\frac{1}{2}s^{*2}_{(j)}\right]\right\}\right] \xrightarrow[J]{a.s.} \beta_1.$$

Proof:

First, note that $\beta^{*}_{(j)}$ can be written as

$$\beta^{*}_{1(j)} = \tilde{\beta}_1 + a^T(X^TX)^{-1}X^T\varepsilon^{*}_{(j)}.$$

Upon application of Theorem 2.1,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\varepsilon^{*}_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\left[E(n,1)\right]\left[-(\tfrac{1}{2}s^2)\right]\right\}.$$

Since

$$a^T(X^TX)^{-1}X^T\left[E(n,1)\right] = 1,$$

the first part of the theorem is then rather straightforward.

The proof for part (ii) follows directly from part (i) and Theorem 5.1.

From Section 5.3, it is known that $E(\hat{\beta}_1) = \left[\beta_1 - \frac{1}{2}\sigma^2\right]$. The remainder of the proof for part (iii) is then straightforward from part (ii). *Q.E.D.*

Theorem 5.4 below concerns the relation between the mean of $\tilde{\mathrm{Var}}\left[B^{*}_{2(j)}\right]$, $(j=1,\ldots,J)$, and $\mathrm{Var}\left[B^{*}_{2}\right]$. On the other hand, Theorem 5.5 concerns the relation between the mean of $\tilde{\mathrm{MSE}}\left[B^{*}_{4(j)}\right]$, $(j=1,\ldots,J)$, and $\mathrm{MSE}\left[B^{*}_{4}\right]$. Both Theorems 5.4 and 5.5 rely upon the results of Theorems 5.1, 5.2 and 5.3.

Theorem 5.4: Let n be finite. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\tilde{\mathrm{Var}}(B^{*}_{2(j)})\right]\right\} \xrightarrow[J]{a.s.} \mathrm{Var}(B^{*}_{2})$$

where

$$\mathrm{Var}(B^{*}_{2}) \equiv \left[\lim_{J\to\infty}\left\{(J-1)^{-1}\sum_{j=1}^{J}\left[B^{*}_{2(j)} - \bar{B}^{*}_{2(j)}\right]^{2}\right\}\right].$$

Proof:

Note that $\mathrm{Var}(B^{*}_{2})$ and $\tilde{\mathrm{Var}}(B^{*}_{2(j)})$ are the bootstrap analogues of $\mathrm{Var}(\hat{B}_2)$ and $\widehat{\mathrm{Var}}(\hat{B}_2)$, respectively. Since $\widehat{\mathrm{Var}}(\hat{B}_2)$ is an unbiased estimate of $\mathrm{Var}(\hat{B}_2)$, it is not surprising that

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\tilde{\mathrm{Var}}(B^{*}_{2(j)})\right]\right\} \xrightarrow[J]{a.s.} \mathrm{Var}(B^{*}_{2}).$$

Nevertheless, the proof is not complete without the applications of Theorems 5.1 (or 5.2) and 5.3. *Q.E.D.*

Theorem 5.5: For n finite,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\tilde{\mathrm{MSE}}(B^{*}_{4(j)})\right]\right\} \xrightarrow[J]{a.s.} \mathrm{MSE}(B^{*}_{4})$$

where

$$\mathrm{MSE}\left[B^{*}_{4}\right] \equiv \left[\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[B^{*}_{4(j)} - \exp(\hat{\beta}_1)\right]^{2}\right\}\right].$$

Proof:

The proof is rather straightforward upon applications of Theorems 5.1 (or 5.2) and 5.3. It is also based upon the fact that $\mathrm{MSE}(B^{*}_{4})$ and $\tilde{\mathrm{MSE}}(B^{*}_{4(j)})$ are the bootstrap analogues of $\mathrm{MSE}(\hat{B}_4)$ and $\widehat{\mathrm{MSE}}(\hat{B}_4)$, respectively, and that $\widehat{\mathrm{MSE}}(\hat{B}_4)$ is an unbiased estimate of $\mathrm{MSE}(\hat{B}_4)$. *Q.E.D.*

When $\hat{\beta}^T = \left[(\hat{\beta}_1+\frac{1}{2}s^2), \hat{\beta}_2, \ldots, \hat{\beta}_k\right]$, it can be shown that

$$\mathrm{Var}(\beta_2^*) = \left\{\exp\left[2\hat{\beta}_1+s^2\right]\right\}\left[2\phi^{-1}\Phi(s^2)-1\right]$$

in which the functions ϕ and Φ are defined in (5.3.3).

It can also be shown that

$$\mathrm{MSE}(B^*_{4(j)}) = \left\{\exp\left[2\hat{\beta}_1+s^2\right]\right\}\left\{\left[\exp(2hs^2)\right]\left[1-2m^{-1}(1-h)s^2\right]^{-m/2}\right.$$
$$\left. - 2\exp\left[\frac{1}{2}(1+h)s^2\right]\left[1-m^{-1}(1-h)s^2\right]^{-m/2} + \exp(s^2)\right\}.$$

Note that $\mathrm{MSE}(B_4^*)$ is similar to $\tilde{\mathrm{MSE}}(\hat{B}_4)$, and Monte Carlo results in Table 5.5 indicate that the latter estimate is a biased estimate of $\mathrm{MSE}(\hat{B}_4)$. This implies that $\widehat{\mathrm{MSE}}(B_4^*)$ is a biased estimate of $\mathrm{MSE}(\hat{B}_4)$. Since $\tilde{\mathrm{Var}}(\hat{B}_2)$ and $\tilde{\mathrm{MSE}}(\hat{B}_4)$ have similar characteristics, this also implies that $\widehat{\mathrm{Var}}(\hat{B}_2)$ is a biased estimate of $\mathrm{Var}(\hat{B}_2)$. Thus, bootstrap confidence intervals obtained from ordered samples of either $B^*_{2(j)}$ or $B^*_{4(j)}$ are biased. The reason for this is that both $\mathrm{Var}(B_2^*)$ and $\mathrm{MSE}(B_4^*)$ contain the expressions $\exp(2\hat{\beta}_1+s^2)$ and $\exp(s^2)$, which happen to be biased estimates of B and $\exp(\sigma^2)$, respectively.

One alternative is to replace $\exp(2\hat{\beta}_1+s^2)$ and $\exp(s^2)$ by unbiased estimates of B and $\exp(\sigma^2)$. Although an unbiased estimate of B can be obtained, it is difficult to obtain an unbiased estimate for $\exp(\sigma^2)$. The second alternative is to define

$$t^{*}_{1(j)} = \left[B^{*}_{2(j)} - \bar{B}^{*}_{2}\right]\left[\hat{Var}(B^{*}_{2})\right]^{-1/2}$$

and $t^{*}_{1}(a)$ and $t^{*}_{1}(1-a)$ be the a'th and (1-a)'th percentiles of the ordered sample of $t^{*}_{1(j)}$, $j=1,2,\ldots,J$. The bootstrap confidence interval of B is then constructed as

$$B^{*}_{2}(z) = \hat{B}_{2} + t^{*}_{1}(z)\left[\hat{Var}(\hat{B}_{2})\right]^{1/2}, \quad z = a, (1-a).$$

In Chapter 4, it is shown that a BCI based upon t^{*}_{1} is accurate to $O(n^{-1/2})$. Confidence intervals obtained in this form would be acceptable for large samples when the difference between $O(n^{-1/2})$ and $O(n^{-1})$ is negligible. For small samples, the accuracy of these confidence intervals can be improved by constructing

$$t^{*}_{2(j)} = \left[B^{*}_{2(j)} - B^{*}_{2}\right]\left[\tilde{Var}(B^{*}_{2(j)})\right]^{-1/2}.$$

The bootstrap confidence interval associated with $t^{*}_{2(j)}$ can be obtained as

$$B^{*}_{2}(z) = \hat{B}_{2}\left\{1-\hat{A}_{1}\left[t^{*}_{2}(z)\right]\right\}^{-1}, \qquad z = a, (1-a).$$

This is because

$$\left[\hat{Var}(\hat{B}_{2})\right]^{1/2} = (\hat{B}_{2})(\hat{A}_{1}).$$

Nevertheless, t^{*}_{2} is applicable, provided that $\left\{1-\hat{A}_{1}\left[t^{*}_{2}(1-a)\right]\right\}$ is restricted to be positive. It can be shown that this restriction always holds for the lower bound. However, it does not always apply to the upper bound and this leads to the possibility of an open upper bound.

Monte Carlo estimates of the empirical coverages of conventional and bootstrap confidence intervals are reported in Table 5.3. The results for the $\hat{B}_2$ and $\hat{B}_4$ are somewhat similar. The simulation results indicate that BCI's based upon $t^*_{2(j)}$ have the best performance in terms of its empirical coverage. In Table 5.3, BCI2 represents one such BCI. The empirical coverages of BCI's based upon $t^*_{1(j)}$, and conventional confidence intervals based upon Student's t-distribution, are not well-balanced. Both coverages are smaller than the nominal coverage for the lower bound but are greater than the nominal coverage for the upper bound. These two confidence intervals are reported in Table 5.3 as BCI1 and CCI1, respectively. Two confidence intervals are also constructed by taking into account the fact that variance of either $\hat{B}_2$ or $\hat{B}_4$ is dependent upon B. They are represented by BCI2 and CCI2, and their coverages are also not balanced. This time, however, the coverage of CCI2 is greater than the nominal coverage for the lower bound but smaller than the nominal coverage for the upper bound. On the other hand, the coverage of BCI2 tends to be more balanced, when compared to the coverage of either BCI1 or CCI2. Thus, the results suggest that one should use BCI2.

5.6 Conclusion

It is shown in this chapter, that Bradu and Mundlak (1970) mistakenly give a wrong estimator for log(B), where B is the constant term of a multiplicative model. Consequently, they erroneously give an incorrect estimator for B. This estimator is shown to be biased downward.

The focus of this chapter is on the estimation of B. Both the correct Bradu-Mundlak's estimate of B and a similar estimate given by

Table 5.3: Empirical Significance Levels of Bootstrap and Conventional Confidence Intervals[1]

Confidence Interval		True	B-M Estimate	S-S Estimate
CCI1	Lower	2.5	0.4	0.4
		5.0	0.8	0.8
		10.0	1.8	1.9
	Upper	2.5	12.3	12.8
		5.0	18.3	18.6
		10.0	24.1	24.5
CCI2	Lower	2.5	4.3	4.3
		5.0	7.0	7.3
		10.0	9.9	10.4
	Upper	2.5	0.4	0.5
		5.0	1.4	1.4
		10.0	5.8	6.0
BCI1	Lower	2.5	2.5	2.6
		5.0	3.6	3.8
		10.0	5.4	5.6
	Upper	2.5	11.6	13.3
		5.0	17.5	19.1
		10.0	26.0	27.5
BCI2	Lower	2.5	2.0	1.9
		5.0	3.3	2.9
		10.0	7.0	6.3
	Upper	2.5	1.1	1.5
		5.0	2.4	2.5
		10.0	5.0	5.1

1. All results are based upon 800 trials and 200 bootstrap replications. Exact values of B and σ^2 are 36.237 and 0.138, respectively.

Srivastava-Singh are re-examined for bias, and confidence intervals have also been constructed. These estimates are denoted by $\hat{B}_2$ and $\hat{B}_4$, respectively. The distributions of both estimates are found to be heavily skewed to the right. It is also found that, when the variance of the error term becomes larger, the bias of $\hat{B}_4$ increases and its distribution shifts to the right.

It is also shown in this chapter, that the MSE of bootstrap estimates of B overestimates the actual MSE $\hat{B}_4$, when the bootstrap estimates are obtained in a similar fashion as $\hat{B}_4$. Similarly, the variance of bootstrap estimates of B overestimates the true variance of $\hat{B}_2$, when the bootstrap estimates are obtained in a similar fashion as $\hat{B}_2$. Consequently, bootstrap confidence intervals obtained by ordering the bootstrap estimates of B are biased and should not be used. This is true when the bootstrap estimates are obtained in a similar fashion as either $\hat{B}_2$ or $\hat{B}_4$. In general, the lengths of these bootstrap intervals will be longer than the length of the exact confidence interval.

Alternative BCI's are proposed and the one based upon the bootstrap t-distribution is found to have the best empirical coverage. It is the only confidence interval whose coverage is relatively balanced. However, it does have its own problems. One such problem is the possibility of obtaining an open upper bound for the true value.

The assumption of lognormality is crucial in both estimates. In this chapter, the bootstrap method is only used for constructing confidence intervals. However, the role of the bootstrap can be enhanced. When the lognormality assumption is violated, both estimates will no longer be reliable and better estimates need to be constructed.

Two feasible alternatives are the jackknife and bootstrap estimates considered in Chaubey and Sim (1988).

In a related area, Rukhin(1986) showed that when the error variance exceeds one, the MSE of the B-M estimate can be significantly reduced by using a Bayesian estimate. However, the results of this chapter will not be significantly affected by not incorporating the Bayesian estimate, because the error variance had been chosen to be less than unity. Nevertheless, future research may incorporate cases when the error variance exceeds one, and it would be interesting to compare the bootstrap confidence intervals of the modified B-M estimate with those of the original B-M estimate.

CHAPTER SIX

BOOTSTRAPPING THE PARAMETER OF AN AR(1) PROCESS

6.1 Introduction

Autoregressive models and the more general ARIMA models have become very useful tools in applied econometrics. These models have many applications, one of which is an application to test whether an observed time series is a random walk. In such an application, one is in fact testing for unit roots. This application itself has become an important topic in time series analysis. For a simple AR(1) process, an estimate ($\hat{\beta}$) of the AR(1) parameter (β) is first obtained and then tested for unity. [See e.g., Evans and Savin (1981, 1984), and Phillips and Perron (1988).]

To test for unit roots (or serial correlation) in an AR(1) model, it is necessary to know the distributions of $\hat{\beta}$ for $\beta=1$ (or $\beta=0$). The distribution of $\hat{\beta}$ has been studied by many authors including Tanaka (1983), Phillips (1984), Durbin (1986) and Phillips and Reiss (1986). It had been shown by Tanaka (1983) and Phillips (1984), that Edgeworth approximations to the exact distribution of $\hat{\beta}$ perform poorly, especially in the tails and when the model is close to the border of nonstationarity. Phillips (1984) proposed a method based upon extended rational approximants (ERA's) which yields very good approximations and requires that the distribution of the time-series be known.

It is known that the OLS estimate of β has a degenerate limiting distribution unless uitably normalized. This normalizing factor is given in Evans and Savin (1984, p. 1245). Unlike conventional normalization which has the serious disadvantage of discontinuity in the limiting distribution, this normalization procedure is continuous. [See e.g. Evans and Savin (1984, pp. 1254-1256.] The normalizing factor is not a serious concern when $0 \leq \beta \leq 0.9$, because the conventional normalization is still continuous within this range. On the other hand, when β is close to the border of nonstationarity, it may be necessary to examine the distribution of the continuous normalized OLS estimate of β. This topic is set aside for future research and the focus of this chapter is on the case when $0 \leq \beta \leq 0.9$.

When the distribution of the time-series is unknown, one may use the bootstrap method to approximate the exact distribution of $\hat{\beta}$. Although the bootstrap method has been applied elsewhere, no one has actually studied its applications for a pure AR(1) processes. In a related area, Kiviet (1984) applies the bootstrap to a linear regression model which has lagged dependent explanatory variables and i.i.d. errors. Based upon Monte Carlo simulation results, Kiviet concludes that bootstrapping is not very useful in lagged-dependent variable models. This view is shared by Veall (1986), for highly trended explanatory variable models with AR(1) errors, and Prescott and Stengos (1987), for a dynamic model with AR(1) errors. This is an interesting area for further research. The results obtained in this chapter may help to further the understanding of the underlying problem(s). This chapter examines the problems associated with bootstrap applications for a simple AR(1) model, and attention is restricted to the ordinary least-squares estimate (OLSE) of β.

Let the time series $y_1, y_2, \ldots, y_n$ be generated by the AR(1) process

$$y_t = \beta y_{t-1} + \varepsilon_t \quad ,(t=1,2,\ldots,n) \tag{6.1.1}$$

in which $\beta^2<1, y_0=\varepsilon_0, E(\varepsilon_t)=0$ and $E(\varepsilon_t^2)=\sigma^2$ $\forall$ $t=0,1,\ldots,n$, and $E(\varepsilon_t\varepsilon_s)=0$ $\forall$ $t\neq s$. The errors $\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_n$ are nonobservable, and the objective is to obtain a consistent estimate of β given $y_1,y_2,\ldots,y_n$. One such estimate is the OLSE of β, which is given by

$$\hat{\beta} = \left\{\sum_{t=2}^{n}\left[y_t y_{t-1}\right]\right\}\left\{\sum_{t=2}^{n}\left[y_{t-1}^2\right]\right\}^{-1}. \tag{6.1.2}$$

When y_0 is fixed and when the errors are normally distributed, $\hat{\beta}$ is also the maximum-likelihood estimate (MLE) of β. On the other hand, when y_0 is also random such that each y_t is $N\{0,\sigma^2/(1-\beta^2)\}$, the MLE of β will be

$$\hat{\beta}_{mle} = \left\{\sum_{t=2}^{n}\left[y_t y_{t-1}\right]\right\}\left\{\sum_{t=2}^{n-1}\left[y_{t-1}^2\right]\right\}^{-1}.$$

The derivation of $\hat{\beta}_{mle}$ is quite tedious and can be found in White (1961, p.86). Since the two estimates are asymptotically the same, only the $\hat{\beta}$ will be considered here.

In the time-series literature, $\hat{\beta}$ is known as the first-order autocorrelation coefficient. Kendall (1944) first employed the terms serial correlation for the calculated value of the correlation in time series and autocorrelation for the population value. [See e.g., Bartlett (1946, p.27).]

When $E(\varepsilon_t)=0$ $\forall$ $t=0,1,\ldots,n$, it can easily be shown that the mean of y is zero, and the ordinary definition of the k'th serial correlation coefficient can be written as

$$r_k = \frac{(n-k)^{-1}\left\{\sum_{t=1}^{n-k}\left[y_t y_{t-k}\right]\right\}}{n^{-1}\left\{\sum_{t=1}^{n}\left[y_t^2\right]\right\}}. \tag{6.1.3}$$

Bartlett (1946) derived a general formula for the variance of r_k, and showed that it depends on all the autocorrelations of the time series. Among several authors, Marriott and Pope (1954), Kendall (1954), Quenouille (1956), White (1961) and Shenton and Johnson (1965) showed that the serial correlation is a biased estimate of the autocorrelation, and that the bias is generally downward. For the Markov scheme defined in (6.1.1), Marriott and Pope (1954, p.394) showed that, when ε is a normal variate with mean zero and unit variance,

$$E(r_k) = \left[\beta^k - 2n^{-1}k\beta^k\right] + O(n^{-2}).$$

When k=1,

$$E(r_1) = \left[\beta - 2n^{-1}\beta\right] + O(n^{-2}).$$

The difference between r and $\hat{\beta}$ lies in the denominator. Writing N_1 for the numerator of (6.1.3) when k=1, the relationship between $\hat{r}_1$ and $\hat{\beta}$ can be written as

$$\hat{\beta}^{-1} = \left\{n\left[(n-1)r_1\right]^{-1} - y_n^2\left[N_1(n-1)\right]^{-1}\right\} \tag{6.1.4}$$

or

$$\hat{\beta} = \left\{\left[(n-1)r_1N_1\right]\left[nN_1+r_1y_n^2\right]^{-1}\right\}.$$

It will be shown, in Section 6.3, that the difference between the expectation of $\hat{\beta}$ and that of r is of $O(n^{-2})$. Therefore, the mean and variance of r can be used to approximate the mean and variance of $\hat{\beta}$, respectively. This follows directly from the work of White (1961).

The organization of this chapter is as follows. The following section gives an overview of the difficulties in obtaining approximations of the exact distribution of $\hat{\beta}$. The bootstrap method is suggested here as a better alternative. Section 6.3 consists of a short discussion on the mean and variance of $\hat{\beta}$, and it is shown that $\hat{\beta}$ is biased downward. An unbiased estimate of β is suggested in Section 6.4. Although this unbiased estimate of β is easy to obtain, its distribution is generally unknown and has to be approximated by using the bootstrap technique of Efron (1979), especially when the error distribution is unknown. Details regarding the bootstrap procedure are also discussed in this section. In Section 6.5, a simulation study is conducted to obtain empirical significance levels of the BCI's of β. This is done for the case when the value of β is unknown. The bootstrap distribution of $\hat{\beta}$ when β=0 is examined in Section 6.6. In Section 6.7, two additional lemmas concerning the properties of selection matrices are derived. A theorem on the mean and variance of bootstrap estimates of β is also obtained in this section.

The simulation results suggest that conventional confidence intervals (CCI's) should be avoided whenever β is positive. The reason for this is that CCI's are no longer reliable, since the normal approximation is poor when β approaches unity. One alternative is to use the method based upon ERA's, which is quite cumbersome to use. Another alternative is the bootstrap procedure, which is computer based and is quite easy to implement. In constructing bootstrap confidence intervals, critical values from the Student's t-distribution are replaced by the corresponding bootstrap t-statistics.

A common situation when confidence intervals are constructed, is one in which the variance of the estimator is not a function of the estimator itself. In the present case, this only happens when $\beta=0$. Otherwise, the variance of $\hat{\beta}$ will depend on the value of β. Simulation results suggest that the relationship between $\hat{\beta}$ and its variance should be accounted for whenever $\beta\neq0$. Only BCI's that account for this relationship should be used. It is then not surprising, that the suggested BCI has the best performance in comparison to other confidence intervals, at both the 5% and 10% significance levels. Narrower BCI's should be avoided, because of possible problems relating to an underlying restriction.

6.2 The Distribution of $\hat{\beta}$

The statistic $\hat{\beta}$ is a non-circular serial correlation coefficient whose distribution is supported over the entire real line; its exact distribution is difficult to approximate. When $\beta=0$ and when the error distribution is normal, the distribution of $\hat{\beta}$ is symmetric. However, as β approaches unity (say, $\beta=0.9$), its distribution becomes heavily skewed to the left. When this happens, the normal approximation will be very poor, and alternative methods which give better approximations have to be used. One such alternative is the approximation by the method of Edgeworth expansions.

The distribution of $\hat{\beta}$ had been studied by many authors including Tanaka (1983), Phillips (1984), Durbin (1986) and Phillips and Reiss

1986). Both Tanaka (1983) and Phillips (1984) show that Edgeworth approximations to the exact distribution of $\hat{\beta}$ perform poorly, especially in the tails, when the model is close to the border of stationarity. They also show that the normal approximation is not very satisfactory for sample sizes of less than or equal to 20. Although the Edgeworth approximation is generally more accurate than normal approximation, its poor performance in the tails makes it unreliable for constructing confidence intervals. One other alternative is the method based upon ERA's. Phillips (1984) shows that the method based upon ERA's yields very close estimates for β when n is greater than or equal to 5. This method requires a special computer function routine which can be found in Phillips and Reiss (1986). The use of this computer routine is restricted to an AR(1) model. For some practitioners, this can be a major inconvenience. An important shortcoming of this approach is that it requires the distribution of the time series be known. When this distribution is unknown, two possible alternatives are the methods based upon the jackknife and the bootstrap. Both are nonparmetric methods and they are both distribution-free. While the jackknife is only applicable when the sample size is large, the bootstrap can be applied to both small and large samples. Thus, when the error distribution is unknown and when the sample size is small, the bootstrap method is the only viable method for approximating the exact distribution of $\hat{\beta}$.

The focus of this chapter will be on the use of Efron's (1979) bootstrap for aproximating the exact distribution of the least-squares estimate of a stationary AR(1) process, which can be Gaussian or non-Gaussian. However, for the Monte Carlo simulations below, only the stationary Gaussian process will be considered.

The problem of approximating the exact distribution of $\hat{\beta}$ is a common interest shared by all applied statisticians, including econometricians. This chapter is essential for a better understanding of the application of bootstrap to this problem. Also, the results of this chapter provide a foundation for future applied econometric work in this area, especially when it involves the bootstrap.

6.3 The Mean and Variance of $\hat{\beta}$

The first and second moments of $\hat{\beta}$ are given by White (1961, p.89) as

$$E(\hat{\beta}) = \left[1 - 2n^{-1}\right]\beta + O(n^{-2}) \qquad (6.3.1)$$

and

$$E(\hat{\beta}^2) = \left\{n^{-1}-n^{-2}+\left[1-5n^{-1}+22n^{-2}\right]\right\}\beta^2+O(n^{-3}). \qquad (6.3.2)$$

Thus,

$$V(\hat{\beta}) = E(\hat{\beta}^2) - \left[E(\hat{\beta})\right]^2 = \left[n^{-1}(1-\beta^2)\right] + O(n^{-2}).$$

The corresponding values for r_1 are given by Marriott and Pope (1954, p.394); $E(r_1)=E(\hat{\beta})$, $V(r_1)=V(\hat{\beta})$. The reason for this similarity is that the difference between $E(\hat{\beta})$ and $E(r_1)$ is $O(n^{-2})$. This may be shown as follows.

Since $E(y_t y_{t-1}) = \beta\left[\text{var}(y_{t-1})\right]$ and $\text{var}(y_{t-1}) = \sigma^2\left[1-\beta^2\right]^{-1}$ (see e.g., Marriott and Pope, 1954 p.394), it follows that

$$E\left[N_1^{-1}y_n^2\right] = \beta^{-1} + O(n^{-2}).$$

Taking expectations on both sides of (6.1.4), the expected value of $\hat{\beta}$ can be written as

$$E(\hat{\beta}) = \left\{n\left[(n-1)E(r_1)\right]^{-1} - (n-1)^{-1}E\left[N_1^{-1}y_n^2\right]\right\}^{-1}.$$

Upon substitution of $E(r_1)$ and $E\left[N_1^{-1}y_n^2\right]$ into this expression and after some algebraic simplification,

$$E(\hat{\beta}) = \left[(n-1)(n-1)(n^2-n+2)^{-1}\right]\beta + O(n^{-2}).$$

Since

$$\left[(n-1)(n-2)(n^2-n+2)^{-1}\right] = \left[(1-2n^{-1}) - 2n^{-2}\right] + O(n^{-3}),$$

it is clear that

$$\left[E(\hat{\beta}) - E(r_1)\right] = O(n^{-2}).$$

6.4 An Almost Unbiased Estimate of β and its Bootstrap Distribution

It follows from (6.3.1) that, to $O(n^{-1})$, an almost unbiased estimate of β is

$$\hat{\beta}_c = \left[1-2n^{-1}\right]\hat{\beta}.$$

The variance of this estimate is, from (6.3.2),

$$V(\hat{\beta}_c) = \left[(1-2n^{-1})^{-2}\right]V(\hat{\beta}) = \left[n(n-2)^{-2}(1-\beta)^2\right] + O(n^{-2}).$$

Since the distribution of $\hat{\beta}_c$ is similar to that of $\hat{\beta}$, it is difficult to approximate this distribution using Edgeworth expansions. As shown by Phillips (1984), this difficulty becomes more apparent, especially in the tails of the distribution, when β approaches unity.

Although the exact distribution of $\hat{\beta}$ is difficult to obtain, a reliable approximation to it may be obtained by applying the bootstrap method. However, before discussing the bootstrap method, it would be useful to note certain properties of the OLS residuals of (6.1.1), since these will eventually be used for bootstrapping.

The OLS residuals of (6.1.1) are given by

$$\hat{\varepsilon}_t = y_t - \hat{\beta} y_{t-1} \qquad (t=2,\ldots,n).$$

Since the model (6.1.1) does not have an intercept, the sample mean of these residuals is generally not equal to zero. Consequently, to be suitable for bootstrapping, they have to be centered at their sample mean. (C.f. Chapter 2). Also, $\hat{\beta}$ is independent of the error variance and the bootstrap estimates of β are independent of any scaling factor. Thus, it is not necessary to rescale the residuals by a constant factor here. Multiplying the residuals by a constant factor here will leave the bootstrap estimates unchanged, regardless of the value of n.

Let $\tilde{\varepsilon}$ a vector of properly conditioned OLS residuals, be used for bootstrapping. This is obtained as

$$\tilde{\varepsilon}_t = \left[\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right],$$

in which $\bar{\hat{\varepsilon}}$ is the sample mean of the OLS residuals. As explained above, it is not necessary to multiply each ε_t by the factor $\left[n(n-1)^{-1}\right]^{1/2}$. For each bootstrap replication j $(j=1,2,\ldots,J)$, the bootstrap errors are obtained as

$$\varepsilon^*_{(j)} = S_{(j)}\tilde{\varepsilon}$$

in which $S_{(j)}$ is an $(n+1)$xn selection matrix and $\varepsilon^{*T}_{(j)} = [\varepsilon^*_{(j)0}, \varepsilon^*_{(j)1}, \ldots, \varepsilon^*_{(j)n}]$. The bootstrap responses are constructed as

$$y^{*}_{(j)t} = \hat{\beta}_c y^{*}_{(j)t-1} + \varepsilon^{*}_{(j)t} \qquad (t=1,\ldots,n) \tag{6.4.1}$$

in which $y^{*}_{(j)0} \equiv \varepsilon^{*}_{(j)0}$.

To obtain a confidence interval for β, the conventional method is to apply the critical values from the standard t-table. However, when the distribution of $\hat{\beta}_c$ is not normal, the conventional method fails to yield reliable estimates for the confidence bounds. One alternative is to replace the critical values from the Student's t-distribution by the corresponding bootstrap t-statistic, which is defined as

$$t^{*}_{c(j)} = \left(\beta^{*}_{c(j)} - \bar{\beta}^{*}_{c}\right)\left(\hat{V}(\beta^{*}_{c(j)})\right)^{-1/2},$$

in which

$$\beta^{*}_{c(j)} = (1-2n^{-1})^{-1}\beta^{*},$$

$$\beta^{*}_{(j)} = \left\{\sum_{t=2}^{n}\left[y^{*}_{(j)t}y^{*}_{(j)t-1}\right]\right\}\left\{\sum_{t=2}^{n}\left[y^{*2}_{(j)t-1}\right]\right\}^{-1},$$

$$\bar{\beta}^{*}_{c} \equiv \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{c(j)}\right]\right\}, \text{ and}$$

$$\hat{V}(\beta^{*}_{c(j)}) = \left[(1-2n^{-1})^{-2}\right]\hat{V}(\beta^{*}_{(j)})$$

$$= \left[n(1-2n^{-1})^{2}\right]^{-1}\left[1-\beta^{*2}_{(j)}\right].$$

Note that $t^{*}_{c(j)}$ is identical to

$$t^{*}_{(j)} = \left[\beta^{*}_{(j)} - \bar{\beta}^{*}\right]\left[\hat{V}(\beta^{*}_{(j)})\right]^{-1/2},$$

in which

$$\bar{\beta}^{*} \equiv \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{(j)}\right]\right\}.$$

This is no surprise, since $\hat{\beta}_c$ is actually $\hat{\beta}$ multiplied by a constant whose value depends only on n.

The critical values of $t^{*}_{(j)}$ can be obtained by arranging all its values in ascending order and taking the values corresponding to the respective percentiles. Let one of these values be t^{*}_{z} corresponding to the z'th percentile. One confidence bound for β based upon t^{*}_{z} will be

$$\hat{\beta}_z = \hat{\beta}_c \pm \left[\hat{V}(\hat{\beta}_c)\right]^{1/2} t^{*}_{z}, \qquad (6.4.2)$$

in which

$$\hat{V}(\hat{\beta}_c) = n(n-2)^{-2}\left[1-\hat{\beta}^2_c\right].$$

However, this confidence bound assumes that the variance of $\hat{\beta}_c$ is independent of $\hat{\beta}_c$. This assumption is not appropriate here, since

$$\hat{V}(\hat{\beta}_c) = (nm)^{-1}\left[1-m\hat{\beta}^2_c\right],$$

in which $$m = \left[1-2n^{-1}\right]^{-2}.$$

Confidence intervals obtained via (6.4.1) are also not reliable, especially when

$$P\left[\ \hat{\beta} \geq 1 \mid \beta\ \right] \geq z.$$

The reason for this is that, whenever $\hat{\beta}\geq 1$, it is usually restricted to be 0.9999 or some other value which is less than but close to unity, to avoid the problem of complex roots. When

$$P\left[\hat{\beta} \geq 1 \mid \beta \right] < z,$$

the reliability of the confidence interval is not affected, because these cases belong to the rejection region. On the other hand, when the proportion of the above cases exceeds the rejection value, confidence intervals with significance levels less than or equal to z serve no useful purpose.

Hence, a better confidence bound for β based upon $t^*_{(j)}$ needs to be constructed. As is explained below, one such confidence bound would be

$$\tilde{\beta}_z = \left\{ -v_b + \left[v_b^2 - 4v_a v_c \right]^{1/2} \right\} \left[2v_a \right]^{-1}, \qquad (6.4.3a)$$

in which

$$v_a = 1 + n(n-2)^{-2} t_z^{*2},$$

$$v_b = -2\hat{\beta}_c, \text{ and}$$

$$v_c = \hat{\beta}_c^2 - n(n-2)^{-2} t_z^{*2}.$$

This is the upper bound, when t^*_z is positive. When t^*_z is negative, the corresponding confidence bound is

$$\tilde{\beta}_z = \left\{ -v_b - \left[v_b^2 - 4v_a v_c \right]^{1/2} \right\} \left[2v_a \right]^{-1}. \qquad (6.4.3b)$$

This procedure is based upon

$$t^*_z = (\hat{\beta}_c - \beta)(1 - \hat{\beta}_c^2)^{-1/2},$$

and the extreme bounds are obtained by solving for β in the quadratic equation

$$v_a\beta^2 + v_b\beta + v_c = 0.$$

Confidence intervals obtained by this latter method are better than those obtained via the earlier method. One improvement is that the second procedure accounts for the dependence of $\hat{V}(\hat{\beta}_c)$ on $\hat{\beta}_c$. Also, the second confidence interval is less restrictive, especially in the tails of the distribution. For any given value of t_z^*, the value of $\hat{\beta}_c$ must not exceed its critical value given by

$$\hat{\beta}_{crit} = \left[1 + n(n-2)^{-2}t_z^{*2}\right]^{1/2}. \tag{6.4.4}$$

This restriction is necessary in order to avoid the problem of complex roots, but poses no problem when the value of t_z^* is large. This is true especially in the tails. However, when the value of t_z^* is small and the corresponding value of $\hat{\beta}_c$ is so large that it exceeds $\hat{\beta}_{crit}$, one would have to restrict the value of $\hat{\beta}_c$ to its critical value. When this happens, the corresponding bootstrap confidence interval becomes unreliable. This usually occurs for large values of $\hat{\beta}_c$, and when the confidence intervals are constructed near the center of the distribution.

6.5 Empirical Significance Levels of Bootstrap Confidence Intervals

In this section, a simulation study will be conducted to obtained empirical significance levels of the various bootstrap confidence intervals, and to compare these levels with the significance levels of conventional confidence intervals. The results are obtained for 500 trials and 200 bootstrap replications, in order to avoid using too much computing time.

The initial value for the double precision SEED is set at 12345 for both the random variable generator and the bootstrap procedure. The Gaussian errors with mean zero and unit variance are generated by the IMSL's GGNPM subroutine. The range of values for β is chosen to lie between zero and unity, because many economic time-series data are known to be positively serially correlated. Two sample sizes, n=10 and n=20 are selected, since they are normally chosen for small sample studies. Finally, the nominal significance levels are 2.5%, 5% and 10% for both tails.

The simulation results are reported in Tables 6.1 and 6.2. For reporting the results, bootstrap confidence intervals based upon (6.4.2) and (6.4.3) are denoted by BCI1 and BCI2, respectively. When the bootstrap t-values are replaced by the corresponding Student's t-values in (6.4.2) and (6.4.3), the resulting conventional confidence intervals are CCI1 and CCI2, respectively.

From the results of Tables 6.1 and 6.2, it can be concluded that BCI1 is inferior to BCI2. BCI1 is also inferior when compared to either CCI1 or CCI2. When compared to CCI2, BCI2 performs better when the value of β lies between zero and 0.75 (both values inclusive), and its performance is marginally inferior to that of CCI2 when β=0.9, especially in the lower tail. When the value of β lies between 0.5 and 0.9, the reliability of either BCI1 or CCI1 is questionable. This is because of the restriction $\left(1-\hat{\beta}^2\right)>0$, which is imposed to avoid the problem of complex roots. As the value of β approaches unity, the

Table 6.1: Empirical Significance Levels of Conventional and Bootstrap Confidence Intervals for β When the Errors are Normal [n=10][1]

Confidence Interval		True %	β 0.0	0.25	0.50	0.75	0.90
CCI1	Lower	2.5	1.5	2.6	4.9	14.3	26.8
		5.0	4.6	4.6	9.0	18.6	30.1
		10.0	9.2	11.7	17.0	25.9	36.5
	Upper	2.5	2.1	1.7	1.3	1.3	1.2
		5.0	4.8	3.7	2.9	2.3	2.4
		10.0	8.7	7.1	6.9	5.8	4.8
CCI2	Lower	2.5	0.4	0.3	0.1	0.1	0.2
		5.0	1.5	1.1	0.6	0.9	2.3
		10.0	7.6	6.5	6.4	7.4	12.5
	Upper	2.5	0.2	0.8	1.4	2.6	3.8
		5.0	2.1	3.4	3.6	4.4	5.3
		10.0	6.7	7.1	8.1	8.7	8.3
BCI1	Lower	2.5	3.4	5.6	8.6	19.0	31.0
		5.0	7.0	9.2	13.0	24.2	34.0
		10.0	11.4	14.4	22.6	33.6	40.6
	Upper	2.5	1.8	0.8	0.4	0.6	0.8
		5.0	4.2	3.2	2.4	1.8	1.6
		10.0	9.6	7.0	6.6	5.2	4.2
BCI2	Lower	2.5	1.8	1.8	0.8	0.4	2.0
		5.0	4.6	5.2	4.0	3.6	8.2
		10.0	9.8	11.0	14.0	18.0	21.0
	Upper	2.5	0.4	0.2	0.6	1.6	1.8
		5.0	3.0	2.8	2.8	3.8	4.8
		10.0	8.2	7.0	7.4	7.0	6.4

1. Results for both CCI1 and CCI2 were based upon 1,000 trials, while results for both BCI1 and BCI2 were based upon 500 trials and 200 bootstrap replications.

Table 6.2: Empirical Significance Levels of Conventional and Bootstrap Confidence Intervals for β When the Errors are Normal [n=20][1]

Confidence Interval		True %	β = 0.0	0.25	0.50	0.75	0.90
CCI1	Lower	2.5	2.8	3.6	4.8	10.3	18.6
		5.0	5.2	5.7	7.4	12.6	23.6
		10.0	10.0	11.2	13.2	20.5	30.0
	Upper	2.5	1.4	1.2	0.8	1.0	1.1
		5.0	3.7	3.2	2.5	2.3	2.7
		10.0	8.2	7.9	7.2	5.9	6.1
CCI2	Lower	2.5	1.7	1.2	0.8	0.2	0.1
		5.0	4.0	3.5	2.5	1.3	2.1
		10.0	9.2	8.1	7.8	9.6	10.0
	Upper	2.5	0.9	1.2	1.6	3.0	4.7
		5.0	2.8	3.4	4.3	5.5	7.0
		10.0	7.4	8.0	8.2	9.0	9.4
BCI1	Lower	2.5	4.0	5.8	6.8	12.4	22.8
		5.0	6.2	8.2	11.2	18.4	30.0
		10.0	11.3	14.4	17.6	25.6	36.4
	Upper	2.5	1.4	1.2	0.8	0.4	0.8
		5.0	3.4	2.6	1.4	1.8	1.8
		10.0	8.6	6.6	5.2	4.6	4.4
BCI2	Lower	2.5	3.2	3.2	2.0	1.0	1.0
		5.0	5.4	5.4	5.4	4.2	4.4
		10.0	10.6	11.6	12.2	12.8	16.8
	Upper	2.5	0.8	1.0	0.8	2.0	3.0
		5.0	2.8	2.8	3.0	3.8	4.6
		10.0	8.0	6.8	6.8	6.4	7.0

1. Results for both CCI1 and CCI2 were based upon 1,000 trials, while results for both BCI1 and BCI2 were based upon 500 trials and 200 bootstrap replications.

proportion of $\hat{\beta}$'s exceeding unity increases. When this proportion exceeds a critical level, the bootstrap confidence bound corresponding to this critical level can no longer serve its purpose. Consequently, BCI's that do not require this restriction would be expected to perform better. This explains why the overall performance of BCI2 is better when compared to BCI1, especially when $0.5 \le \beta \le 0.9$.

When $\beta=0$, the simulation results indicate that CCI1 is better than CCI2, and BCI1 is better than BCI2. This is because, when $\beta=0$, the variance of $\hat{\beta}$ is independent of β. The normal approximation is good in this case and, hence, there is little to gain from bootstrapping. When the value of β approaches one, CCI2 is better than CCI1, and BCI2 is better than BCI1. This result is expected, because both CCI2 and BCI2 account for the dependence of $V(\hat{\beta}_c)$ on $\hat{\beta}_c$, as mentioned in Section 6.4 above.

The bootstrap confidence interval BCI2 generally performs well, especially in the five percent tail. Its performance in the 2.5 percent tail is poor because BCI's admit errors of $O(n^{-1})$. Lastly, its poor performance in the ten percent tail can be attributed to the above restriction that $\hat{\beta}_c$ not exceeds $\hat{\beta}_{crit}$, whose value is given by equation (6.4.4). This type of BCI is generally not needed because, in the ten percent tail, CCI's are adequate approximations of the exact confidence intervals.

Generally, for $0 \le \beta \le 0.9$, BCI2 exhibits the best performance in terms of its empirical coverage at both the 5% and 10% levels. At the 20% level, the empirical coverage of CCI2 is closest to the nominal coverage, among the confidence intervals considered here. The empirical coverage of a confidence interval of β is the observed frequency that the true value of β lies in that interval. When $\beta=0$, both CCI1 and BCI1 are fairly accurate. Since BCI1 requires more computation, there is no advantage to be gained from bootstrapping when the autocorrelation is suspected to be zero, especially when the error distribution is normal.

6.6 Bootstrap Distribution of $\hat{\beta}_c$ When β is Known

Both the confidence intervals discussed above are based upon $\hat{\beta}_c$, and the null hypothesis that $\beta=\beta_0$ will be rejected, when the value of β_0 lies outside the relevant confidence bounds constructed at the z'th level of significance. Let this set of confidence intervals be known as the set of confidence intervals based upon the alternative hypothesis that $\beta\neq\beta_0$. One other set of confidence intervals is the set of confidence intervals based upon the null hypothesis. This latter set of confidence intervals is based upon β_0, and the null hypothesis will be rejected when the value of $\hat{\beta}_c$ lies outside the calibrated confidence bounds.

When the distributions of $n^{1/2}\left(\hat{\beta}_c-\beta\right)$ under each of the two hypothesis are identical, the two approaches will be equivalent. This happens only when the distribution of $n^{1/2}\left(\hat{\beta}_c-\beta\right)$ is invariant to a change in location. Otherwise, the two approaches will be different.

In the first approach, the distribution of $\hat{\beta}_c$ has $\hat{\beta}_c$ as its location parameter. Subsequently, the bootstrap distribution of $\hat{\beta}_c$ when β is unknown, must also have $\hat{\beta}_c$ as its location parameter. On the other hand, the distribution of $\hat{\beta}_c$ is obtained around β_0, when it is known that $\beta=\beta_0$. Hence, the corresponding bootstrap distribution of $\hat{\beta}_c$ when β is known, must have β_0 as its location parameter.

When the first set of confidence intervals become unreliable, it is natural to seek for alternative confidence intervals. One such alternative is to assume that $\beta=\beta_0$ under the null hypothesis, and use the second set of confidence intervals. When $\beta_0=0$, the normal approximation can be used for constructing these intervals. The normal approximation will be poor when $\beta_0\neq 0$, and it will become poorer as β_0 approaches the unit circle. When this happens, the bootstrap distribution of $\hat{\beta}_c$, with β_0 as its location parameter, can be used for constructing confidence intervals. In the event when β_0 is known, the bootstrap distribution of $\hat{\beta}_c$ can easily be obtained. This distribution has β_0 as its location parameter. The only change to the above bootstrap procedure is to replace $\hat{\beta}_c$ by β_0 in equation (6.4.1).

Under the assumption that the bootstrap estimates $\left[\beta^{*}_{c(j)}, (j=1,\ldots,J)\right]$ are unbiased estimates of β_0,

$$\bar{\beta}^{*}_{c} \equiv \left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^{*}_{c(j)}\right]\right\} \xrightarrow[J]{a.s.} \beta_0.$$

Nevertheless, preliminary simulation results indicated that there is a small upward bias in $\beta^{*}_{c(j)}$. This is probably due to the over-correction of the initial underestimation in $\hat{\beta}$. In other words,

$$\bar{\beta}^{*}_{c} \xrightarrow[J]{a.s.} a\beta_0$$

in which a>1. Hence, an unbiased estimate of the bias of $\beta^{*}_{c(j)}$, $(j=1,\ldots,J)$, is

$$\hat{\text{Bias}}\left[\bar{\beta}^{\bullet}_{c}\right] = \left[\bar{\beta}^{\bullet}_{c} - \beta_{0}\right],$$

and an unbiased bootstrap estimate of β can be written as

$$\beta^{\bullet}_{u(j)} = \beta^{\bullet}_{c(j)} - \hat{\text{Bias}}\left[\bar{\beta}^{\bullet}_{c}\right].$$

In the subsequent section, it will be shown that the bias in the variance of the bootstrap estimates is $O(n^{-2})$. The same order of bias also holds for the mean of the bootstrap estimates. When β is known, the bias in the mean (i.e. location bias) of $\hat{\beta}_c$ can easily be corrected. On the other hand, it is rather cumbersome to adjust for the bias in the variance of the bootstrap estimates. Since this bias is generally $O(n^{-2})$, it will be ignored, and this will not affect the results significantly.

From the analytical results of White (1961, p.90), the exact location of the mean of $\hat{\beta}_c$ is approximately $\left[n^{-2}(4\beta+2\beta^3+2\beta^5)\right]$ above β_0. Preliminary simulation results showed that the empirical significance level of the bootstrap confidence intervals is very sensitive to this bias. This is true especially for the upper confidence bound which happens to have a larger variance than the lower confidence bound. The results reported in Table 6.3 have been corrected for bias.

From Table 6.3, it is without doubt that the bootstrap confidence interval can serve as a reasonably good approximation to the exact confidence interval. Thus, the bootstrap distribution of $\hat{\beta}_c$ may be used for approximating the exact distribution of $\hat{\beta}_c$. Since the method based on ERA's to obtain the exact distribution of $\hat{\beta}_c$ is difficult to implement, the bootstrap procedure should appeal to both applied econometricians and statisticians.

Table 6.3: Empirical Significance Levels of Bootstrap Confidence Intervals for β when the errors are Normal and When the Value of β is Known[1]

Confidence Interval		True %	β 0.0	0.25	0.50	0.75	0.90
BCI3 [n=10]	Lower	2.5	2.0	2.4	2.4	3.0	3.0
		5.0	5.0	4.8	5.2	5.6	5.2
		10.0	9.6	9.2	10.0	10.4	9.8
	Upper	2.5	1.2	1.0	1.2	1.4	1.0
		5.0	3.2	3.8	3.8	3.0	5.2
		10.0	8.8	8.8	8.4	10.8	11.6
BCI3 [n=20]	Lower	2.5	2.0	2.4	1.6	2.2	2.8
		5.0	4.6	4.6	4.0	4.4	4.8
		10.0	8.4	6.6	8.6	8.4	8.0
	Upper	2.5	3.2	3.0	2.4	2.8	2.2
		5.0	4.8	5.8	5.8	6.0	5.4
		10.0	8.8	9.8	11.8	12.0	11.4

1. Results for BCI3 were based upon 500 trials and 200 bootstrap replications.

6.7 The Mean and Variance of the Bootstrap Estimates of β

Let β be known and $\hat{\varepsilon}$ be the OLS residuals. Also, let the bootstrap responses be constructed as

$$y^{*}_{(j)t} = \beta y^{*}_{(j)t-1} + \varepsilon^{*}_{(j)t}, \qquad (t=1,\dots,n), \qquad (6.7.1)$$

in which $y^{*}_{(j)0}=\varepsilon^{*}_{(j)0}$, $\varepsilon^{*}_{(j)}=S_{(j)}\tilde{\varepsilon}$ and $S_{(j)}$ is an $(n+1)$xn selection matrix, as defined in Chapter Two. The bootstrap estimates of β in (6.7.1) are

$$\beta^{*}_{(j)} = \left\{\sum_{t=2}^{n}\left[y^{*}_{(j)t}y^{*}_{(j)t-1}\right]\right\}\left\{\sum_{t=2}^{n}\left[y^{*2}_{(j)t-1}\right]\right\}^{-1}, \quad (j=1,\dots,J). \qquad (6.7.2)$$

In matrix notation, (6.7.1) can be written compactiy as

$$y^{*}_{(j)} = AS_{1(j)}\tilde{\varepsilon} + S_{2(j)}\tilde{\varepsilon}, \qquad (6.7.3)$$

in which A is the following lower triangular matrix,

$$A = \begin{bmatrix} \beta & & & & \\ \beta^2 & \beta & & & \\ \beta^3 & \beta^2 & \beta & & \\ \vdots & \vdots & \ddots & \ddots & \\ \beta^n & \beta^{n-1} & \dots & \beta^2 & \beta \end{bmatrix},$$

$S_{1(j)}$ is an nxn selection matrix whose components are the first n rows of $S_{(j)}$, and $S_{2(j)}$ is also an nxn selection matrix whose components are the last n rows of $S_{(j)}$. Also, let $\bar{A}$ and $\underline{A}$ be $(n-1)$xn matrices whose components are the first and last $(n-1)$ rows of A, respectively.

Further, let $S_{3(j)}$ and $S_{4(j)}$ be (n-1)xn selection matrices, such that the components of $S_{3(j)}$ are the last (n-1) rows of $S_{(j)}$, and that $S_{4(j)}$ be $S_{(j)}$ with the first and last rows removed. The bootstrap estimates given by (6.7.2) can now be written as

$$\beta^*_{(j)} = \frac{\left[\bar{A}S_{1(j)}\tilde{\varepsilon} + S_{3(j)}\tilde{\varepsilon}\right]^T\left[\bar{A}S_{1(j)}\tilde{\varepsilon} + S_{4(j)}\tilde{\varepsilon}\right]}{\left[\bar{A}S_{1(j)}\tilde{\varepsilon} + S_{4(j)}\tilde{\varepsilon}\right]^T\left[\bar{A}S_{1(j)}\tilde{\varepsilon} + S_{4(j)}\tilde{\varepsilon}\right]} . \tag{6.7.4}$$

Note that equation (6.7.4) can also be written as

$$\beta^*_{(j)} = \beta + \left(V^*_{(j)}W^{*-1}_{(j)}\right) , \tag{6.7.5}$$

in which $V^*_{(j)} = \left(V^*_1+V^*_2\right)$, $W^*_{(j)} = \left(W^*_1+W^*_2+W^*_3\right)$,

$$V^*_1 = \left(\tilde{\varepsilon}^T S^T_{3(j)}\bar{A}S_{1(j)}\tilde{\varepsilon}\right),$$

$$V^*_2 = \left(\tilde{\varepsilon}^T S^T_{3(j)}\bar{A}S_{4(j)}\tilde{\varepsilon}\right),$$

$$W^*_1 = \left(\tilde{\varepsilon}^T S^T_{1(j)}\bar{A}^T\bar{A}S_{1(j)}\tilde{\varepsilon}\right),$$

$$W^*_2 = \left(2\tilde{\varepsilon}^T S^T_{1(j)}\bar{A}^T S_{4(j)}\tilde{\varepsilon}\right), \text{ and}$$

$$W^*_3 = \left(\tilde{\varepsilon}^T S^T_{4(j)}S_{4(j)}\tilde{\varepsilon}\right).$$

The subscript j in V^*_1 , V^*_2 , W^*_1 $\quad ^*$ and W^*_3 has been dropped for the ease of notation.

For comparison purposes, let

$$\hat{\beta} = \beta + (\hat{V}/\hat{W}) ,$$

in which $\hat{V} = \left\{\sum_{t=2}^{n}\left[y_t\varepsilon_t\right]\right\}$, $\hat{W} = \left\{\sum_{t=2}^{n}\left[y_{t-1}^2\right]\right\}$ and $E(\hat{V}/\hat{W})$ is the bias of $\hat{\beta}$. From the results of White (1961, p.90), it is known that the ratio $(\hat{V}/\hat{W})$ is of $O(n^{-1})$, and both $\hat{V}$ and $(n^{-1}\hat{W})$ are of $O(1)$. Also, let $\upsilon=E(\hat{V})$ and $\omega=E(\hat{W})$. Further, for any bootstrap variable $Z^{*}_{(j)}$ corresponding to the j'th $(j=1,\ldots,J)$ replication, let the notation $E_n(Z^{*})$ represents the limit of $E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[Z^{*}_{(j)}\right]\right\}$ as J goes to infinity for finite n. This assumes that the corresponding limit exists. It will be shown that,

(i) $E_n(V^{*}) = \upsilon + O(n^{-1})$,

(ii) $E_n(n^{-1}W^{*}) = (n^{-1}\omega) + O(n^{-1})$.

Note that both V and V^{*} are $O(1)$, whereas both W and W^{*} are $O(n)$. The following results are not surprising, since it is well known that bootstrap estimates of the above type admit errors $O(n^{-1})$.

Lemma 6.1: Let $A=\{a_{hi}\}$ be a real nxm matrix of finite elements and let $S_{(j)}(j=1,\ldots,J)$ be an mxr selection matrix. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}AS_{(j)}\right]\right\} \xrightarrow[J]{a.s.} \left\{\sum_{h=1}^{m}\left[a_{hh}\right]\right\}\left[(r^{-1})I_r\right] + \left\{\sum_{h\neq i}^{m}\left[a_{hi}\right]\right\}(r^{-2})\left[E(r,r)\right].$$

Proof:

Note that this lemma is a slight variant of Lemma 6.1. Let the elements of $S_{(j)}$ be $\{S_{hp}\}$. The subscript (j) is dropped to avoid confusion. For each j $(j=1,\ldots,J)$, the element corresponding to the p'th row and q'th column of the matrix $\left[S^{T}_{(j)}AS_{(j)}\right]$ can be written as

$$\left[S^T_{(j)}AS_{(j)}\right]_{p,q} = \left\{\sum_{h,i}^{m}\left[S_{hp}a_{hi}S_{iq}\right]\right\} = \left\{\sum_{\substack{h,i\\i=h}}^{m}\left[S_{hp}a_{hi}S_{iq}\right]\right\} + \left\{\sum_{\substack{h,i\\i\neq h}}^{m}\left[S_{hp}a_{hi}S_{iq}\right]\right\}.$$

When $i=h$, Prob. $\left[S_{hp}=S_{iq}=1\right] = (r^{-1})$ when $p=q$, zero otherwise. Consequently, when $p=q$, $(S_{hp}a_{hi}S_{iq})=a_{hi}$ and zero with probabilities (m^{-1}) and $(1-r^{-1})$, respectively. When $p\neq q$, $(S_{hp}a_{hi}S_{iq}) = 0$ with probability one. Thus,

$$\left\{\sum_{h,i}^{m}\left[S_{hp}a_{hi}S_{iq}\right]\right\} = \begin{cases}\left\{\sum_{h=1}^{m}(a_{hh})\right\} \text{ with probability } (r^{-1}) \text{ when } p=q \\ 0 \text{ with probabilities } (1-r^{-1}) \text{ and one, when } p=q \\ \qquad \text{and } p\neq q\text{, respectively.}\end{cases}$$

On the other hand, when $i\neq h$,

$$\text{Prob.}\left[S_{hp}=1,\ S_{iq}=1\right] = (m^{-2})\ \forall\ p,q$$

and

$$\left\{\sum_{\substack{h,i\\i\neq h}}^{m}\left[S_{hp}a_{hi}S_{iq}\right]\right\} = \left\{\sum_{\substack{h,i\\i\neq h}}^{m}\left[a_{hi}\right]\right\} \text{ with probability } (r^{-2})\ \forall\ p,q.$$

Therefore,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^T_{(j)}AS_{(j)}\right]_{p,q}\right\} \xrightarrow[J]{a.s.} \begin{cases}\left\{\sum_{h=1}^{m}\left[r^{-1}a_{hh}\right]\right\} + \left\{\sum_{\substack{h,i\\i\neq h}}^{m}\left[r^{-2}a_{hi}\right]\right\} \text{ when } p=q, \\ \left\{\sum_{\substack{h,i\\i\neq h}}^{m}\left[r^{-2}a_{hi}\right]\right\} \text{ when } p\neq q,\end{cases}$$

which is the desired result. *Q.E.D.*

Lemma 6.2: Let $S_{p(j)}$ and $S_{q(j)}$ be any two selection matrices of finite order mxr such that the r'th $(r=1,\ldots,n)$ rows both matrices are different. Then,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{p(j)}^{T}S_{q(j)}\right]\right\} \xrightarrow[J]{a.s.} \begin{cases}\left[(mr^{-1})I_r\right] & \text{when } p=q,\\ \left\{(mr^{-2})\left[E(r,r)\right]\right\} & \text{when } p\neq q.\end{cases}$$

Proof:

The proof is straightforward from the proofs of Lemmas 2.2 and 2.3.

The notation of Lemmas 6.1 and 6.2 will be used in both Lemma 6.3 and Theorem 6.1 below.

Lemma 6.3: Let m=n and n be finite. Then,

(i) $E_n(V_1^*) = E(\hat{V}_1) + O(n^{-1})$,

(ii) $E_n(V_2^*) = E(\hat{V}_2) = 0$,

(iii) $E_n(n^{-1}W_1^*) = E(n^{-1}\hat{W}_1) + O(n^{-1})$,

(iv) $E_n(n^{-1}W_2^*) = E(n^{-1}\hat{W}_2) + O(n^{-1})$,

(v) $E_n(n^{-1}W_3^*) = E(n^{-1}\hat{W}_3) + O(n^{-1})$.

Proof:

For the second statement, first note that

$$E_n(V_2^*) = E\left[\tilde{\varepsilon}^T\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{3(j)}^{T}S_{4(j)}\right]\right\}\tilde{\varepsilon}\right].$$

Upon application of Lemma 6.2 and let $m=(n^{-1})$ and r=n,

$$E_n(V_2^*) = E\left[\tilde{\varepsilon}^T\left[E(n,n)\right]\tilde{\varepsilon}^T\right] = E\left[n\bar{\tilde{\varepsilon}}^2\right] = 0$$

since $\bar{\tilde{\varepsilon}}\equiv n^{-1}\left[\sum_{i=1}^{n}\tilde{\varepsilon}_i\right]=0$. The application of Lemma 6.2 is also required to show that

$$E_n(n^{-1}W_3^*) = E\left[n^{-1}\tilde{\varepsilon}^T\tilde{\varepsilon}\right] = F(n^{-1}\hat{W}_3) + O(n^{-1}).$$

To prove the third statement, one needs to apply Lemma 6.1, and obtains

$$E_n(n^{-1}W_1^{*}) = E\left[b_1\tilde{\varepsilon}^T\tilde{\varepsilon} + b_2\tilde{\varepsilon}^T\Big[E(n,n)\Big]\ \tilde{\varepsilon}\right]$$

$$= E\left[b_1\tilde{\varepsilon}^T\tilde{\varepsilon} + b_2\bar{\tilde{\varepsilon}}^2\right]$$

$$= E\left[b_1\tilde{\varepsilon}^T\tilde{\varepsilon}\right] = E(n^{-1}\hat{W}_1) + O(n^{-1}) \ ,$$

in which $\bar{A}^T\bar{A} = \{a_{hi}\}$, $b_1 = n^{-1}\left(\sum_{h=1}^{n} a_{hh}\right)$ and $b_2 = n^{-2}\left(\sum_{\substack{h,i \\ i\neq h}}^{n} a_{hi}\right)$.

For the proofs of both the first and fourth statements, slight variants of Lemma 6.1 are required and these modifications are as follow. Let $r = h + \delta$, then

$$\text{(i)} \quad \left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{1(j)}^T\bar{A}^T S_{3(j)}\right]p,q\right\} \xrightarrow[J]{a.s.} \begin{cases} \left\{\sum_{h=1}^{n-2}\left[n^{-1}a_{rh}\right] + \left\{\sum_{\substack{h=1 \\ h\neq(i+\delta)}}^{n}\ \sum_{i=1}^{n-1}\left[n^{-2}a_{hi}\right]\right\}\right. & \text{when } p=q \\ \left\{\sum_{\substack{h=1 \\ h\neq(i+\delta)}}^{n}\ \sum_{i=1}^{n-1}\left[n^{-2}a_{hi}\right]\right\} & \text{when } p\neq q \end{cases}$$

in which $\delta = 2$,

$$\text{(ii)} \quad \left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{1(j)}^T\bar{A}^T S_{4(j)}\right]p,q\right\} \xrightarrow[J]{a.s.} \begin{cases} \left\{\sum_{h=1}^{n-1}\left[n^{-1}a_{rh}\right] + \left\{\sum_{\substack{h=1 \\ h\neq(i+\delta)}}^{n}\ \sum_{i=1}^{n-1}\left[n^{-2}a_{hi}\right]\right\}\right. & \text{when } p=q \\ \left\{\sum_{\substack{h=1 \\ h\neq(i+\delta)}}^{n}\ \sum_{i=1}^{n-1}\left[n^{-2}a_{hi}\right]\right\} & \text{when } p\neq q \end{cases}$$

in which $\delta = 1$.

Upon applications of the above moditications,

$$E_n(n^{-1}V_1^{*}) = E\left[b_3\tilde{\varepsilon}^T\tilde{\varepsilon} + b_4\tilde{\varepsilon}^T\Big[E(n,n)\Big]\tilde{\varepsilon}\right]$$

$$= E\left[b_3\tilde{\varepsilon}^T\tilde{\varepsilon}\right] = E(n^{-1}\hat{V}_1) + O(n^{-1})$$

and

$$E_n(n^{-1}W_1^*) = 2E\left[b_5\tilde{\varepsilon}^T\tilde{\varepsilon} + b_6\tilde{\varepsilon}^T\left[E(n,n)\right]\tilde{\varepsilon}\right]$$

$$= 2E\left[b_5\tilde{\varepsilon}^T\tilde{\varepsilon}\right] = E(n^{-1}\hat{W}_2) + O(n^{-1}),$$

in which $b_3 = n^{-1}\left[\sum_{h=1}^{n-2} a_{(h+2),h}\right]$, $b_4 = \left\{\sum_{\substack{h=1\\ h\neq(l+2)}}^{n}\sum_{l=1}^{n-1}\left[n^{-2}a_{hl}\right]\right\}$,

$$b_5 = n^{-1}\left\{\sum_{h=1}^{n-1}\left[a_{(h+1),h}\right]\right\}, \text{ and } b_6 = \left\{\sum_{\substack{h=1\\ h\neq(l+1)}}^{n}\sum_{l=1}^{n-1}\left[n^{-2}a_{hl}\right]\right\}.$$

Q.E.D.

Lemma 6.4: Let $\upsilon_n^* \equiv E_n(\upsilon^*)$ and $\omega_n^* \equiv E_n(\omega^*)$. Note that V_n^* and W_n^* are O(1) and O(n), respectively. Also, let n be finite. Then,

(i) $\upsilon_n^* = \upsilon + O(n^{-1})$,

(ii) $(n^{-1}\omega_n^*) = (n^{-1}\omega) + O(n^{-1})$.

Proof:

The proof is straightforward upon application of Lemma 6.3. *Q.E.D.*

With Lemma 6.4, it now becomes possible to state and prove the following theorem:

Theorem 6.1: Let n be finite. Then,

(i) $E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^*_{(j)}\right]\right\} \xrightarrow[J]{a.s.} E(\hat{\beta}) + O(n^{-2})$,

(ii) $E\left\{\frac{1}{J}\sum_{j=1}^{J}\left[\beta^*_{(j)}-\bar{\beta}^*\right]^2\right\} \xrightarrow[J]{a.s.} V(\hat{\beta}) + O(n^{-2})$,

in which

$$\bar{\beta}^{*} \equiv \left\{ \frac{1}{J} \sum_{j=1}^{J} \left[\beta^{*}_{(j)} \right] \right\}.$$

Proof:

Using the results of Marriott and Pope (1954, p.392), the following approximation is accurate to $O(n^{-2})$;

$$E_n\left[V^{*}/W^{*}\right] = \left\{ \upsilon^{*}_{n}/\omega^{*}_{n} \right\} \left\{ 1 - COV(V^{*}, W^{*}) \left[\omega^{*}_{n} \upsilon^{*}_{n} \right]^{-1} + Var(\omega^{*}) \left[\omega^{*}_{n} \right]^{-2} \right\}.$$

Since the second and third terms are already of $O(n^{-2})$, the focus will be on the ratio $(\upsilon^{*}_{n}/\omega^{*}_{n})$. Upon application of Lemma 6.4, it can easily be shown that

$$(\upsilon^{*}_{n}/\omega^{*}_{n}) = (\upsilon/\omega) + O(n^{-2})$$

and the first statement becomes obvious.

In a similar fashion, a proof of the second statement can also be obtained. *Q.E.D.*

In this section, the mean and variance of the bootstrap estimates of β are examined with the help of some new notation. Two additional lemmas concerning the selection matrices are also given in this section. Lemma 6.1 deals with the quadratic form of a selection matrix, whereas Lemma 6.2 deals with the product of two different selection matrices. Both lemmas are used to examine the mean and variance of the bootstrap estimates of β, and this is done in Lemmas 6.3 and 6.4. The final theorem is given in Theorem 6.1, which states that the difference between the mean of $\hat{\beta}$ and the mean of bootstrap estimates of β is $O(n^{-2})$. The difference between the variance of $\hat{\beta}$ and the variance of bootstrap estimates of β is also $O(n^{-2})$.

6.8 Conclusion

In this chapter, attention is focused on $\hat{\beta}$, the least-squares estimate of the parameter of the first-order stationary Gaussian process. Unlike most least-squares parameter estimates, the variance of $\hat{\beta}$ depends on β, which is the unknown parameter. It is important to take this into account when constructing confidence intervals for β. When this is not taken into account, confidence intervals of β will not be reliable, especially when $0.5 \le \beta \le 0.9$. This is true for both the conventional and bootstrap confidence intervals.

Conventional confidence intervals are generally not reliable and should be avoided when $\beta > 0.5$. On the other hand, bootstrap confidence intervals provide fairly good approximations to the exact confidence interval of β, when $0 \le \beta \le 0.9$ and when $n \ge 20$. When $n=10$, the bootstrap approximation is fairly good when $0 \le \beta < 0.9$, but it is poor for the case when $\beta = 0.9$, especially in the lower tail. Thus, bootstrap confidence intervals should be used with care when $n \le 10$.

The simulation results show that bootstrap confidence intervals are best used for constructing ninety percent confidence intervals. For ninety-five percent confidence intervals, the bootstrap confidence intervals are still better than conventional confidence intervals, especially when the parameter value is close to unity. The simulation results also indicate that bootstrap confidence intervals should not be used for constructing eighty percent confidence intervals. This suggests that narrower bootstrap confidence intervals should also be avoided. However, these confidence intervals are seldom needed, and can be replaced by conventional confidence intervals in empirical applications.

When $\beta=\beta_0$ and β_0 is some known constant, bootstrap distributions of $\hat{\beta}_c$ with β_0 as the location parameter can be reasonably good approximations to the exact distribution of $\hat{\beta}_c$. Judging from the simulation results, the bootstrap method may be suggested as a simpler alternative method to the procedure based on ERA's for approximating the exact distribution of $\hat{\beta}$. Although the simulation results are based on a stationary Gaussian process, the general results can be extended to include stationary processes which are not Gaussian, except for the case when β is zero. When $\beta=0$, the normal approximation will be poor if the underlying distribution is significantly different from the normal distribution. This is an additional reason favouring application of the bootstrap method.

In this chapter, β is restricted to lie between zero and 0.9 because for $0.9<\beta\le1.0$, the proportion of $\hat{\beta}$ exceeding unity will be large. When $\hat{\beta}>1.0$, the estimate of its variance given by $(1-\hat{\beta}^2)$ will be negative, and the t-statistic given by $\left[\hat{\beta}(1-\hat{\beta}^2)^{-1/2}\right]$ will take on imaginary values. To avoid this problem, $\hat{\beta}$ is assigned the value 0.99 whenever $\hat{\beta}\ge0.99$. When $0\le\beta<0.9$, this restriction poses no problem because the proportion of $\hat{\beta}\ge0.99$ will be small. However, this restriction becomes a serious problem when $0.9<\beta\le1.0$, especially when $\beta=1.0$.

Also, in the chapter, a theorem is given on the mean and variance of bootstrap estimates of β. It is shown that the difference between the mean of $\hat{\beta}$ and the mean of the bootstrap estimates of β is $O(n^{-2})$. It is also shown that the same $O(n^{-2})$ applies to the difference between the variance of $\hat{\beta}$ and the variance of the boctstrap estimates of β. This is done with the help of two additional lemmas concerning the selection matrices, which are also obtained in this chapter.

To test for unit roots in an AR(1) model, it is necessary to know the distribution of $\hat{\beta}$ for $\beta=1$. A bootstrap distribution of $\hat{\beta}$ can be obtained by setting $\beta=1$, and by bootstrapping the regression residuals to obtain estimates of β. This is a direct method whose results are inferior, when compared to the results obtained by using an alternative method. [C.f. Chapter 4]. This is done by obtaining the bootstrap distribution of

$$\hat{\tau} = s^{-1}\left\{\sum_{t=2}^{n}\left[y_{t-1}^{2}\right]\right\}^{1/2}(\hat{\beta}-1)$$

in which

$$s^{2} = (n-2)^{-1}\left\{\sum_{t=2}^{n}\left[y_{t}-\hat{\beta}y_{t-1}\right]^{2}\right\}.$$

A detailed explanation of the statistic $\hat{\tau}$ can be found in Fuller (1976, p.372). A table for the distribution of $\hat{\tau}$, when $\beta=1$ and when the errors are normally distributed, is also available in Fuller (1976, p.373). When the errors are not normal, the distribution of $\hat{\tau}$ is generally unknown, and the better way to approximate this distribution is by bootstrapping. It is noted here that the method based upon ERA's cannot be used in this case, because this method requires that the error distribution be known. On the other hand, the bootstrap method does not impose this requirement.

At this point, it is important to note that at $\beta=1$ the usual asymptotics do not apply. [See e.g., White (1958, 1959).] In particular, assumptions A.2.1 and A.2.2 will be violated when $\beta=1$. As noted earlier in this chapter, $\hat{\beta}$ has a degenerate limiting distribution unless suitably normalized. Also, the usual t-statistics do not apply in this case and the usual bootstrap procedure would be inappropriate. However, this does not prevent one from applying the bootstrap to this problem. One alternative method is to obtain the bootstrap distribution of $\hat{\tau}$ by assuming apriori that $\beta=1$. For continuity in the limiting distribution of $\hat{\tau}$, one may apply the continuous normalization factor suggested by Evans and Savin (1984). This should be a subject of future research and no claim is made here about the accuracies of bootstrap approximations of the exact distribution of $\hat{\tau}$.

One other alternative method is the method of jackknife, which is also a distribution-free non-parametric method. For its application to a regression model, a good reference can be found in Hinkley (1978). For future work, it may be useful to compare the relative performances of the bootstrap and jackknife techniques. This is possible only when n is large.

CHAPTER SEVEN

SUMMARY AND DISCUSSION

7.1 Introduction

This dissertation provides a systematic study of bootstrap estimates of regression coefficients. It examines the moments of these bootstrap estimates and compares the bootstrap confidence intervals with conventional intervals. It also studies how the use of different types of regression residuals for bootstrapping affects these bootstrap estimates. In particular, it looks into how the moments of bootstrap estimates are affected, and it also looks into how to obtain a more accurate bootstrap confidence interval. The regression residuals studied in this dissertation are the OLS, inflated OLS, BLUS and Stine's (1985) residuals.

In the literature, two main approaches are used to investigate the properties of bootstrap estimates. The first approach uses asymptotic theory, while the second makes inferences from Monte Carlo simulations. The current approach is different from both existing approaches. It first examines the finite sample moments of bootstrap estimates, and then infers from these moments, the characteristics of bootstrap distributions of the OLS estimates. The derivations of the finite sample results are made possible with the use of a selection matrix. With the selection matrix, one is able to obtain exact finite sample results, without resorting to an extensive use of Monte Carlo simulations. It is a major innovation of this dissertation. The present approach is also able to reproduce some of the existing asymptotic results concerning bootstrap estimates.

The selection matrix is defined in Chapter 2, and four lemmas concerning its properties are also given in this chapter. Chapter 2 then focusses on the first two moments of bootstrap estimates, when OLS, inflated OLS, BLUS and Stine's residuals are used for bootstrapping. The higher moments of these bootstrap estimates are examined in Chapter 3. A relationship between the error moments and moments of OLS estimates of the regression coefficients is established in Chapter 3. A similar relationship is also established in this chapter, for the sample moments of regression residuals used for bootstrapping and moments of bootstrap estimates of the regression coefficients. Chapter 4 deals mainly with the bootstrap confidence intervals, and its objective is to find the most suitable method for constructing bootstrap confidence intervals. The approach used here is also different from the existing approaches, in that both Edgeworth expansions and sample mements of bootstrap estimates of the regression coefficients are used to compare bootstrap confidence intervals with conventional intervals. Nevertheless, in Chapters 2, 3 and 4, attention is restricted to the case of a simple linear regression model. An old problem in multiplicative models is studied, in the bootstrap context, in Chapter 5. This problem concerns the estimation of the constant term. First, an unbiased estimate of this term has to be obtained. Secondly, a confidence interval for the constant term has to be constructed. Lastly, in Chapter 6, the problem of constructing a confidence interval for the AR(1) parameter is examined.

7.2 Some Properties of the Selection Matrix

The selection matrix plays an important role in this dissertation. Let $S_{(j)}$ be an mxn selection matrix, and let A and B be real nxn and mxm matrices, respectively, of finite elements. Then, some of the main results concerning the properties of the selection matrix are:

Lemma 2.1: $$\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}\right]\right\} = n^{-1}E(m,n).$$

Lemma 2.4: $$\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)}AS^{T}_{(j)}\right]\right\} = \left\{c_1I_m+c_2\left[E(m,m)-I_m\right]\right\},$$

in which $c_1 = n^{-1}\left[\sum_{i=1}^{n}a_{ii}\right]$ and $c_2 = n^{-2}\left[\sum_{h,i}^{n}a_{hi}\right]$.

Lemma 6.1 $$\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{(j)}BS_{(j)}\right]\right\} = \left\{c_3I_n+c_4E(n,n)\right\},$$

in which $c_3 = n^{-1}\left[\sum_{i=1}^{m}b_{ii}\right]$ and $c_4 = n^{-2}\left[\sum_{i\neq h}^{m}b_{hi}\right]$.

Note that both Lemmas 2.2 and 5.1 are special cases of Lemma 6.1, when m=n. In addition, Lemma 2.2 sets $B=I_n$. When $A=I_n$, Lemma 2.4 becomes Lemma 2.3. The case for two different selection matrices can be found in Lemma 6.2. In particular, let both $S_{p(j)}$ and $S_{q(j)}$ be two different mxn selection matrices. Then,

Lemma 6.2: $$\lim_{J\to\infty}\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S^{T}_{p(j)}S_{q(j)}\right]\right\} = \left[mn^{-2}\right]E(n,n) \text{ when } p\neq q.$$

When p=q, Lemma 6.2 becomes Lemma 2.2.

The application of these results is not restricted to a linear regression model. They can be applied to obtain the means of bootstrap estimates of sample means, sample variances and covariances, and sample correlation coefficients. They can also be applied to obtain the variance of bootstrap estimates of the sample mean. For the variances of bootstrap estimates of sample variances and covariances, and sample correlation coefficients, extensions of the above results are needed. These extensions are left for future research.

7.3 Moments of Bootstrap Estimates in the Linear Regression Context

Asymptotically, Freedman (1981) has shown that OLS residuals can be used for bootstrapping. This is because bootstrapping based upon OLS residuals leads to consistent estimates of the exact moments of OLS estimates of regression coefficients. Let β be the vector of regression coefficients and let $\hat{\beta}$ be its OLS estimate. Then, when OLS residuals are used for bootstrapping, the sample moments of bootstrap estimates of β are consistent estimates of the exact moments of $\hat{\beta}$. However, the biases in the sample moments of bootstrap estimates of β can be substantial, especially when the sample size is small and when the number of regression coefficients is large.

It is now common knowledge among many authors that OLS residuals are not suitable for bootstrapping when the sample size is small. However, most of the inferences come from Monte Carlo simulations. Thus, the cause of the problem(s) associated with the use of OLS residuals for bootstrapping is still unknown. One conjecture is that the problem stems from the fact that the OLS residuals do not have a scalar variance-covariance matrix. [See e.g., Stine (1985)]. This conjecture is proven to be incorrect in Chapter 2. It is demonstrated in Chapters 2 and 3, that the problem with using OLS residuals for bootstrapping is due to the fact that the sample moments of these residuals underestimate the exact moments of $\hat{\beta}$. Thus, when one is interested in only the first two moments of bootstrap estimates of β, the conjecture of Freedman and Peters (1984a) is correct. Freedman and Peters suggest that the OLS residuals be transformed, prior to bootstrapping, either by the BLUS procedure or by multiplying the OLS residuals by the factor $\left\{n(n-K)^{-1}\right\}^{1/2}$. It is shown in Theorems 2.12 and 2.14, that this conjecture is correct, provided that only the first two moments of bootstrap estimates of β are required. For higher moments of these bootstrap estimates, this conjecture is incorrect. It is shown in Chapter 3, that when the errors are normally distributed, the use of BLUS residuals for bootstrapping may lead to better estimates of the fourth moment of $\hat{\beta}$. In this case, the use of either BLUS or inflated OLS residuals for bootstrapping leads to unbiased estimates of the third moment of $\hat{\beta}$.

The main contribution of Chapters 2 and 3 is that they provide a systematic understanding of the finite sample properties of the bootstrap estimates of β. The results of these two chapters enable one systematically to unveil and understand the problems of bootstrapping, especially in the regression context. However, these results are not restricted to the regression problem. Difficulties do exist in applications of the bootstrap to other areas of statistical estimation. The above results can be applied, with possible modifications or extensions, to resolve these difficulties.

7.4 A Note on Bootstrap Confidence Intervals

A general result, concerning bootstrap confidence intervals (BCI's) of an unknown parameter, suggests that these intervals admit errors of $O(n^{-1})$. This result is given by Efron (1979, 1985), Singh (1981), and Abramovitch and Singh (1985). It applies only to a special case in the regression context, and this happens when the true errors are observable. Otherwise, this result is not applicable to BCI's of β. This is demonstrated in Chapter 4. Depending on the type of regression residuals used for bootstrapping and depending on the method of constructing these BCI's, the accuracies of these intervals vary; they can admit errors as small as $O(n^{-3})$ or as large as $O(n^{-1})$. BCI's which admit errors of $O(n^{-1})$ are of no practical value, since conventional confidence intervals, which are easy to construct, also admit errors of the same order.

It is also shown in Chapter 4, that the ideal BCI is based upon the bootstrap distribution of t_2, where $t_2=\left[(\hat{\beta}-\beta)s_{\beta}^{-1}\right]$ and s_{β}^{2} is an unbiased estimate of the variance of $\hat{\beta}$. Either the OLS or BLUS residuals can be used for obtaining the bootstrap distribution of t_2, depending on the sample size and on the number of regression coefficients. When the sample size is small and when there are many coefficients in a regression model, BLUS residuals should be used. On the other hand, when the sample size is large and when there are a few coefficients to be estimated, OLS residuals should be used. The use of either OLS or inflated OLS residuals leads to the same bootstrap distribution of t_2.

The naive BCI, obtained by ordering the bootstrap estimates of β, should be avoided because it can be extremely short when compared with the exact confidence interval of β, especially when the sample size is small. This problem becomes more serious as the number of regression coefficients increases, while holding the sample size constant.

7.5 Consequences of Using OLS Residuals for Bootstrapping

Since OLS residuals are easy to compute, an unsuspecting researcher is likely to use these residuals for bootstrapping. This researcher is also likely to use the naive BCI's for constructing confidence intervals of β. Marais (1984) and Veall (1987) are two recent examples.

In Marais (1984), which studies the two parameter capital asset pricing model, the consequence is not serious. This is because when the sample size is 50, the variance of bootstrap estimates of β underestimates the variance of $\hat{\beta}$ by 4 percent; when the sample size is 200, this bias is only 1 percent. This is evident from Theorems 2.9 and 2.10. On the other hand, the consequence is serious in Veall (1987). In this example, there are two models of peak electricity demand. The first model has two parameters, while the second model has four. For both models, the sample size is 20. As is evident from Theorems 2.9 and 2.10, the variance of bootstrap estimates of β underestimates the variance of $\hat{\beta}$ by 10 and 20 percent in the first and second models, respectively.

Consequently, in Veall's (1987, pp.210-211) Table 1, the correct values are much smaller (larger) than those reported for the lower (upper) tails. Improvements can be made to the results in Marais' (1984, p.48) Table 4, especially for n=50. This can easily be accomplished by using inflated OLS residuals for bootstrapping. On the other hand, this is not adequate in Veall's case. The proper approach, then, is to use the bootstrap distribution of t_2 for the construction of BCI's of peak electricity demands.

7.6 The Role of Bootstrapping in a Multiplicative Model

In double-logarithmic models, which are commonly used in economics, an important difficulty is the estimation of the constant term and its standard error, especially when the errors are multiplicative. Let this constant term be B. When the error distribution is lognormal with unit mean and finite variance, this difficulty is partially resolved. This is due to fact that an unbiased estimate of B can be obtained by using Bradu and Mundlak's (1970) estimator. It is shown in Chapter 5, that the original Bradu-Mundlak estimator has a downward bias. Consequently, an unbiased estimator is duly given in this chapter. This estimator is denoted by $\hat{B}_2$. However, it remains difficult to obtain a confidence interval for B. For this reason, BCI's are suggested here as practical alternatives.

It is shown in Chapter 5, that the variance of bootstrap estimates of B overestimates the actual variance of the $\hat{B}_2$, when the bootstrap estimates are obtained in a similar fashion as $\hat{B}_2$. Consequently, one should not use the naive BCI's for constructing confidence intervals of B. The consequence of using the naive BCI's here is much more serious, as compared to the case of linear regression coefficients. Alternative BCI's are proposed in this chapter, and the one based upon the bootstrap t-distribution is found to have the best performance, in terms of its empirical coverage. It is the only confidence interval whose coverage is relatively well-balanced. However, it does have a problem, which is the possibility of obtaining an open upper bound for B.

One other estimate of B, which is suggested by Srivastava and Singh (1989), is also examined in Chapter 5. This estimate is simpler to compute than $\hat{B}_2$, but it has a bias, whose magnitude depends on the variance of the error term. However, this bias is easy to correct, when the error distribution is lognormal. When the lognormality assumption is violated, both estimates become unreliable and better estimates need to be constructed. Two feasible alternatives are the jackknife and bootstrap estimates considered in Chaubey and Sim (1988). Thus, the role of the bootstrap is not restricted to constructing confidence intervals of B. Its role can be enhanced to obtain unbiased estimates of B, especially when the error distribution is not lognormal and when the error variance is large.

7.7 Bootstrap Confidence Intervals of the AR(1) Parameter

In Chapter 6, the focus is on the least-squares estimate of the AR(1) parameter. Let β be the AR(1) parameter, and let its least-squares estimate be $\hat{\beta}$. It is demonstrated in this chapter that, for constructing confidence intervals of β, one should take into account the fact that the variance of $\hat{\beta}$ depends upon β. It is also shown in this chapter, that conventional confidence intervals of β are unreliable in the tails, especially when $n \le 20$. Bootstrap confidence intervals are suggested here as simple alternatives. Another alternative is the method based upon ERA's introduced by Phillips (1983, 1984). However, this method is rather cumbersome to use, and it also requires that the distribution of the underlying process be known. On the other hand, the bootstrap method is simple to use, and it is a distribution-free method.

With regard to bootstrap confidence intervals, the following are in order. First, for constructing confidence intervals of B, one should use the bootstrap t-distribution, and one should also account for the fact that the variance of $\hat{\beta}$ depends upon β. Secondly, when $n \geq 20$ and when $0 \leq \beta \leq 0.9$, BCI's provide fairly good approximations to the exact confidence interval. When $n=10$, this observation is restricted to the case when $0 \leq \beta \leq 0.9$. Thirdly, BCI's are most reliable when used for constructing ninety percent confidence intervals. For ninety-five percent confidence intervals, BCI's are still better than conventional intervals, especially when β is close to unity.

A bootstrap theorem is also given, in Chapter 6, on the mean and variance of bootstrap estimates of β. This is done with the aid of two additional lemmas concerning the properties of selection matrices. This theorem and its corresponding lemmas are not restricted to the current example. They can also be used to extend the current application of the bootstrap to the Durbin-Watson statistic, and statistical tests like the F, LM, Wald and unit root tests.

It is also shown in Chapter 6, that bootstrap approximations to the exact distribution of $\hat{\beta}$, when β is some known constant, are reasonably good. Thus, it may be suggested as a simple alternative to the method based upon ERA's for approximating the exact distribution of $\hat{\beta}$. This can be very useful in the case when one is testing for unit roots, especially when the underlying distribution is unknown.

7.8 Applications of Bootstrapping in Econometrics

The bootstrap can be applied to a general class of econometric problems. This is demonstrated by the growing literature on the application of bootstrap in economics. However, the bulk of the literature consists mainly of applications and Monte Carlo results. Only Freedman (1981, 1984) gives asymptotic results for regression models; namely, single equation and two stage least-squares models. Earlier applications of the bootstrap to a system of linear seemingly unrelated regression equations can be found in Freedman and Peters (1984a, 1984b) and Korajcyk (1985). Other earlier applications to linear regression models include Peters (1983), Daggelt and Freedman (1984) and Marais (1984). Flood (1985) applies the bootstrap to a system of seemingly unrelated Tobit equations.

Recently, bootstrapping has been applied to many areas in economics. For examples, see Taylor *et al* (1986), Green *et al* (1987) and Vinod and Raj (1988). Other recent examples can also be found in Economics Letters Volume 22. The applied areas include demand homogeneity, standard errors for elasticities and economic issues in system divestiture (Vinod and Raj, 1988). Applications of bootstrap to forecasting elasticity demand can be found in Bernard and Veall (1987) and Veall (1987).

Among the theoretical applications, Hsu *et al* (1986) use the bootstrap as a bias reduction method for two stage least-squares estimates. Application of bootstrapping in regression models with dependent errors is also an interesting area. Recent studies in this area include Kiviet (1984), Veall (1986) and Prescott and Stengos

(1987). One other interesting problem in econometrics is the case when the design matrix is ill-conditioned. Recent examples of the application of bootstrapping in this area can be found in Delaney and Chatterjee (1986, 1987), and Nebebe and Sim (1989).

7.9 Conclusion

This chapter summarizes the main results obtained in this dissertation. It also presents a discussion of some of these results. The main contributions of this dissertation are:

(1) It introduces a new method, based upon the selection matrix, for studying the finite sample properties of bootstrap estimates. The selection matrix itself is a new innovation, at least in the bootstrap context.

(2) It locates the problem associated with bootstrap estimates of regression coefficients and, consequently, suggests the appropriate measure for correcting this problem.

(3) It explains why naive bootstrap confidence intervals are unreliable, especially when the sample size is small and when the parameter to be estimated is other than a simple linear regression coefficient. Hence, more reliable bootstrap confidence intervals are also suggested.

(4) The results obtained in this dissertation are not restricted to the current examples, and have general applications to a wider class of statistical and econometric problems.

<u>**References**</u>

Abramovitch, L. and K. Singh (1985). "Edgeworth corrected pivotal statistics and the bootstrap." <u>Annals of Statistics</u>, 13(1), 116-132.

Anderson, T.W. (1971). <u>The Statistical Analysis of Time Series</u>. [New York, John Wiley and Sons, 1971].

Babu, G. J. and K. Singh (1983) "Inference on means using the bootstrap." <u>Annals of Statistics</u>, 11(3), 999-1003.

Barlett, M.S. (1935). "The effect of non-normality on the t distribution." <u>Biometrika</u>, 21, 259-296.

Bartlett, M.S. (1946). "On the theoretical specification and sampling properties of autocorrelated time-series." <u>Supplement to the Journal of Royal Statistical Society</u>, 8, 27-41.

Beran, R. (1984). "Jackknife approximations to bootstrap estimate." <u>Annals of Statistics</u>, 12, 101-118.

Bernard, J.T. and M.R. Veall (1987). " The probability distribution of future demand." <u>Journal of Business and Economic Statistics</u>, 5(3), 417-424.

Bickel, P.J. and D.A. Freedman (1981). "Some asymptotic theory for the bootstrap." <u>Annals of Statistics</u>, 9, 1196-1217.

Bickel, P.J. and J.A. Yahav (1988). "Richardson extrapolation and the bootstrap." <u>Journal of the American Statistical Association</u>, 83, 387-393.

Bodkin, R.G. and L.R. Klein (1967). "Nonlinear estimation of aggregate production functions." <u>Review of Economics and Statistics</u>, 49, 28-44.

Bradu, D. and Y. Mundlak (1970). "Estimation in log-normal linear models." Journal of the American Statistical Association, 65, 198-211.

Bronshtein, I. N. and K. A. Semendyayev (1985). Handbook of Mathematics. English Translation by K. A. Hirsch. [Frankfurt; Verlag Harr: Deutsch].

Chaubey, Y.P. and A.B. Sim (1988). "A Comparison of Bootstrap, Jackknife, Maximum Likelihood and Some other methods for estimation in Multiplicative models when the errors are inverse-Gaussian." Mimeographed. August, 1988. Department of Mathematics and Statistics. Concordia University.

———————— and B. Singh (1988). "Almost unbiased estimation in multiplicative models." Metrika, 35, 13-28.

Cressie, N. (1980). "Relaxing assumptions in the one-sample t-test." Australian Journal of Statistics, 22, 143-153.

Daggett, R.S. and D.A. Freedman (1984). "Econometrics and the Law: A case study in the proof of antitrust damages." Technical Report No. 23, University of California, Berkeley, Dept. of Statistics.

Daniels, H.E. (1954). "Saddle point approximations in statistics." Annals of Mathematical Statistics, 25, 631-650.

Davison, A.C., D.V. Hinkley and E. Schechtman (1986). "Efficient bootstrap simulation." Biometrika, 73(3), 555-566.

Delaney, N.J. and S. Chatterjee (1986). "Use of the bootstrap and Cross-Validation in Ridge Regression." Journal of Business and Economic Statistics, 4(2), 255-262.

———————— (1987). "A note on collinearity, bootstrapping, and cross-validation." Economics Letters, 25, 335-338.

Dhrymes, P.J. (1962). "On devising unbiased estimators for the parameters of the Cobb-Douglas production function." Econometrica, 30, 297-304.

Diciccio, T. and R. Tibshirani (1987). "Bootstrap Confidence Intervals and Bootstrap approximations." Journal of the American Statistical Association, 82, 163-170.

Durbin, J. (1986). "Approximate distributions of Student's t-statistics for autoregressive coefficients caluculated from regression residuals." Journal of Applied Probability, 23A, 173-185.

Efron, B. (1979). "Bootstrap methods: Another look at the jackknife." Annals of Statistics, 7, 1-26. 26.

———— (1985). "Bootstrap confidence intervals for a class of parametric problems." Biometrika, 72(1), 45-58.

———— (1987). "Better Bootstrap Confidence intervals" (With discussions). Journal of the American Statistical Association, 82, 171-200.

Evans. G.B.A. and N.E. Savin (1981). "Testing for Unit Roots: 1" Econometrica, 49, 753-759.

———————— (1984). "Testing for Unit Roots: 2" Econometrica, 52, 1241,1269.

Evans, J.G. and S.A. Shaban (1976). "Point estimation in multiplicative models." Econometrica, 44, 467-73.

Finney, D.J. (1951). "On the distribution of a variate whose logarithm is normally-distributed." Supplement to the Journal of the Royal Statistical Society, 7, 155-161.

Flood, L. (1985). "Using bootstrap to obtain standard errors of system Tobit coefficients." Economics Letters, 19, 339-342.

Frangos, C.C. (1987). "An updated bibliography on the jackknife method" Communications in Statistics-Theory and Methods, 16(6), 1543-1584.

Freedman, D.A. (1981). "Bootstrapping regression models". Annals of Statistics, 9, 1218-1228.

Freedman, D.A. (1984). "On bootstrapping two-stage least-squares estimates in stationary linear models". Annals of Statistics, 12(3), 827-842.

Freedman, D.A. and S.C. Peters (1984a). "Bootstrapping a regression equation: Some empirical results". Journal of the American Statistical Association, 79, 97-106.

Freedman, D.A. and S.C. Peters (1984b). "Bootstrapping an econometric model: Some empirical results". J. Bus. Econ. Statist., 2, 150-158.

Fuller, W. (1976). Introduction to Statistical Time Series [New York, John Wiley and Sons].

Fuss, M., D. McFadden and Y. Mundlak (1978). "A survey of functional forms in the economic analysis of production." in M. Fuss and D. McFadden (eds.), The Econometric Analysis of Production [Amsterdam, North-Holland], 219-268.

Gayen, A.K. (1949). "The distribution of "Student's" t in random samples of any size drawn from non-normal universes." Biometrika, 36, 353-369.

Geary, R.C. (1936). "The distribution of Student's ratio for non-normal samples." Supplement to the Journal of the Royal Statistical Society, 3, 178–184.

Goldberger, A S. (1968). "The interpretation and estimation of Cobb-Douglas functions." Econometrica, 36, 464-72.

Green, R., W. Hahn and D. Rocke (1987). "Standard errors for elasticities: A comparison of bootstrap and asymptotic standard errors." Journal of Business & Economic Statistics, 5(1), 145-149.

Guilkey, D.K., C.A.K. Lovell and R.C. Sickles (1983). "A comparison on the performance of three flexible functional forms." International Economic Review, 24(3), 591-616.

Hall, Peter (1983). "Inverting an Edgeworth expansion." Annals of Statistics, 11(2), 569-576.

——————— (1986). "On the bootstrap and confidence intervals." Annals of Statistics, 14, 1431-1452.

——————— (1987). "On the bootstrap and continuity correction." Journal of the Royal Statistical Society, Series B, 49(1), 82-89.

Hampel, F.R. (1971). "A general qualitative definition of robustness." Annals of Mathematical Statistics, 42, 1887-1896.

——————— (1973). "Some small sample asymptotics." Proc. Prague Symposium on Asymptotic Statistics, Prague.

Helen, D.M. (1968). "A note on log-linear regression." Journal of the American Statistical Association, 63, 1034-8.

Hinkley, D. (1976). "On jackknifing in unbalanced situations". Technical Report No. 22, Division of Biostatistics, Stanford University.

Hinkley, D. (1977). "On jackknifing in unbalanced situations." Technometrics, 19, 285-292.

Hsu, Y-S., K.N. Lau, H.G. Fung and E.F. Ulveling (1986). "Monte Carlo studies on the effectiveness of the bootstrap bias reduction method on 2SLS estimates." Economics Letters, 20, 233-239.

Huber, P.J. (1981). Robust Statistics [New York, John Wiley & Sons].

Jackson, L.F. (1984). "Hierarchic demand and the Engel curve for variety." Review of Economics and Statistics, 66(1), 8-15.

Kadiyala, K.R. (1970). "Testing for the independence of regression disturbances." Econometrica, 38(1), 97-117.

Kendall, M.G. (1944). "On autoregressive time-series." Biometrika, 33, 105.

Kendall, M.G. and A. Stuart (1977). The Advanced Theory of Statistics. Volume 1. Fourth Edition. [London: Charles Griffin & Company].

————————, A. Stuart and K.K. Ord (1983). The Advanced Theory of Statistics, Volume 3. [London, Charles Griffin & Co., 1983].

Kiviet, J.F. (1984). "Bootstrap inference in lagged-dependent variable models." (Faculty of Actuarial Science and Econometrics, University of Amsterdam, Amsterdam).

Koerts, J. (1967). "Some further notes on disturbance estimates in regression analysis." Journal of the American Statistical Association, 62, 169-183.

Korajcyk, R.A. (1985). "The pricing of future contracts for foreign exchange." Journal of Political Economy, 93, 346-368.

Loh, W.Y. (1987). "Calibrating confidence coefficients." Journal of the American Statistical Association, 82, 155-162.

Lubrano, M., R.G. Perse and J-F. Richard (1986), "Stability of a U.K. Money Demand Equation: A Bayesian Approach to testing exogeneity." Review of Economic Studies, 53, 603-634.

MaCurdy, T.E. and J.H. Pencavel (1986). "Testing between competing models of wage and employment determination in Unionized markets." Journal of Political Economy, 94(3, part2), S3-S39.

Marais, M.L. (1984). "An application of the bootstrap method to the analysis of squared, standardized market model prediction errors." Journal of Accounting Research, 22 supplement, 34-54.

Maritz, J.S. (1981). Distribution-Free Statistical Methods. [New York, Chapman and Hall].

Marriott, F.H.C. and J.A. Pope (1954). "Bias in the estimation of autocorrelations" Biometrika, 41, 390-402.

Miller, R.G. (1974). "An unbalanced jackknife." Annals of Statistics, 2, 880-891.

Nebebe, F. and A.B. Sim (1989). " Improved ridge estimators based on the bootstrap and the jackknife versus empirical Bayes estimators: Some Monte Carlo evidence." Unpublished manuscript, May 1989. Concordia University.

Neudecker, H. (1969). "A note on BLUS estimation." Journal of the American Statistical Association, 64, 949-952.

Parr, W.C. (1983). "A note on the jackknife, the bootstrap and the delta method estimators of bias and variance." Biometrika, 70, 719-722.

Patel, J.K. and C.B. Read (1982). Handbooks of the Normal Distribution. [New York, Marcel Dekker Inc.].

Pearson, E.S. (1929). "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations." Biometrika, 21, 259-296.

Peters, S.C. (1983). Bootstrapping a Regression Equation: Some Empirical Results, unpublished Ph.D. Dissertation, Stanford University, Dept. of Statistics.

Phillips, P.C.B. (1978). "Edgeworth and Saddlepoint approximations in the first order non-circular autoregression." Biometrika, 65, 71-78.

———— (1983). "ERA's: A new approach to small sample theory." Econometrica, 51, 1505-1526.

———— (1984). "Finite sample econometrics using ERA s." Journal of Japan Statistical Society, 14, 107-124.

———— (1985). "The exact distribution of SUR estimator." Econometrica, 53, 745-756.

———— and P.C. Reiss (1986). "Testing for serial correlation and unit roots using a computer function routine based on ERA's" in Advances in Statistical Analysis and Statistical Computing, Vol. 1 edited by Roberto Mariano. [Jai Press, 1986].

———— and P. Perron (1988). "Testing for a unit root in time series regression." Biometrika, 75(2), 335-346.

Pollak, R.A. and T.J. Wales (1980). "Comparison of the quadratic expenditure system and translog demand systems with alternative specifications of demographic effects." Econometrica, 48(3), 595-612.

Prescott, D. and T. Stengos (1987). "Hypothesis testing in regression models with AR(1) errors and a lagged dependent variable: Bootstrapping and artificial regression compared." Economics Letters, 24, 237-242.

Quenouille, M. (1949). "Approximate tests of correlation in time series." Journal of the Royal Statistical Society., Series B, 11, 68-84.

———————— (1956). "Notes on bias in estimations." Biometrika, 43, 353-360.

Rao, J.N.K. and C.F.J. Wu (1988). "Resampling inference with complex survey data." Journal of the American Statistical Association, 83, 231-241.

Rukhin, A.L. (1986). "Improved estimation in Lognormal Models." Journal of the American Statistical Association, 81, 1046-1049.

Sargan, J.D. (1976). "Econometric Estimators and the Edgeworth Approximations." Econometrica, 44, 421-448.

———————— and Y.K. Tse (1979). "Edgeworth Approximations to the Distributions of various test statistics." Discussion Paper No. A22, London School of Economics, Econometrics Programme, 1979.

Sargan, J.D. and S.E. Satchell (1986). "The theorem of validity for Edgeworth expansions." Econometrica, 54(1), 189-213.

Schemper, M. (1987). "Non parametric estimation of variance, skewess and kurtosis of distribution of a statistic by jackknife and bootstrap techniques." Statistica Neerlandica, 41, 59-64.

Shenton, L.R. and W.L. Johnson (1966). "Moments of a serial correlation coefficient". Journal of the Royal Statistical Society, Series B, 27, 308-320.

Singh, K. (1981). "On asymptotic accuracy of Efron's bootstrap". Annals of Statistics, 9, 1187-1195.

Srivastava, M.S. and B. Singh (1989). "Bootstrapping in multiplicative models." Journal of Econometrics, forthcoming.

Stine, R.A. (1985). "Bootstrap prediction intervals for regression" Journal of the American Statistical Association, 80, 1026-1031.

Strassen, V. (1965). "The existence of probability measures with given marginals." Annals of Mathematical Statistics, 36, 423-439.

Swanepoel, Jan W.H. (1985). "Bootstrap selection procedures based on robust estimators". The Frontiers of Modern Statistical Inference Procedures. American Science Press Inc., 43-58.

Tanaka, K. (1983). "Asymptotic expansions associated with AR(1). model with unknown mean". Econometrica, 51(4), 1221-1231.

Taylor, T.G., J.S. Shonkwiller and H. Theil (1986). "Monte Carlo and bootstrap testing of demand homogeneity." Economics Letters, 20, 55-57.

Teeken, R. and J. Koerts (1972). "Some statistical implications of the log transformations of multiplicative model." Econometrica, 40, 793-819.

Theil, H. (1965). "The analysis of disturbances in regression analysis". Journal of the American Statistical Association, 60, 1067-1079.

Tibshirani, R. (1988). "Variance stabilization and the bootstrap." Biometrika, 75(3), 433-444.

Tukey, J.W. (1958). "Bias and confidence in not quite large samples". Annals of Mathematical Statistics, 29, 614.

Veall, M.R. (1986). "Bootstrapping regression estimators under first-order serial correlation." Economics Letters, 21, 41-44.

——————— (1987). "Bootstrapping the probability distribution of peak electricity demand." International Economic Review, 28(1), 203-212.

Vinod, H.D. and B. Raj (1988). "Economic Issues in Bell System divestiture: a bootstrap application." Applied Statistics, 37(2), 251-261.

White, Halbert (1984). Asymptotic Theory For Econometricians. [New York, Academic Press, 1984].

White, J.S. (1958). "The limiting distribution of the serial correlation in the explosive case." Annals of Mathematical Statistics, 29, 1188-1197.

——————— (1959). "The limiting distribution of the serial correlation in the explosive case II." Annals of Mathematical Satistics, 30, 831-834.

——————— (1961). "Asymptotic expansions for the mean and variance of the serial correlation coefficient". Biometrika, 48, 85-94.

Wu, C.F.J. (1986). "Jackknife, bootstrap and other resampling methods in regression analysis". Annals of Statistics, 14, 1261-1295.

Appendix A

Alternative Proof for Lemma 2.2: Let

$$S^T_{(j)} = \left[S^T_{(j)}, \ldots, S^T_{(j).m}\right]$$

and

$$\left[S^T_{(j)} S_{(j)}\right] = \left[S_{(j)hi}\right] \quad , \ (h,i=1,\ldots,n),$$

$$= \left\{\sum_{r=1}^{m} \left[S^T_{(j).r} S_{(j)r.}\right]\right\}.$$

First note that for each r $(r=1,\ldots,m)$, there are n ways to place one unity in $S_{(j)r.}$. In other words, the elements of $S_{(j)r.}$ can be arranged in n different ways. Consequently, there will be (n^m) ways to arrange the elements in $S_{(j)}$. Of these (n^m) arrangements, there will be

(i) $C_n = 1$ matrix with unity in the first column,

(ii) $({}^mC_{m-1})(n-1)$ matrices with unity in $(m-1)$ positions and 1 zero in the first column,

(iii) $({}^mC_{m-2})(n-1)^2$ matrices with unity in $(m-2)$ positions and 2 zeroes in the first column,

(iv) $({}^mC_1)(n-1)^{m-2}$ matrices with unity in 2 positions and $(m-2)$ zeroes in the first column,

(v) $({}^mC_1)\ (n-1)^{m-1}$ matrices with unity in one position and $(n-1)$ zeroes in the first column.

Secondly, whenever unity is in q positions in the first column of $S_{(j)}$, $S_{(j)11}=q$ for all $q=1,\ldots,m$. Thus,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)11}\right]\right\} \xrightarrow[J]{a.s.} \left[\left\{\sum_{i=1}^{m}\left[({}^{m}C_{i})(n-1)^{m-i}(i)(n^{-m})\right]\right\}\right].$$

However,

$$\left\{\sum_{i=1}^{m}\left[({}^{m}C_{i})(n-1)^{m-i}(i)\right]\right\} = mn^{m-1}$$

and the R.H.S. becomes $(n^{-1}m)$. The same conclusion can be made for the other columns of S . Thus,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)hh}\right]\right\} \xrightarrow[J]{a.s.} (n^{-1}m) \ , \ (h=1,\ldots,m).$$

The proof is then straightforward. *Q.E.D.*

Alternative Proof for Lemma 2.3: Let the mxm matrices

$$\left[S_{(j)}S^{T}_{(j)}\right] = \left[S_{(j)hi}\right] \ , \ (h,i=1,\ldots,n).$$

First note that $S_{(j)hh}=1$ for all $h=1,\ldots,m$. When h and i are different, $S_{(j)hi}=1$ if $S_{(j)h.}=S_{(j)i.}$. Since there are n columns in $S_{(j)}$, this can occur in n ways. Suppose that the unit values are in the first columns of $S_{(j)h.}$ and $S_{(j)i.}$. Then, there will be (n^{m-2}) arrangements for the remaining (m-2) rows. Subsequently, there will be $\left[n(n^{m-2})\right]$ and $\left[n(n-1)(n^{m-2})\right]$ ways of arranging $S_{(j)}$ so that $S_{(j)hi}=1$ and $S_{(j)hi}=0$, respectively. Thus, when $h\neq i$,

$$\left\{\frac{1}{J}\sum_{j=1}^{J}\left[S_{(j)hh}\right]\right\} \xrightarrow[J]{a.s.} \left\{n(n^{m-2})n^{-m}\right\}$$

and the R.H.S. reduces to (n^{-1}) The proof then becomes straightforward. *Q.E.D.*