

BOOTSTRAPPING A TIME SERIES MODEL

By

MUN-SHIG SON

Bachelor of Economics
Sung Kyun Kwan University
Seoul, Korea
1975

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1982

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1984

Thesis
1984D
S 698b
Cop. 2



BOOTSTRAPPING A TIME SERIES MODEL

Thesis Approved:

Don Holbert

Thesis Adviser

W. H. Stewart

William H. Stewart

John D. Rea

Norman D. Durham

Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my appreciation to my major adviser, Dr. Donald Holbert, for suggesting this topic and for all of the time and effort he has devoted to this and other studies. Words cannot express the appreciation I feel for his assistance and his friendship.

Appreciation is also expressed to Dr. David L. Weeks, Dr. William H. Stewart, and Dr. John D. Rea for serving on my advisory committee. I am especially grateful to Dr. William D. Warde and Dr. Michael J. Applegate for replacing Dr. David L. Weeks and Dr. John D. Rea in their absence during my oral examination.

I would also like to thank my dear friend, David Christopher, who directed the writing of the computer program.

I gratefully acknowledge indebtedness for the constant support and encouragement from my uncle and aunt in Osaka, Japan, through the years of my study in the United States. Without their assistance I could not have pursued the degree. I would also like to acknowledge the financial support I received the past four years from the Department of Statistics and the Department of Mathematics.

I would like to dedicate this dissertation in memory of my parents. The devotion and selfless love they provided me with in my formative years was essential to my growth and an inspiration to me in later years. They gave me so much and more than I know.

Special thanks to Mrs. Sandra Ireland and Miss Barbara Newport for

the exceptional job they did in typing my thesis.

Finally, I want to say thanks to my wife, daughter, brother and sisters for their love, strength, sacrifice, endurance, patience, and never-ending confidence in me.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION AND GENERAL STATEMENT OF THE PROBLEM	1
II. REVIEW OF LITERATURE.	6
III. OUTLINE OF THE BOOTSTRAP PRINCIPLE IN GENERAL	10
The Bootstrap.	10
Nonparametric Confidence Intervals via Bootstrap	13
The Bootstrap Estimate of Bias	16
IV. BOOTSTRAPPING A TIME SERIES MODEL	17
Chapter Introduction	17
Numerical Example--Daily Readings of the Viscosity. . . .	19
Conventional Least Squares Estimates	19
Conventional Maximum Likelihood Estimates.	22
Bootstrapping AR(2) Model.	24
Comparison With the Least Squares Estimates	24
Comparison With the Maximum Likelihood Estimates. .	30
V. FORECASTING AN AR(2) MODEL USING THE BOOTSTRAP.	34
Chapter Introduction	34
Conventional Forecasting Procedures.	37
The Bootstrap Forecasting Procedure.	39
VI. QUALITY OF THE BOOTSTRAP ESTIMATES.	46
Chapter Introduction	46
Simulation Procedure for the Trial and Bootstrap	
Experiment	48
Results of the Simulation Experiment	55
Evaluation of Random Number Generators	57
VII. BOOTSTRAPPING AN ECONOMETRIC MODEL.	59
Chapter Introduction	59
Conventional Least Squares Estimates	63
Bootstrapping Oklahoma Unemployment Rate Model	64
Bootstrapping Oklahoma Individual Income Tax Model . . .	71

Chapter	Page
VIII. SUMMARY AND FURTHER CONSIDERATIONS.	75
A SELECTED BIBLIOGRAPHY.	78
APPENDICES	82
APPENDIX A - COMPUTATIONAL DETAILS.	83
APPENDIX B - A SIMPLE MATHEMATICAL EXPLANATION.	85
APPENDIX C - TABLES	88

LIST OF TABLES

Table	Page
I. Daily Readings of the Viscosity of Chemical Product XB-77-5.	20
II. Bootstrap Result for the AR(2) Model (There are 100 Bootstrap Pseudo-Random Samples. Estimation is by Least Squares, Using the First Seed.)	25
III. Computations for the Entries in Table II (There are 100 Bootstrap Pseudo-Random Samples.)	29
IV. Bootstrap Result for the AR(2) Model (There are 100 B/S Pseudo-Random Samples. Estimation is by Maximum Likelihood, Using the First Seed.)	32
V. Bootstrap Forecast Experiment for Model (V-7) (There are 100 B/S Replications. Estimation is by Least Squares, Using the First Seed.)	42
VI. Bootstrap Forecast Experiment for Model (V-7) (There are 100 B/S Replications. Estimation is by Maximum Likelihood, Using the First Seed.)	43
VII. Computations for the Trial and Bootstrap Experiment (There are 100 Trials and 100 B/S Samples for Each Trial.)	52
VIII. Trial and Bootstrap Experiment (There are 100 Trials and 100 B/S Samples for Each Trial. Estimation is by Maximum Likelihood.)	54
IX. Data for the Oklahoma Unemployment Rate Equation	65
X. Bootstrap Result for the Unemployment Rate Equation (There are 200 Bootstrap Replications. Estimation is by Least Squares, Using the First Seed.)	67
XI. Data for the Oklahoma Individual Income Tax Equation	72
XII. Bootstrap Result for the Individual Income Tax Equation (There are 200 Bootstrap Replications. Estimation is by Least Squares, Using the First Seed.)	73

Table	Page
XIII. Bootstrap Result for the AR(2) Model (There are 100 Bootstrap Pseudo-Random Samples. Estimation is by Least Squares, Using the Second Seed.)	89
XIV. Bootstrap Result for the AR(2) Model (There are 100 B/S Pseudo-Random Samples. Estimation is by Maximum Likelihood, Using the Second Seed.)	89
XV. Bootstrap Forecast Experiment for Model (V-7) (There are 100 B/S Replications. Estimation is by Least Squares, Using the Second Seed.)	90
XVI. Bootstrap Forecast Experiment for Model (V-7) (There are 100 B/S Replications. Estimation is by Maximum Likelihood, Using the Second Seed.)	91
XVII. Bootstrap Result for the Unemployment Rate Equation (There are 200 Bootstrap Replications. Estimation is by Least Squares, Using the Second Seed.)	92
XVIII. Bootstrap Result for the Individual Income Tax Equation (There are 200 Bootstrap Replications. Estimation is by Least Squares, Using the Second Seed.)	93

LIST OF FIGURES

Figure	Page
1. Bootstrap Model of the Daily Readings of Viscosity.	27
2. Bootstrap Forecast Model.	41
3. Generating 100 Trials and 100 B/S Pseudo-Random Samples for Each Trial.	49
4. Model of the Oklahoma Unemployment Rate	68

CHAPTER I

INTRODUCTION AND GENERAL STATEMENT OF THE PROBLEM

The bootstrap, a computer-intensive methodology, is a recently introduced statistical technique, which checks the accuracy of the asymptotics, and makes alternative estimates of the standard errors; see Efron (1979a and 1982). The bootstrap method is to resample the original data at hand in an appropriate way, to collect "pseudo-random samples" on which the estimator of interest is computed. More specifically, the theoretical error distribution is approximated by the empirical residual distribution. This bootstrap idea gives us a very useful answer to the question: What can we do if we do not know the form of the observational distribution? The bootstrap avoids the difficulty of finding a large sample approximation to the sampling distribution by replacing the unknown distribution by the empirical distribution. Freedman and Peters (1983a) showed that the bootstrap method was appreciably better than the conventional asymptotics, when applied to a particular finite-sample situation, in the setting of a single econometric equation.

In this paper an attempt is being made to compare the performance of conventional asymptotic estimates of standard error to the performance of a bootstrap procedure in the setting of a time series model. Also this paper is concerned with the way the bootstrap develops standard errors for multi-period forecasts. This paper is the first application

of the bootstrap in a time series setting.

A time series is a collection of observed values generated sequentially in time order. Time series occur in a variety of fields, such as business, economics, sociology, physics and engineering, medicine and meteorology. The methods devised to obtain a concise description of the features of a particular time series process are important. Also a forecasting procedure for the behavior of the series in the future based on a knowledge of the past is important. The methodology developed by Box and Jenkins (1976) represents a popular systematic approach to modeling and forecasting time series. This paper is mainly concerned with estimating standard errors of fitted coefficients, obtained by both the least squares and maximum likelihood estimation procedure, using bootstrap methodology. These standard errors are compared to those obtained using the conventional approach.

In the next few paragraphs, we outline four approaches to the problem of estimating standard errors, in order to illustrate how the bootstrap procedures fit in to the broader framework of standard error estimation. The methods outlined are summarized in the table on page 4.

A general statement of the problem of error estimation is the following: Suppose we observe $X_1 = x_1, X_2 = x_2, \dots, X_i = x_i, \dots, X_n = x_n$, where the X_i are independent and identically distributed according to some probability distribution F . The X_i may be simple real-valued, two-dimensional, or take values in a more complicated space. Let $\hat{\theta} = T(X_1, X_2, \dots, X_n)$ be a statistic. For example, $T(X_1, X_2, \dots, X_n) = \sum_{i=1}^n X_i / n$ or $\text{Med}\{X_i\}$ are statistics.

Let $\sigma[T(X_1, X_2, \dots, X_n)]$ be a measure of precision that we would use if the distribution F were known, for example, $\sigma[T(X_1, X_2, \dots, X_n)] =$

$SD_F(\hat{\theta})$, the standard deviation of $\hat{\theta}$ when $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$. How do we calculate $\sigma[T(X_1, X_2, \dots, X_n)]$?

A. When the distribution F is known

Direct calculation:

$$\begin{aligned}\sigma^2[T(X_1, X_2, \dots, X_n)] &= E_F[T(X_1, X_2, \dots, X_n) - E_F\{T(X_1, X_2, \dots, X_n)\}]^2 \\ &= \int_{\mathbb{R}^n} [T(x_1, x_2, \dots, x_n) - \int_{\mathbb{R}^n} T(y_1, y_2, \dots, y_n) \\ &\quad \cdot F(dy_1) \dots F(dy_n)]^2 F(dx_1) \dots F(dx_n).\end{aligned}$$

The variance $\sigma^2[T(X_1, X_2, \dots, X_n)] = \sigma^2/n$ for the statistic $T(X_1, X_2, \dots, X_n) = \bar{X}$ would be a well known example when the distribution F is $N(\mu, \sigma^2)$. But in many problems of interest, the mathematics is intractable.

Possible approximations when the mathematics is intractable would be:

a) Simulation Method--In order to estimate $\sigma^2[T(X_1, X_2, \dots, X_n)]$, we employ simulation methods. An appropriate number of samples $X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)}$, $j = 1, 2, \dots, N$ is drawn from F repeatedly, and for each sample the statistic $T^{(j)}(X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)}) = T^{(j)}(\tilde{X}^{(j)})$ is computed:

$$\left. \begin{array}{l} X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)} \rightarrow T^{(1)}(\tilde{X}^{(1)}) \\ X_1^{(2)}, X_2^{(2)}, \dots, X_n^{(2)} \rightarrow T^{(2)}(\tilde{X}^{(2)}) \\ \vdots \\ X_1^{(N)}, X_2^{(N)}, \dots, X_n^{(N)} \rightarrow T^{(N)}(\tilde{X}^{(N)}) \end{array} \right\} \frac{1}{N} \sum_{j=1}^N T^{(j)}(\tilde{X}^{(j)}) \xrightarrow[N \rightarrow \infty]{a.s.} E_F T(X_1, X_2, \dots, X_n),$$

Here the symbol " $\xrightarrow[N \rightarrow \infty]{a.s.}$ " stands for convergence almost surely.

Then $\sigma^2[T(X_1, X_2, \dots, X_n)]$ is estimated by $\frac{1}{N-1} \sum_{j=1}^N [T^{(j)}(\tilde{X}^{(j)}) - T^{(\cdot)}(\tilde{X}^{(\cdot)})]^2$,
 where $T^{(\cdot)}(\tilde{X}^{(\cdot)}) = \frac{1}{N} \sum_{j=1}^N T^{(j)}(\tilde{X}^{(j)})$.

b) Large Sample Theory--By the Classical Central Limit Theorem, under appropriate regularity conditions, the distribution of the quantity

$$\sqrt{n} [T(X_1, X_2, \dots, X_n) - E_F T(X_1, X_2, \dots, X_n)]$$

tends to $N(0, \Sigma(F, T))$ in law, where $\Sigma(F, T)$ is the asymptotic variance of the statistic $T(X_1, X_2, \dots, X_n)$ under the distribution F . So, in this situation, if we can compute $\Sigma(F, T)$, then $\sigma^2[T(X_1, X_2, \dots, X_n)] \approx \frac{1}{n} \Sigma(F, T)$.

B. When the distribution F is unknown

We can estimate the unknown distribution F in two ways:

a) Parametric Method--Suppose the distribution of the X_i 's is assumed to be one member of a parametric family indexed by θ , say $\{F_\theta; \theta \in H\}$. We first use the data to obtain a suitable estimate of θ , say $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$. Then we estimate the unknown F by $F_{\hat{\theta}}$.

b) Nonparametric Method--The nonparametric estimate of F is $\hat{F} = \hat{F}_n$, the sample distribution function.

Now let us combine the two methods of standard error assessment with the above two methods of estimating the unknown distribution F .

Estimation of F

		Parametric	Nonparametric
Assessment of Standard Error	Simulation	Parametric Bootstrap	Nonparametric Boot- strap (Efron 1979)
	Large Sample Approximation	Parametric Large Sample Theory	Jackknife (Quenouille 1949, Tukey 1958) Infinitesimal Jack- knife (Jackel 1972)

Some advantages of nonparametric approaches are: i) The methods require little in the way of modeling and related assumptions; ii) the nonparametric methods can be applied in both simple and complicated settings.

Fisher (1915) employed a bell-shaped surface to estimate how much the correlation coefficient varies from sample to sample using only the information in a single sample. This familiar theory of Fisher is itself a "bootstrap theory", carried out in a parametric framework. The non-parametric bootstrap was first introduced by Efron in 1979.

The balance of the next six chapters is organized as follows: Chapter II gives a review of literature concerning the bootstrap and estimation of standard errors in a time series setting. In Chapter III, the bootstrap idea is outlined and bootstrap confidence intervals and bias estimation procedures are discussed. Chapter IV applies the bootstrap to a time series model, and also gives a brief review of conventional least squares and maximum likelihood estimation procedures. The bootstrap is used to attach standard errors to multi-period forecasts in Chapter V. Chapter VI presents a simulation experiment to assess the quality of the bootstrap. Chapter VII applies the bootstrap to two econometric models which are in current use for developing forecasts for the state of Oklahoma. In the final chapter, we summarize the results of this research and make recommendations for further study.

CHAPTER II

REVIEW OF LITERATURE

The bootstrap is a relatively newly introduced statistical methodology. The "bootstrap" is the name given by Efron (1979a) to the method for estimation of the variability in an estimate by reusing the sample at hand. Efron's bootstrap idea is based on the jackknife technique for reduction of bias of parametric estimators as suggested by Quenouille (1949) and Tukey (1958). In his original paper, Efron (1979a) showed that the bootstrap methods work satisfactorily on a variety of estimation problems. He also showed that bootstrap methods are more widely applicable than the jackknife, and also more dependable. The jackknife is shown to be a linear expansion method for approximating the bootstrap.

Since the work of Efron, there have been several articles in the literature on the theory (Bickel and Freedman, 1981, 1983; Freedman, 1981; Singh, 1981; Efron, 1981b; Beran, 1982, 1984) and specific applications (Efron, 1979b, 1981a; Gong, 1982; Efron and Gong, 1983; Diaconis and Efron, 1983; Freedman and Peters, 1983a, 1983b, 1984; Theil, Rosalsky, and Finke, 1983) of the bootstrap. A much longer review and book-length treatment of nonparametric estimation, including related topics such as the jackknife estimate of bias, the bootstrap estimate of bias, cross-validation, resampling, random

subsampling, half-sampling, nonparametric confidence intervals, and influence functions, appears in Efron (1982).

Bickel and Freedman (1981) showed that Efron's bootstrap method of distribution approximation is asymptotically valid in a large number of situations, including t-statistics, von Mises functionals, and the empirical process. Singh (1981) showed that the bootstrap approximation of the distribution of the standardized sample mean is asymptotically more accurate than approximation by the limiting normal distribution in the nonlattice case. Freedman (1981) discussed the regression and correlation models. He showed that the bootstrap approximation to the distribution of the least squares estimates is valid. Bickel and Freedman (1983) showed that the bootstrap approximation to the distribution of contrasts is valid, provided p/n is small, where n and p are the numbers of data points and parameters, respectively, in the regression model. Efron (1981b) investigated several nonparametric methods; the bootstrap, the jackknife, and the delta method. He considered nonparametric methods for estimating standard errors and confidence intervals, including the case of confidence intervals for the median.

In 1981, Efron (1981a) considered an example, including setting standard errors and confidence intervals for the parameters of an unknown distribution, when the data is subject to right censoring. Gong (1982) studied the estimates of excess error, or the difference between the true and apparent errors. She dealt with the probability of incorrectly predicting the outcome of a new patient when a prediction rule based on a set of patients is given. Freedman and Peters (1983a) presented the bootstrap procedure for determining the

variability of parameter estimates in a system of econometric equations designed to forecast demand for energy by industry. They found that the conventional asymptotic formulae for estimating standard errors are too optimistic by factors of nearly three, when applied to a finite-sample problem. The bootstrap can also be used to judge standard errors for multi-period forecasts, and to choose among competing equations. These methods were discussed in Freedman and Peters (1983b). Also in 1984, they applied the bootstrap method to the Berndt-Wood econometric model which describes the demand for capital, labor, energy, and materials. In sharp contrast with previous results, they found that the traditional coefficient estimates and the estimated standard errors perform very well.

The discrete bootstrap is by no means critical, and the normal bootstrap could be used. Theil, Rosalsky, and Finke (1983) compared the bootstrap with asymptotic standard errors of maximum likelihood coefficient estimates, using the continuous and the discrete bootstrap. The results confirmed those of Freedman and Peters. They found that the continuous bootstrap does not yield any gain over the discrete bootstrap.

Next, let us review the literature on methods of parameter and standard error estimation, and on some papers related to the bootstrap for the time series model.

Hurwicz (1950) demonstrated analytically that the use of least squares and maximum likelihood estimates tends to biased estimators for the autoregressive model parameters. Alternative estimators for parameters in the second order autoregressive models were presented by Salem and Kline (1978), using the Monte Carlo technique. Mikhail

(1975) reported bias in standard errors for econometric estimators. Marquardt (1963) developed a maximum neighborhood method which is used for the nonlinear least squares iterations. Ansley (1979) developed a method for calculating the exact likelihood for an autoregressive-moving average process.

The work of Copas (1966) is closer in spirit to the bootstrap, but quite different in detail. His paper compared, by direct Monte Carlo methods, the performance of a simple likelihood estimator with conventional least squares, and with maximum likelihood estimates. Dent and Min (1978) and Ansley and Newbold (1980) analyzed the properties of a variety of proposed estimators of unknown parameters in the autoregressive moving average models by simulation. They judged that maximum likelihood estimation is the preferred method for autoregressive moving average models.

The new computer-based bootstrap methods make it possible to explore statistical properties numerically, even though these procedures have undergone very little theoretical development since they have been computationally practical for a comparatively short time. What is needed is independent theoretical justification by theoreticians that the bootstrap estimate of precision remains as valid in complex settings as it is for simpler problems.

CHAPTER III

OUTLINE OF THE BOOTSTRAP PRINCIPLE IN GENERAL

The Bootstrap

As mentioned earlier, the bootstrap methodology was invented by Efron (1979). This method estimates standard errors by resampling the data in a suitable way. The unknown underlying distribution, F , is replaced by the empirical distribution, \hat{F}_n , of the data. Pseudo-random samples are then constructed to obtain a Monte Carlo distribution for the estimator of interest. The usual empirical c.d.f. $\hat{F}_n(x)$ is the following:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

where $I(A)$ is the indicator function. The bootstrap technique is distribution-free, and investigates the appropriate finite-sample behavior for the estimates. An important practical advantage of the bootstrap method in statistics is its direct and nonasymptotic character. $\hat{F}_n(x)$ is used to approximate $F(x)$ and it is well-known that $\sup |\hat{F}_n - F|$ approaches zero with probability one as $n \rightarrow \infty$ (see section 6f in Rao (1973)).

Suppose we have a real-valued statistic $\hat{\theta}(X_1, X_2, \dots, X_n)$, where X_i are independent and identically distributed according to some unknown probability distribution, i.e.,

$$X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} F.$$

The X_i may be real-valued, two-dimensional, or take values in a more complicated space. Having observed

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n,$$

we wish to assign some measure of precision to $\hat{\theta}(X_1, X_2, \dots, X_n)$. The true standard error is a function of the unknown F , say

$$\sigma(F) = \sigma_F(\hat{\theta}),$$

which denotes the standard error of the statistic in question, under sampling from F .

The bootstrap estimate of the standard error, $\hat{\sigma}_{\text{BOOT}}$, is simply

$$\hat{\sigma}_{\text{BOOT}} = \sigma(\hat{F}_n) = \sigma_{\hat{F}_n}(\hat{\theta}),$$

where \hat{F}_n is the empirical probability distribution

$$\hat{F}_n: \text{mass } \frac{1}{n} \text{ on } x_i, i = 1, 2, \dots, n.$$

Example 1: Let $\Omega = \mathbb{R}^1$ and $\hat{\theta} = \bar{X} = \sum_{i=1}^n X_i/n$, in which case

$$\sigma(F) = \left[\int_{-\infty}^{\infty} (x - E_F X)^2 dF(x) \right]^{1/2}.$$

Then,

$$\hat{\sigma}_{\text{BOOT}} = \left[\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 / n \right\} / n \right]^{1/2} = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}.$$

In fact, it is usually difficult to express the function $\sigma(\hat{F}_n)$ in simple form, and in order to calculate $\hat{\sigma}_{\text{BOOT}}$, it is usually necessary to employ a Monte Carlo simulation algorithm in the following way:

Step 1. Draw a "bootstrap pseudo-random sample" from \hat{F}_n ,

$$X_1^*, X_2^*, \dots, X_n^* \stackrel{\text{iid}}{\sim} \hat{F}_n,$$

in which each X_i^* independently takes value x_i with probability $\frac{1}{n}$, $i = 1, 2, \dots, n$. In other words, $X_1^*, X_2^*, \dots, X_n^*$ is a random sample of size n drawn with replacement from the set of observations $\{x_1, x_2, \dots, x_n\}$.

Step 2. Calculate a corresponding bootstrap value

$$\hat{\theta}^* = \theta(X_1^*, X_2^*, \dots, X_n^*) \text{ from the sample obtained in Step 1.}$$

Step 3. Independently repeat Steps 1 and 2, a large number of times, say B , obtaining bootstrap values

$$\hat{\theta}^{*(1)}, \hat{\theta}^{*(2)}, \dots, \hat{\theta}^{*(B)}.$$

Step 4. The bootstrap estimate of standard error for the estimator $\hat{\theta}(X_1, X_2, \dots, X_n)$ is

$$\begin{aligned} \hat{\sigma}_{\text{BOOT}} &= [\text{Var } \hat{\theta}(X_1^*, X_2^*, \dots, X_n^*)]^{1/2} \\ &= \left[\frac{\sum_{b=1}^B \{\hat{\theta}^{*(b)} - \hat{\theta}^{*(\cdot)}\}^2}{B-1} \right]^{1/2}, \end{aligned}$$

$$\text{where } \hat{\theta}^{*(\cdot)} = \frac{\sum_{b=1}^B \hat{\theta}^{*(b)}}{B}.$$

Example 2. Let $\Omega = \mathbb{R}^2$, the plane, and $\hat{\theta} = \hat{\rho}(X_1, X_2, \dots, X_n)$, the Pearson correlation coefficient for the observed sample. The bootstrap estimate of the standard error of the parameter ρ , $\hat{\sigma}_{\text{BOOT}}$, is simply

$$\hat{\sigma}_{\text{BOOT}} = \sigma(\hat{F}_n) = \left[\frac{\sum_{b=1}^B (\hat{\rho}^*(b) - \hat{\rho}^*(\cdot))^2}{B-1} \right]^{1/2}, \quad \hat{\rho}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\rho}^*(b)}{B}.$$

The $\hat{\rho}^*(b)$, $b = 1, \dots, B$, are generated according to steps 1, 2, and 3 of the algorithm just described.

Example 3. Suppose X_1, X_2, \dots, X_n is a random sample with unknown cumulative distribution function F . Let

$$S_n(x, F) = P_F \{ \sqrt{n} \sup |\hat{F}_n - F| \leq x \}$$

be the c.d.f of the Kolmogorov-Smirnov type statistic. Notice that $S_n(x, \hat{F}_n)$ is the c.d.f. of the Kolmogorov-Smirnov type statistic if we put the actual c.d.f. of each observation at \hat{F}_n rather than F . The bootstrap estimate of $S_n(x, F)$ is $S_n(x, \hat{F}_n) = P_{\hat{F}_n} \{ \sqrt{n} \sup |\hat{F}_n^* - \hat{F}_n| \leq x \}$.

Beran (1984) makes the following points about the bootstrap procedure:

1. The bootstrap methodology provides a way of estimating nuisance parameters in many statistical situations.
2. We don't need to be mathematically sophisticated to use the bootstrap methods, we only need to know that the method is valid and how to perform the resampling procedure.
3. We can use the bootstrap technique to solve some statistical problems for which the conventional approach fails to work satisfactorily.

Nonparametric Confidence Intervals via Bootstrap

So far we have dealt with estimating the standard deviation of a point estimator $\hat{\theta}(X_1, X_2, \dots, X_n)$. In applied work, estimates are

often given in the form of the estimate plus or minus a certain amount. One of our purposes of estimating a standard error is to find confidence intervals for the unknown real parameter θ . With the assumption of normality, the typical confidence intervals usually used are: $\hat{\theta} \pm Z_{\alpha} \hat{\sigma}$ or $\hat{\theta} \pm t_{\alpha} \hat{\sigma}$, where Z_{α} is the $100(1-\alpha)$ percentile point of a standard normal distribution and t_{α} is the $100(1-\alpha)$ percentile point of the student t distribution. In small-sample parametric situations, confidence intervals are quite often asymmetric about the point estimate $\hat{\theta}(X_1, X_2, \dots, X_n)$. We shall outline some nonparametric methods of setting confidence intervals, that try to use the correct degree of asymmetry, from a discussion in Efron (1981b), and also Chapter 10 of Efron (1982).

a. The Percentile Method--This method assigns approximate confidence intervals to our parameter θ , based on the empirical bootstrap distribution of $\hat{\theta}$. Let

$$\hat{c}(x) = \text{Prob}_{*}\{\hat{\theta}^{*} \leq x\}$$

be the bootstrap cumulative distribution function of $\hat{\theta}^{*}$. The "Prob_{*}" indicates the bootstrap probability. In the Monte Carlo simulation method, this "Prob_{*}" will be approximated by

$$\frac{\#\{\hat{\theta}^{*}(b) \leq x\}}{B}.$$

For a given α , define $\hat{\theta}_L(\alpha)$ and $\hat{\theta}_U(\alpha)$ as follows:

$$\hat{\theta}_L(\alpha) \equiv \hat{c}^{-1}(\alpha), \quad \hat{\theta}_U(\alpha) \equiv \hat{c}^{-1}(1-\alpha),$$

then the percentile method takes

$$\theta \in [\hat{\theta}_L(\alpha), \hat{\theta}_U(\alpha)]$$

as an approximate $1-2\alpha$ central confidence interval for the unknown parameter θ . As we see $\alpha = \hat{c}(\hat{\theta}_L)$ and $1-\alpha = \hat{c}(\hat{\theta}_U)$, therefore the percentile method confidence interval approximates the central $1-2\alpha$ percentile of the bootstrap cumulative distribution function. Thus the percentile method results in a statement

$$P_*[\hat{\theta}^* \in \{\hat{\theta}_L(\alpha), \hat{\theta}_U(\alpha)\}] = 1-2\alpha,$$

which is a substitute for the statement

$$P_F[\theta \in \{\hat{\theta}_L(\alpha), \hat{\theta}_U(\alpha)\}] = 1-2\alpha.$$

b. The Bias-Corrected Percentile Method--If the bootstrap distribution is median biased, in other words, if $\text{Prob}_*\{\hat{\theta}^* \leq \hat{\theta}\} \neq .50$, then we turn to the bias-corrected percentile method. Let $\Phi(z)$ be the cumulative distribution function for a standard normal variate and define

$$z_0 \equiv \Phi^{-1}(\hat{c}(\hat{\theta})),$$

where $\hat{c}(\hat{\theta}) = \text{Prob}_*\{\hat{\theta}^* \leq \hat{\theta}\}$ as in the previous definition. Then the bias-corrected percentile method takes

$$\theta \in [\hat{c}^{-1}\{\Phi(2z_0 - z_\alpha)\}, \hat{c}^{-1}\{\Phi(2z_0 + z_\alpha)\}]$$

as an approximate $1-2\alpha$ central confidence interval for θ . Here z_α is the upper α point for a standard normal $\Phi(z_\alpha) = 1-\alpha$. Note that if $\hat{\theta}$ lies at the median of the bootstrap distribution (i.e., $\hat{c}(\hat{\theta}) = 0.5$), then z_0 is zero and there is no correction. In this case, the bias-corrected percentile method reduces to the percentile method.

The Bootstrap Estimate of Bias

The jackknife originally was introduced by Quenouille (1949) as a means of reducing the bias in an estimator (see Miller (1974)). We wish to estimate the bias of a statistic $\hat{\theta} = \theta(\hat{F}_n)$, then the bias is well defined,

$$\beta = E\{\theta(\hat{F}_n) - \theta(F)\}.$$

How would we estimate bias in the context of bootstrapping? The bootstrap estimate of bias is

$$\hat{\beta}_{\text{BOOT}} = E_{*}\{\theta(\hat{F}^{*}) - \theta(\hat{F}_n)\},$$

where E_{*} and \hat{F}^{*} denote expectation in terms of bootstrap sampling and the bootstrap empirical probability distribution, respectively. In practice, the bootstrap estimate of bias is approximated by Monte Carlo methods. The steps 1, 2, and 3 are the same as in Section 1. At step 4, we calculate

$$\hat{\beta}_{\text{BOOT}} = \hat{\theta}^{*}(\cdot) - \hat{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*}(b) - \hat{\theta}.$$

We would use this to correct the estimator for bias in the following way:

$$\hat{\theta}_{\text{CORRECTED}} = \hat{\theta} - \hat{\beta}_{\text{BOOT}}.$$

CHAPTER IV

BOOTSTRAPPING A TIME SERIES MODEL

Chapter Introduction

This chapter is mainly concerned with presenting the bootstrap in the context of the second-order autoregressive model

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

where ε_t are errors or white noise series, and the model is stationary. We use the bootstrap technique to make alternative estimates of the standard errors of fitted coefficients. We compare the performance of conventional least squares and maximum likelihood estimates to the performance of a bootstrap methodology in the setting of a time series model. As it turns out, the conventional estimates appear to over-estimate the true standard errors when applied to a particular finite sample. We can also attach standard errors to forecasts, and to the variability indicated by the bootstrap, using the bootstrap technique.

The bootstrap is a methodology for estimating standard errors by resampling the data to obtain a Monte Carlo distribution for the resulting random variable. The idea is that we resample the residuals, keeping the stochastic structure, so that standard errors are generated preserving the model's own assumptions.

Now then, consider the well-known autoregressive moving average model of order (p,q),

$$y_t = y_t(\delta, \underline{\phi}, \underline{\theta}) = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} . \quad (\text{IV-1})$$

In this ARMA(p,q) model, δ , $\underline{\phi}' = (\phi_1, \phi_2, \dots, \phi_p)$ and $\underline{\theta}' = (\theta_1, \theta_2, \dots, \theta_q)$ are coefficient vectors, to be estimated from the data, subject to some given conditions. The coefficients are estimated as $\hat{\delta}$, $\hat{\underline{\phi}}$, and $\hat{\underline{\theta}}$, usually by the least squares or maximum likelihood estimation procedure. After δ , $\underline{\phi}$, and $\underline{\theta}$ are estimated, residuals are

$$\hat{\varepsilon}_t = y_t(\delta, \underline{\phi}, \underline{\theta}) - \hat{y}_t(\hat{\delta}, \hat{\underline{\phi}}, \hat{\underline{\theta}}) , \quad t = p+1, p+2, \dots, n.$$

i.e., the difference between the actual observations and the predicted observations. Let \hat{F}_n be the empirical distribution function of the residuals, putting mass $\frac{1}{n-p}$ on each of ε_t , $t = p+1, p+2, \dots, n$.

Next let us set the coefficients of model (IV-1) at $\hat{\delta}$, $\hat{\underline{\phi}}$, and $\hat{\underline{\theta}}$ with common empirical error distribution \hat{F}_n . We can then generate "the bootstrap pseudo-data" using this model. We can draw a bootstrap sample $y_1^*, y_2^*, \dots, y_n^*$ by independent random sampling from \hat{F}_n . The construction is recursive: $y_1^* = y_1$, $y_2^* = y_2$, ..., $y_p^* = y_p$, and

$$y_t^* = \hat{\delta} + \hat{\phi}_1 y_{t-1}^* + \hat{\phi}_2 y_{t-2}^* + \dots + \hat{\phi}_p y_{t-p}^* + \varepsilon_t^* + \hat{\theta}_1 \varepsilon_{t-1}^* + \hat{\theta}_2 \varepsilon_{t-2}^* + \dots + \hat{\theta}_q \varepsilon_{t-q}^* , \quad \forall t = p+1, p+2, \dots, n,$$

where the ε_t^* are independently distributed according to \hat{F}_n .

Now using the previous least squares or maximum likelihood estimation procedure, compute the estimates $\hat{\underline{\phi}}^*$ and $\hat{\underline{\theta}}^*$ from the pseudo-data. We can compute the distribution of the pseudo-errors $\hat{\underline{\phi}}^* - \hat{\underline{\phi}}$, $\hat{\underline{\theta}}^* - \hat{\underline{\theta}}$ to approximate the distribution of the real errors $\underline{\phi} - \hat{\underline{\phi}}$, $\underline{\theta} - \hat{\underline{\theta}}$ by the Monte Carlo simulation method.

Numerical Example--Daily Readings
of the Viscosity

The 95 daily readings of the viscosity of a chemical product XB-75-5 (see Bowerman and O'Connell (1979), example 10.2) were selected as an example to illustrate the use of bootstrap methodology. The 95 daily readings, y_1, y_2, \dots, y_{95} are given in Table I. Since the daily readings of the time series seem to fluctuate around a constant mean, they seem to be stationary. The analysis in Bowerman and O'Connell shows that, for the original time series, the sample autocorrelation function dies down in a damped sine-wave fashion, and the sample partial autocorrelation function cuts off after lag 2, identifying the appropriate model as autoregressive of order 2, AR(2).

After we identify an appropriate time series model, we need to estimate parameters of our model. Two basic methods are available for estimating these parameters in the Box-Jenkins method. One such method is the least squares method; the other is the maximum likelihood method.

Conventional Least Squares Estimates

The least squares estimates can be obtained through standard computer programs. It is required that we specify preliminary estimates of the unknown parameters as input values. These preliminary estimates are the starting values in an iterative search procedure to compute least squares estimates of the parameters. These preliminary estimates are obtained by equating the sample autocorrelations to the parameter representation of the population autocorrelations

TABLE I
DAILY READINGS OF THE VISCOSITY OF CHEMICAL
PRODUCT XB-77-5

t	y _t	t	y _t	t	y _t
1	25.000	33	34.4337	65	32.2754
2	27.000	34	35.4844	66	33.2214
3	33.5142	35	33.2381	67	34.5786
4	35.4962	36	36.1684	68	32.3448
5	36.9029	37	34.4116	69	31.5316
6	37.8359	38	33.7668	70	37.8044
7	34.2654	39	33.4246	71	36.0536
8	31.8978	40	33.5719	72	35.7297
9	33.7567	41	35.9222	73	36.7991
10	36.6298	42	33.2125	74	34.9502
11	36.3518	43	37.1668	75	33.5246
12	40.0762	44	35.8138	76	35.1012
13	38.0928	45	33.6847	77	35.9774
14	34.5412	46	33.2761	78	38.0977
15	34.8567	47	38.8163	79	33.4598
16	34.5316	48	42.0838	80	32.9278
17	32.3851	49	40.0069	81	36.5121
18	32.6058	50	33.4514	82	37.4243
19	34.8913	51	30.8413	83	35.1550
20	38.2418	52	30.0655	84	34.4797
21	36.8926	53	37.0544	85	33.2898
22	33.8942	54	39.0982	86	33.9252
23	34.1710	55	37.9075	87	36.1036
24	35.4268	56	36.2393	88	36.7351
25	38.5831	57	34.9535	89	35.4576
26	34.6184	58	33.2061	90	37.5924
27	33.9741	59	34.4261	91	34.4895
28	30.2072	60	37.4511	92	39.1692
29	30.5429	61	37.3335	93	35.8242
30	34.8686	62	38.4679	94	32.3875
31	35.8892	63	33.0976	95	31.2846
32	35.2035	64	32.9285		
$\bar{y} = 34.93$					

and then solving for each individual parameter. Also the AR(2) model requires starting values for the data points y_0 and y_{-1} . These are obtained using a technique called "backcasting", which is fully explained in Box and Jenkins (1976).

Most non-linear least squares routines, such as the one used by SAS, employ a Taylor Series expansion of ε_t about the initial parameter value estimates, say δ_0 , $\phi_{1,0}$, and $\phi_{2,0}$. Terminating the expansion after the first term, we have

$$\begin{aligned}\varepsilon_t &= \varepsilon_t(\delta, \phi_1, \phi_2) = y_t - \delta - \phi_1 y_{t-1} - \phi_2 y_{t-2} \\ &\approx \varepsilon_t(\delta_0, \phi_{1,0}, \phi_{2,0}) - [(\delta - \delta_0) \left(-\frac{\partial \varepsilon_t}{\partial \delta}\right) + (\phi_1 - \phi_{1,0}) \left(-\frac{\partial \varepsilon_t}{\partial \phi_1}\right) \\ &\quad + (\phi_2 - \phi_{2,0}) \left(-\frac{\partial \varepsilon_t}{\partial \phi_2}\right)].\end{aligned}$$

This equation can be rewritten in the familiar form of linear least squares:

$$\varepsilon_t(\delta_0, \phi_{1,0}, \phi_{2,0}) = (\delta - \delta_0)x_{1,t} + (\phi_1 - \phi_{1,0})x_{2,t} + (\phi_2 - \phi_{2,0})x_{3,t} + \varepsilon_t,$$

where $x_{1,t} = -\frac{\partial \varepsilon_t}{\partial \delta}$, $x_{2,t} = -\frac{\partial \varepsilon_t}{\partial \phi_1}$, and $x_{3,t} = -\frac{\partial \varepsilon_t}{\partial \phi_2}$. Now combine these equations for $t = 1, \dots, n$ to obtain

$$\underline{\varepsilon}_0 = X \cdot \underline{\beta} + \underline{\varepsilon},$$

where $\underline{\varepsilon}_0$ is the $n \times 1$ vector of values of $\varepsilon_t(\delta_0, \phi_{1,0}, \phi_{2,0})$, X is the $n \times 3$ matrix whose t^{th} row is $(x_{1,t}, x_{2,t}, x_{3,t})$, $\underline{\beta}' = [(\delta - \delta_0), (\phi_1 - \phi_{1,0}), (\phi_2 - \phi_{2,0})]$, and $\underline{\varepsilon}' = (\varepsilon_1, \dots, \varepsilon_n)$. The coefficients can then be estimated by

$$\begin{pmatrix} \widehat{\delta - \delta}_0 \\ \widehat{\phi_1 - \phi}_{1,0} \\ \widehat{\phi_2 - \phi}_{2,0} \end{pmatrix} = (X'X)^{-1}X'\varepsilon_0.$$

The initial estimates can now be updated:

$$\begin{pmatrix} \delta_1 \\ \phi_{1,1} \\ \phi_{2,1} \end{pmatrix} = \begin{pmatrix} \delta_0 \\ \phi_{1,0} \\ \phi_{2,0} \end{pmatrix} + \begin{pmatrix} \widehat{\delta - \delta}_0 \\ \widehat{\phi_1 - \phi}_{1,0} \\ \widehat{\phi_2 - \phi}_{2,0} \end{pmatrix},$$

and the process repeated until convergence is obtained. The partial derivatives needed for the entries in the X matrix can be estimated numerically at each stage. For example,

$$\frac{\partial \varepsilon_t}{\partial \delta} \hat{=} \frac{\varepsilon_t(\delta_0 + h, \phi_{1,0}, \phi_{2,0}) - \varepsilon_t(\delta_0, \phi_{1,0}, \phi_{2,0})}{h},$$

where h is some sufficiently small positive number.

At the last stage of the iteration, when the convergence criterion is attained, the conventional standard errors are the square roots of the diagonal entries of the matrix

$$V = s^2(X'X)^{-1}, \quad \text{where } s^2 = \frac{1}{n-3} \sum_{t=1}^n \hat{\varepsilon}_t^2.$$

More details can be found in Box and Jenkins (1976).

Conventional Maximum Likelihood Estimates

An alternative approach is to calculate maximum likelihood estimates. The maximum likelihood estimates presuppose distributional properties of the disturbances ε_t . The assumption is that the disturbances ε_t are nor-

mally distributed with mean zero and variance σ_ε^2 .

It is shown in Box and Jenkins (1976) that the exact likelihood for the AR(2) model is

$$L(\delta, \phi, \sigma_\varepsilon^2 | \text{data}) = (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} |V|^{\frac{1}{2}} \exp \left\{ -\frac{S(\delta, \phi)}{2\sigma_\varepsilon^2} \right\}$$

$$\text{where } V = v_{ij} = \sigma_\varepsilon^2 \begin{bmatrix} \text{Var}(Y_t) & \text{Cov}(Y_t, Y_{t-1}) \\ \text{Cov}(Y_t, Y_{t-1}) & \text{Var}(Y_{t-1}) \end{bmatrix}^{-1}$$

$$\text{and } S(\delta, \phi) = \sum_{i=1}^2 \sum_{j=1}^2 v_{ij} y_i y_j + \sum_{t=3}^n (y_t - \delta - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2.$$

Ansley and Newbold (1980) describe an efficient algorithm for maximizing the likelihood numerically to produce the maximum likelihood estimates $\hat{\delta}$, $\hat{\phi}_1$, $\hat{\phi}_2$, and $\sigma_\varepsilon^2 = S(\hat{\delta}, \hat{\phi})/n$.

Under appropriate regularity conditions, discussed for example in Rao (1973), the inverse of the information matrix supplies the asymptotic variance/covariance matrix of the maximum likelihood estimates. The information matrix in the present case is

$$I(\delta, \phi) = -E \begin{bmatrix} \frac{\partial^2 \ell}{\partial \delta^2} & \frac{\partial^2 \ell}{\partial \delta \partial \phi_1} & \frac{\partial^2 \ell}{\partial \delta \partial \phi_2} \\ \frac{\partial^2 \ell}{\partial \phi_1 \partial \delta} & \frac{\partial^2 \ell}{\partial \phi_1^2} & \frac{\partial^2 \ell}{\partial \phi_1 \partial \phi_2} \\ \frac{\partial^2 \ell}{\partial \phi_2 \partial \delta} & \frac{\partial^2 \ell}{\partial \phi_2 \partial \phi_1} & \frac{\partial^2 \ell}{\partial \phi_2^2} \end{bmatrix},$$

where $\ell(\delta, \phi, \sigma_\varepsilon^2)$ is the log-likelihood function.

Estimates of the variances and covariances of the parameters may be obtained by evaluating these derivatives numerically and replacing δ , ϕ_1 , and ϕ_2 by their maximum likelihood estimates. This evaluation is a part of the Ansley and Newbold algorithm referred to earlier.

Bootstrapping AR(2) Model

Comparison with the Least

Squares Estimates

The main finding is that for the least squares estimates in the Box-Jenkins methodology, the true standard errors appear to be overestimated when the empirical error distribution is the true error generating process. The appropriately identified model was

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad t = 1, 2, \dots, 95$$

where ε_t is a white noise process, and δ , ϕ_1 , and ϕ_2 are parameters to be estimated.

The stochastic assumptions in the above model concern the stochastic white noise process ε_t , which are unobservable random variables. They state that each of the white noise terms has a zero mean, i.e., $E(\varepsilon_t) = 0$, $\forall t$, and that ε_t , $t = 1, 2, \dots, 95$, are independently and identically distributed in time.

To get started on the bootstrap method, let $\hat{\mu}$, $\hat{\phi}_1$, and $\hat{\phi}_2$ be the least squares estimates reported in Table II. Note that we reported the estimated mean of the AR(2) process $\hat{\mu}$ instead of the estimated constant coefficient $\hat{\delta}$. One of the conditions for stationarity in a time series process is to attain a constant mean over time. The mean of the second-order autoregressive model is a function of the parameters δ , ϕ_1 , and ϕ_2

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2}.$$

From this point, we will report $\hat{\mu}$ and its standard error rather than $\hat{\delta}$ and its standard error.

TABLE II

BOOTSTRAP RESULT FOR THE AR(2) MODEL (THERE ARE 100 BOOTSTRAP PSEUDO-RANDOM SAMPLES. ESTIMATION IS BY LEAST SQUARES, USING THE FIRST SEED.)

Col. Parm.	Least Squares Estimates		Bootstrap		
	(1) Estimate	(2) Conv. SE	(3) B/S Mean	(4) B/S SD	(5) RMS Conv. SE
μ	34.9039	0.2978	34.9116	0.2564	0.3009
ϕ_1	0.613551	0.0971	0.645318	0.0763	0.0956
ϕ_2	-0.383048	0.0975	-0.404626	0.0759	0.0961

Consider the residual errors,

$$\hat{\varepsilon}_t = y_t - \hat{\delta} - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2}.$$

Let $\hat{\varepsilon}$ be the 93-vector $(\hat{\varepsilon}_3, \hat{\varepsilon}_4, \dots, \hat{\varepsilon}_{95})$ of residual errors. Now simulate the model AR(2) putting mass $\frac{1}{93}$ on each of the $\{\hat{\varepsilon}_t: t=3, 4, \dots, 95\}$.

- i. y_1 and y_2 are held fixed.
- ii. The parameters are set at their least squares estimates, $\hat{\delta}$, $\hat{\phi}_1$, and $\hat{\phi}_2$.
- iii. The errors are independent with common empirical error distribution \hat{F}_n .

In other words, $\{\varepsilon_t^*: t=3,4,\dots,95\}$ is a random sample of size 93 drawn with replacement from \hat{F}_n . The bootstrap pseudo-data can be collected iteratively: $y_1^* = y_1$, $y_2^* = y_2$, and for $t=3,4,\dots,95$,

$$y_t^* = \hat{\delta} + \hat{\phi}_1 y_{t-1}^* + \hat{\phi}_2 y_{t-2}^* + \varepsilon_t^*$$

These are artificial data for the model AR(2). Now then we can obtain the bootstrap parameter estimates $\hat{\delta}^*$, $\hat{\mu}^*$, $\hat{\phi}_1^*$, and $\hat{\phi}_2^*$ from the above artificial data using the previous least squares estimation procedure. These steps are repeated, independently, 100 times (see Figure 1). On each repetition, a new set of starred residuals was generated, hence a new set of bootstrap pseudo-data, and then a new set of starred parameter estimates. Thus we obtain $\hat{\delta}^{*(b)}$, $\hat{\mu}^{*(b)}$, $\hat{\phi}_1^{*(b)}$, and $\hat{\phi}_2^{*(b)}$ for $b=1,2,\dots,100$.

Now go back to Table II. The first column in this table shows the conventional least squares estimates, fit to the second-order autoregressive model. In the second column, the standard errors are displayed from the least squares estimation method, and will be called the "conventional standard errors." The bootstrap methodology gives an alternative method for approximating the standard errors, and a means for estimating the statistical accuracy of the conventional standard errors.

We already mentioned that the conventional standard error estimates for the Box-Jenkins method appear to overestimate the true standard errors. To compare this conventional method with the bootstrap technique, we built up a bootstrap simulation model (see Figure 1), where the parameters and the distribution of the residual errors are all known. Column (4) shows the bootstrap estimates of variability in

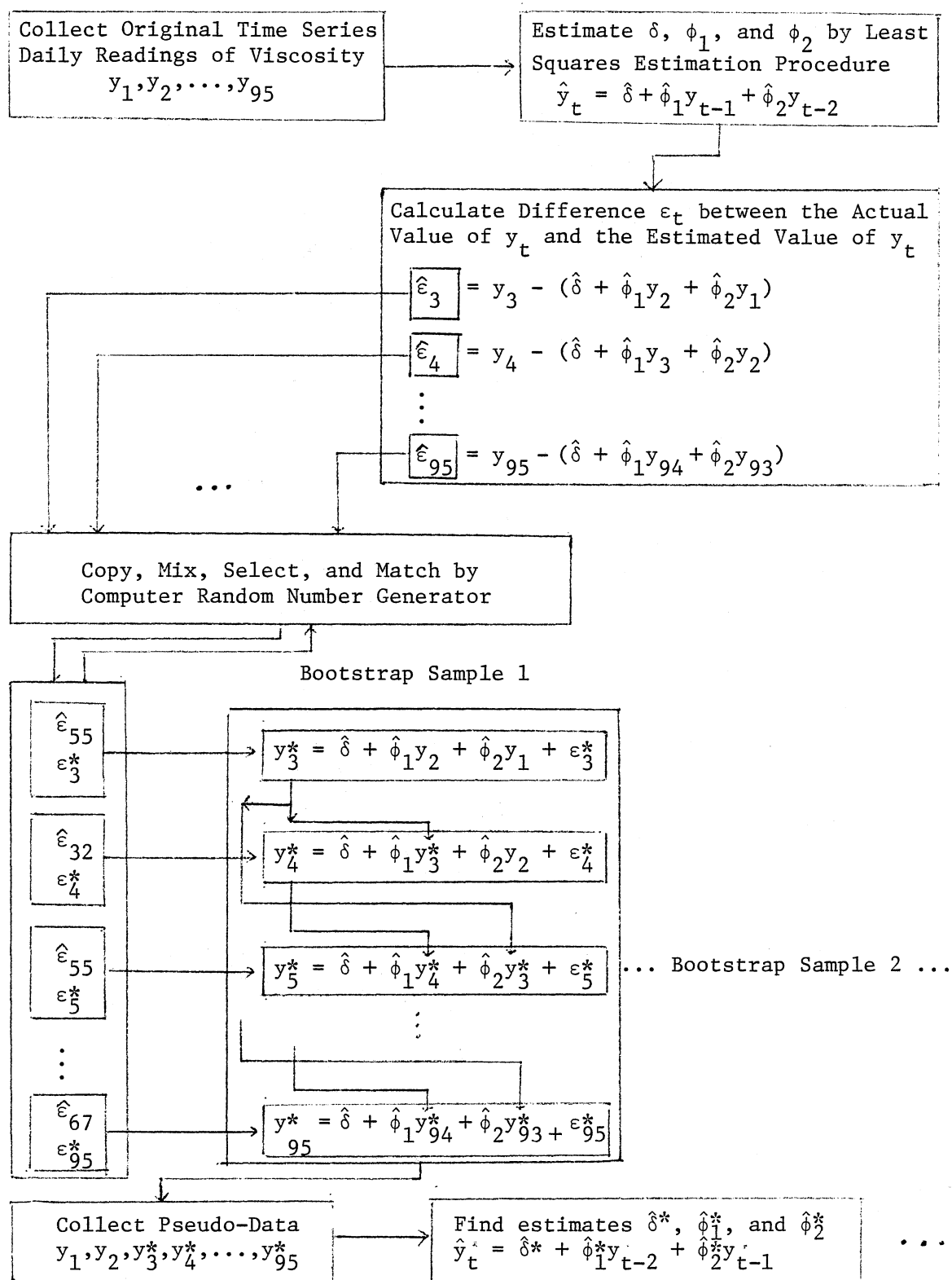


Figure 1. Bootstrap Model of the Daily Readings of Viscosity

the parameter estimates. This "bootstrap SD" shows the "real" variability in the simulation world of the bootstrap. This variability was determined empirically. In the same bootstrap simulation world, how good are the conventional standard errors? As shown in the "RMS Conventional SE" column, the conventional standard errors are large. The root mean square of the standard errors is the square root of the average of the variances. This RMS is

$$\left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2 \right]^{\frac{1}{2}},$$

where SE_b^2 is the conventional estimate of the variance of the parameter in question, for the b^{th} set of the pseudo-data. (See Table III for the computations.)

At each of the 100 bootstrap repetitions, the conventional standard errors are calculated using the conventional least squares estimation on the starred data set. The root mean square of these standard errors is column (5). Let us look at the coefficient ϕ_1 . In the bootstrap simulation methodology, the "real" variability for this parameter is 0.0763 as shown in column (4). But the typical variability using the conventional method is 0.0956. Column (5) is uniformly larger than column (4), indicating the standard errors from the Box-Jenkins method are overestimating the real variability.

Now let us investigate the bias in the least squares coefficient estimates. Look at columns (1) and (3) from Table II. For example, the coefficient ϕ_1 was set to the estimated value 0.613551 in the building of the bootstrap pseudo-data. However, the 100 bootstrap coefficients $\hat{\phi}_1^*$ have a sample average of 0.645318, giving the difference of 0.031767. A standard error for the difference can be

TABLE III

COMPUTATIONS FOR THE ENTRIES IN TABLE II (THERE ARE 100 B/S BOOTSTRAP PSEUDO-RANDOM SAMPLES)

Col. Parm.	LSE		Bootstrap		
	(1) Estimate	(2) Conv. SE	(3) B/S Mean	(4) B/S SD	(5) RMS Conv. SE
μ	$\hat{\mu}$	$se(\hat{\mu})$	$\hat{\mu}^*(\cdot) = \frac{1}{100} \sum_{b=1}^{100} \hat{\mu}^*(b)$	$\sigma_B(\hat{\mu}) = \left[\frac{1}{100-1} \sum_{b=1}^{100} (\hat{\mu}^*(b) - \hat{\mu}^*(\cdot))^2 \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2(\hat{\mu}^*) \right]^{\frac{1}{2}}$
ϕ_1	$\hat{\phi}_1$	$se(\hat{\phi}_1)$	$\hat{\phi}_1^*(\cdot) = \frac{1}{100} \sum_{b=1}^{100} \hat{\phi}_1^*(b)$	$\sigma_B(\hat{\phi}_1) = \left[\frac{1}{100-1} \sum_{b=1}^{100} (\hat{\phi}_1^*(b) - \hat{\phi}_1^*(\cdot))^2 \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2(\hat{\phi}_1^*) \right]^{\frac{1}{2}}$
ϕ_2	$\hat{\phi}_2$	$se(\hat{\phi}_2)$	$\hat{\phi}_2^*(\cdot) = \frac{1}{100} \sum_{b=1}^{100} \hat{\phi}_2^*(b)$	$\sigma_B(\hat{\phi}_2) = \left[\frac{1}{100-1} \sum_{b=1}^{100} (\hat{\phi}_2^*(b) - \hat{\phi}_2^*(\cdot))^2 \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2(\hat{\phi}_2^*) \right]^{\frac{1}{2}}$

calculated from the standard deviation of the $\hat{\phi}_1^*$ divided by the square root of the number of the bootstrap replications, $0.0763/\sqrt{100} = 0.0076$. The calculated t-statistic is $0.031767/0.0076 = 4.16$ with 99 degrees of freedom, so the bias is significant. The coefficient ϕ_2 is also significantly biased, the term μ is not.

The statistics to test for bias in the least squares coefficient estimates are:

- i. for μ ; $t = (34.9116 - 34.9039)/(0.2564/\sqrt{100}) = 0.30$,
- ii. for ϕ_1 ; $t = (0.645318 - 0.613551)/(0.0763/\sqrt{100}) = 4.16$, and
- iii. for ϕ_2 ; $t = \{-0.404626 - (-0.383048)\}/(0.0759/\sqrt{100}) = -2.84$.

We obtained consistent bootstrap experiment results from the second seed, the results of which are reported in Table XIII (see Appendix C). The last column of this table gives the calculated t-values for checking the biases. A seed number is required for the initialization of a random number stream.

Comparison with the Maximum

Likelihood Estimates

We apply the bootstrap to the viscosity data using the alternative estimation procedure, namely maximum likelihood estimation, in the Box-Jenkins method. Our well-fitted model for the viscosity data was the AR(2) model,

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad t=1,2,\dots,95$$

To do the bootstrap, all ingredients are known:

- i. y_1 and y_2 are held fixed,
- ii. The parameters are fixed at their maximum likelihood estimated values; $\hat{\delta} = 26.237$, $\hat{\phi}_1 = 0.682098$, and $\hat{\phi}_2 = -0.432882$.

iii. The disturbance terms ε_t are independent with common distribution \hat{F}_n .

The bootstrap pseudo-random samples were generated in a similar manner, and the maximum likelihood estimation procedure was applied to each pseudo-random sample to obtain the fitted parameters. The fluctuation of these fitted parameters showed the variability of the maximum likelihood estimates as a statistical estimate for the viscosity data under the assumption that the error distribution is \hat{F}_n . The results are presented in Table IV. Each column might be interpreted as before:

Col. (1): Maximum likelihood estimates of parameters, $\hat{\mu}$, $\hat{\phi}_1$, and $\hat{\phi}_2$.

Col. (2): Variability of the estimates of parameters by MLE, $se(\hat{\mu})$, $se(\hat{\phi}_1)$, and $se(\hat{\phi}_2)$.

Col. (3): Sample mean of the estimates of parameters in the simulation world of the bootstrap, i.e., under the assumption that $\varepsilon_t \sim \hat{F}_n$.

Col. (4): The "real" variability of the maximum likelihood estimates for parameters μ , ϕ_1 , and ϕ_2 in this simulation world. This column gives the bootstrap estimates of standard error for each parameter.

Col. (5): The variability indicated by the conventional maximum likelihood estimation formulae, as applied to starred data sets.

The comparison between RMS Conventional SEs and B/S SDs reveals that conventional SEs overestimate the real variability. Let us see the t statistics to check the bias in the maximum likelihood coefficient estimates:

TABLE IV
 BOOTSTRAP RESULT FOR THE AR(2) MODEL (THERE ARE 100
 B/S PSEUDO-RANDOM SAMPLES. ESTIMATION IS BY
 MAXIMUM LIKELIHOOD, USING THE
 FIRST SEED.)

Parm.	Col. Max. Likelihood Estimates		Bootstrap		
	(1) Estimate	(2) Conv. SE	(3) B/S Mean	(4) B/S SD	(5) RMS Conv. SE
μ	34.9461	0.2962	34.9690	0.2636	0.3010
ϕ_1	0.682098	0.0983	0.776016	0.0859	0.0932
ϕ_2	-0.432882	0.0944	-0.512179	0.0865	0.0892

i. for μ ; $t = (34.9690 - 34.9461) / (0.2636 / \sqrt{100}) = 0.87$,

ii. for ϕ_1 ; $t = (0.776016 - 0.682098) / (0.0859 / \sqrt{100}) = 10.94$,

iii. for ϕ_2 ; $t = \{-0.512179 - (-0.432882)\} / (0.0865 / \sqrt{100}) = -9.17$.

The biases in coefficients ϕ_1 and ϕ_2 are significant, but the bias in μ is not. The result from the second seeds is reported in Table XIV (see Appendix C). The results differ in that the RMS conventional SE for ϕ_2 is a bit smaller than the bootstrap SD. The biases for ϕ_1 and ϕ_2 are significant, as with the first seed.

Since the conventional standard error formula appears to underestimate the true standard errors, these results cast doubt on the validity of conventional procedures when the true error distribution is \hat{F}_n . The biases in the estimates of $\hat{\phi}_1$ and $\hat{\phi}_2$ have also been shown under this error distribution.

To do the bootstrap, the first two observations y_1 and y_2 were held fixed. In an autoregressive model, this assumption is not met since lagged values of the dependent variable cannot be fixed in repeated sampling. The bootstrap variability will thus be smaller than it would if y_1 and y_2 were allowed to vary from one bootstrap sample to another. Another problem is that the residuals $\hat{\varepsilon}_t$ tend to be smaller than the error terms ε_t , due to the effect of fitting. Maybe we need some inflation of the residuals to compensate for the deflation. All of these points are considered in Chapter VI.

CHAPTER V

FORECASTING AN AR(2) MODEL USING THE BOOTSTRAP

Chapter Introduction

After we adopt an appropriate model, we may wish to use it to forecast future values of the time series. In this chapter, we will apply the bootstrap method to the AR(2) model to determine the multi-period bootstrap forecasting error.

We use the same daily readings of the viscosity data set to indicate how to apply the bootstrap to develop standard errors for multi-period forecasts in the second-order autoregressive model;

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad t=1,2,\dots,85$$

Note that we use the data points only up to $t=85$. The observed values beyond 85 will be used to indicate the future actuals. The model's bootstrap forecasts show small bias and small random errors.

Let us look at the autoregressive moving average process of order (p,q) ;

$$\begin{aligned} y_t = & \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \\ & \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \end{aligned} \quad (V-1)$$

An appropriate ARMA(p,q) model may be chosen from the iterative cycle of identification, estimation, and diagnostic checking

procedures. Its coefficients may be estimated by the least squares and/or maximum likelihood estimation procedures, and the residuals are observed, namely the difference between actual and fitted values.

Based on the assumption that the fitted model is the true model for the data generating process, the bootstrap resampling method generates pseudo-random samples both for the past and for the future. Now then we refit the model using the past pseudo-random samples, and use them to forecast the pseudo-random samples for the future. We can observe the forecast errors through this artificial simulation experiment. We analyze the Monte Carlo distribution of such errors to approximate the distribution of the unobservable stochastic errors in the real forecasts. The bootstrap forecast procedure is this approximation.

Data are available for $t=1,2,\dots,n$. The coefficients are estimated as $\hat{\delta}$, $\hat{\phi}$, and $\hat{\theta}$ by some statistical procedure mentioned earlier. Assume that we are at the time origin n and wish to forecast y_{n+h} , which is the value of the time series to be observed at time $n+h$. Generally, we have to build up the forecasts of y_{n+h} recursively from the forecasts of y_{n+1} , y_{n+2} , ..., y_{n+h-1} . The forecast \hat{y}_{n+h} is obtained in the following way.

- i. Use the observed values y_{n+i} for $i \leq 0$,
- ii. Replace the unobserved values y_{n+i} by their forecasts \hat{y}_{n+i} for $i > 0$,
- iii. Use the known values $\hat{\varepsilon}_{n+i}$ for $i \leq 0$,
- iv. Since ε_t is white noise, set ε_{n+i} equal to zero for $i > 0$.

We now generate the bootstrap pseudo-random samples for the past, namely periods $t=1,2,\dots,n$, through the bootstrap simulation procedure. Let us denote these pseudo-random samples by stars:

$$y_1^*, y_2^*, \dots, y_n^* \quad (V-2)$$

In the same simulation world, we generate the pseudo-random samples for the future:

$$y_{n+1}^*, y_{n+2}^*, \dots, y_{n+h}^* \quad (V-3)$$

The construction is recursive: $y_1^* = y_1, y_2^* = y_2, \dots, y_p^* = y_p$, and

$$y_t^* = \hat{\delta} + \hat{\phi}_1 y_{t-1}^* + \hat{\phi}_2 y_{t-2}^* + \dots + \hat{\phi}_p y_{t-p}^* + \hat{\varepsilon}_t^* + \hat{\theta}_1 \varepsilon_{t-1}^* + \hat{\theta}_2 \varepsilon_{t-2}^* + \dots + \hat{\theta}_q \varepsilon_{t-q}^*, \quad \forall t = p+1, p+2, \dots, n+h \quad (V-4)$$

where the ε_t^* are independent with common distribution \hat{F}_n , which is the empirical distribution of the residuals.

We suppose the pseudo-random samples (V-2) come from the model (V-1) with unknown coefficient parameters. Now use the previous estimation methods to find the estimated coefficients from the pseudo-random samples (V-2). These estimates will be denoted by stars: $\hat{\delta}^*$, $\hat{\phi}^*$, and $\hat{\theta}^*$. Next, we use the previous forecasting procedure to generate simulated forecasts for period $n+h$; let us denote these by \hat{y}_{n+h}^* . Then we can find the distribution of the following pseudo-random errors

$$y_{n+h}^* - \hat{y}_{n+h}^* \quad (V-5)$$

This approximation is the bootstrap. We repeat the procedure some number of times to compute the distribution of (V-5) and see what happens.

Conventional Forecasting Procedures

We already mentioned that we selected 85 daily readings of the viscosity (see Table I) for an illustration of the bootstrap procedure for determining forecast errors. The sample autocorrelation function dies down in a damped sine-wave fashion, and the sample partial autocorrelation function cuts off after lag 2. Thus we conclude that the model AR(2)

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad t=1,2,\dots,85 \quad (V-7)$$

adequately describes the viscosity data. The following two equations are the estimated models from the Box-Jenkins methodology using the least squares and maximum likelihood estimation procedures, respectively.

$$\begin{aligned} \hat{y}_t &= 26.7167 + 0.646054 y_{t-1} - 0.412669 y_{t-2}, \\ \hat{\sigma}_\varepsilon^2 &= 4.92357 \text{ (LSE)} \end{aligned} \quad (V-8)$$

$$\begin{aligned} \hat{y}_t &= 26.1421 + 0.725174 y_{t-1} - 0.474156 y_{t-2}, \\ \hat{\sigma}_\varepsilon^2 &= 4.55495 \text{ (MLE)} \end{aligned} \quad (V-9)$$

These forecasting equations use historical data up to and including the current time period $n=85$.

Suppose we wish to forecast the time series value for period 86 in the viscosity example using equation (V-8). Since y_{86} is unknown, we assume ε_{86} to be equal to its expected value of zero. The results of the forecast calculations are given by

$$\begin{aligned} \hat{y}_{86} &= \hat{\delta} + \hat{\phi}_1 y_{85} + \hat{\phi}_2 y_{84} \\ &= 26.7167 + (0.646054)(33.2898) - (0.412669)(34.4787) = 33.9950. \end{aligned}$$

Similarly,

$$\begin{aligned}\hat{y}_{87} &= \hat{\delta} + \hat{\phi}_1 y_{86} + \hat{\phi}_2 y_{85} \\ &= 26.7167 + (0.646054)(33.8950) - (0.412669)(33.2898) = 34.9416.\end{aligned}$$

$$\begin{aligned}\vdots \\ \hat{y}_{97} &= \hat{\delta} + \hat{\phi}_1 y_{96} + \hat{\phi}_2 y_{95} \\ &= 26.7167 + (0.646054)(34.8489) - (0.412669)(34.8665) = 34.8426.\end{aligned}$$

We obtain the maximum likelihood forecast values using equation (V-9) in the same way (see Table I, and column (2) of Tables V and VI).

As a second step, let us look at the way we obtain the forecast error variance from the conventional method. The ARMA(p,q) process can be written as

$$\phi_p(B)y_t = \theta_q(B)\varepsilon_t,$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, and

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

are operator polynomials in the backward shift operator B, where $B^r y_t = y_{t-r}$. Now the ARMA(p,q) process $\phi_p(B)y_t = \theta_q(B)\varepsilon_t$ can be written as an infinite moving average $y_t = c(B)\varepsilon_t$, where $c(B) = c_0 + c_1 B + c_2 B^2 + \dots$. Knowing the values of the ϕ 's and θ 's in a particular process, the weights, c's are obtained by equating coefficients of B^j , $j=1,2,\dots$, in $\phi_p(B)c(B) = \theta_q(B)$:

$$c_0 = 1$$

$$c_1 = \phi_1 + \theta_1$$

$$c_2 = \phi_1 c_1 + \phi_2 + \theta_2$$

$$\vdots$$

$$c_j = \phi_1 c_{j-1} + \dots + \phi_p c_{j-p} + \theta_q.$$

Both the weight c_j and the variance estimate $\hat{\sigma}_\varepsilon^2$ are then combined to calculate the estimated forecast error variance for the h^{th} period in the future, $\hat{v}(h)$:

$$\hat{v}(h) = \hat{\sigma}_\varepsilon^2 \sum_{j=0}^{h-1} c_j^2, \quad c_0 = 1.$$

This is the method used to generate the conventional forecast standard errors shown in column (3) of Tables V and VI. For mathematical details, see Chapter 5 of Box and Jenkins (1976), also Chapter 4 of Granger and Newbold (1977).

The Bootstrap Forecasting Procedure

In this section we will use the bootstrap forecasting procedure to develop standard errors for multi-period forecasts. Through this bootstrap simulation experiment, we can observe both the simulated future actuals and the simulated forecast values. Thus we can find the forecast error, which is the difference between the two. We will use the AR(2) model to illustrate this idea, with a historical period of $t=1,2,\dots,85$, and a forecast period of $t=86,87,\dots,97$.

As a first step in the bootstrap forecasting procedure, we estimate the coefficients δ , ϕ_1 , and ϕ_2 using the historical data from $t=1,2,\dots,85$, and find the residuals $\hat{\varepsilon}_t$,

$$\begin{aligned} \hat{\varepsilon}_t &= y_t - \hat{y}_t \\ &= y_t - (\hat{\delta} + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2}), \quad t=3,4,\dots,85. \end{aligned}$$

Let \hat{F}_n be the empirical distribution of the residuals, $\{\hat{\varepsilon}_t, t=3,4,\dots,85\}$. Make 95 independent draws of the starred residuals ε_t^* , $t=3,4,\dots,97$ from \hat{F}_n . Construct the bootstrap forecast pseudo-random samples with the resampled residuals:

- i. $y_1^* = y_1, y_2^* = y_2$
- ii. $y_t^* = \hat{\delta} + \hat{\phi}_1 y_{t-1}^* + \hat{\phi}_2 y_{t-2}^* + \varepsilon_t^*$ for $t=3,4,\dots,97$

For $t=1,2,\dots,85$, the y_t^* are the simulated historical data. For $t=86,87,\dots,97$, the y_t^* are the simulated future actuals. Find the starred estimates $\hat{\delta}^*$, $\hat{\phi}_1^*$, and $\hat{\phi}_2^*$ using the starred simulated historical data. We now find the bootstrap forecasts from the simulated historical data instead of the real viscosity data (see Figure 2).

These \hat{y}_t^* are the simulated bootstrap forecasts for $t=86,87,\dots,97$. Notice that \hat{y}_t^* do not incorporate $\hat{\varepsilon}_t^*$ while y_t^* do. The bootstrap forecast error is the difference between the simulated future actuals y_t^* and the simulated forecasts \hat{y}_t^* . We repeat the above procedure 100 times, getting new starred residuals on each repetition, to construct the distribution of

- i. The simulated future actuals y_t^* , $t=86,87,\dots,97$,
- ii. The simulated forecasts \hat{y}_t^* , $t=86,87,\dots,97$,
- iii. The simulated bootstrap forecast error $y_t^* - \hat{y}_t^*$, $t=86,87,\dots,97$.

Tables V and VI show the results of the bootstrap forecasting experiment using least squares and maximum likelihood estimates, respectively. Column (1) shows the real viscosity data. Note that two more actual values were added for $t=96$ and 97 in addition to the observations in Table I. Columns (2) and (3) show the forecast values and standard error of forecasts obtained by the conventional formulae. In doing the bootstrap forecasting procedure, we have set

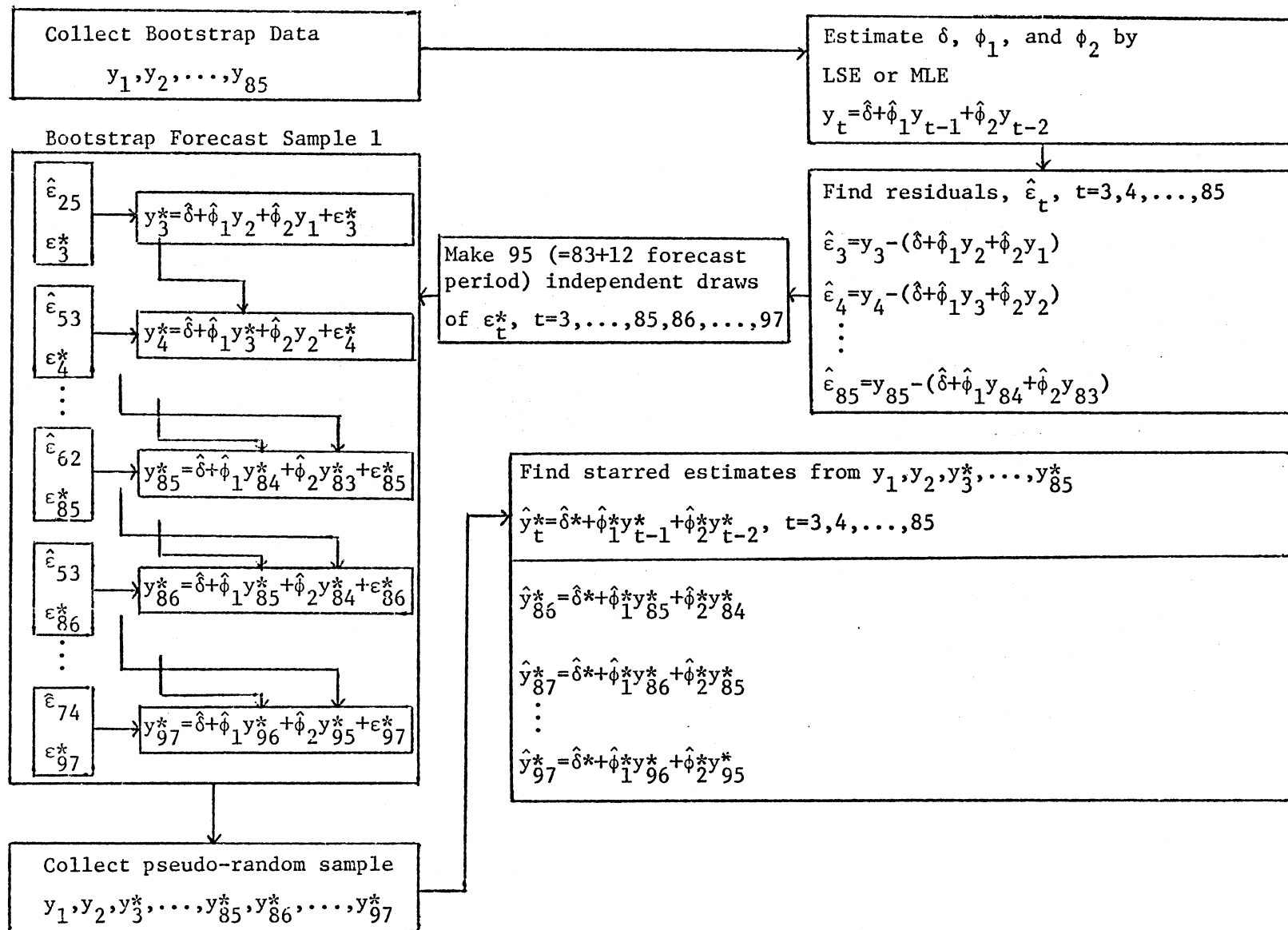


Figure 2. Bootstrap Forecast Model

TABLE V

BOOTSTRAP FORECAST EXPERIMENT FOR MODEL (V-7). (THERE ARE 100 B/S REPLICATIONS.
ESTIMATION IS BY LEAST SQUARES, USING THE FIRST SEED.)

t=n+h n=85 h=1,2,...,12	Least Square Estimates			Bootstrap		
	(1) Actuals y_t , t=86,87,...,97	(2) Forecasts	(3) Standard Error of Forecasts	(4) Sample Mean Actuals	(5) Sample Mean Forecasts	(6) SD of Forecast Error
86	33.9252	33.9950	2.2189	35.0765	34.9173	2.1052
87	36.1036	34.9416	2.6417	35.0191	34.8388	1.9603
88	36.7351	35.2622	2.6417	35.0800	34.9003	1.7064
89	35.4576	35.0786	2.7057	35.0532	34.9765	1.9578
90	37.5924	34.8278	2.7325	35.2126	34.9044	1.9323
91	34.4895	34.7414	2.7325	35.1945	34.9743	1.8441
92	39.1692	34.7892	2.7369	35.2773	34.8604	2.0055
93	35.8242	34.8557	2.7388	35.2611	34.9318	2.1618
94	32.3875	34.8789	2.7388	35.2874	34.9027	1.6989
95	31.2846	34.8665	2.7391	35.1778	34.9532	1.9110
96	33.5576	34.8489	2.7392	35.1670	34.8884	2.0635
97	35.6008	34.8426	2.7392	35.1850	34.9180	1.8753

TABLE VI

BOOTSTRAP FORECAST EXPERIMENT FOR MODEL (V-7). (THERE ARE 100 B/S REPLICATIONS.
ESTIMATION IS BY MAXIMUM LIKELIHOOD, USING THE FIRST SEED.)

	Maximum Likelihood Estimates			Bootstrap		
	(1)	(2)	(3)	(4)	(5)	(6)
t=n+h n=85 h=1,2,...,12	Actuals y_t , t=86,87,...,97	Forecasts	Standard Error of Forecasts	Sample Mean Actuals	Sample Mean Forecasts	SD of Forecast Error
86	33.9252	33.9342	2.1342	34.8389	34.7564	2.0261
87	36.1036	34.9657	2.6363	34.9997	34.9058	2.0896
88	36.7351	35.4082	2.6387	35.3103	35.0558	2.2506
89	35.4576	35.2399	2.7184	35.4441	35.1825	1.7883
90	37.5924	34.9081	2.7690	35.1412	35.1389	1.7089
91	34.4895	34.7473	2.7699	34.8930	34.8229	2.0086
92	39.1692	34.7880	2.7769	35.1036	34.7790	2.0738
93	35.8242	34.8938	2.7826	34.8300	35.1133	1.8539
94	32.3875	34.9512	2.7828	34.7458	34.7113	1.9261
95	31.2846	34.9426	2.7834	34.9128	34.8121	1.9379
96	33.5576	34.9092	2.7840	34.1224	34.9928	2.0664
97	35.6008	34.8891	2.7841	35.2322	35.0389	1.8264

up a bootstrap simulation model, where the parameters and the distribution of the error terms are all known. The sample mean of the simulated future actuals,

$$\frac{1}{100} \sum_{b=1}^{100} y_t^{*(b)}, \quad t=86,87,\dots,97,$$

the sample mean of the simulated forecasts,

$$\frac{1}{100} \sum_{b=1}^{100} \hat{y}_t^{*(b)}, \quad t=86,87,\dots,97.$$

and the standard deviation of the simulated forecast errors,

$$\left[\frac{1}{100-1} \sum_{b=1}^{100} (D_t^{*(b)} - D_t^{*(\cdot)})^2 \right]^{\frac{1}{2}}, \quad D_t^{*(b)} = y_t^{*(b)} - \hat{y}_t^{*(b)},$$

$$D_t^{*(\cdot)} = \frac{1}{100} \sum_{b=1}^{100} D_t^{*(b)}, \quad t=86,87,\dots,97,$$

are displayed in columns (4), (5), and (6). The bootstrap measures of random error in forecasts are the standard deviations in column (6).

It is obvious that as we look further into the future, bigger forecast standard errors are obtained from the conventional methods as shown in column (3) of Tables V and VI. But the bootstrap estimates of forecast standard errors do not show this pattern as can be seen in column (6). This is somewhat disconcerting, although the theoretical forecast error remains relatively flat beyond a lead time of about eight time periods.

First, let us look at Table V. Column (4) is uniformly a bit larger than column (5), indicating a small bias in the bootstrap forecasting. For example, the forecasts tend to be only 1.2% smaller for $t=92$ from the comparison of column (4) and (5). The other

comparisons are less than this. The standard deviations in column (6) are fairly low, compared with the mean values. Column (3) is uniformly bigger than column (6), suggesting that the conventional standard errors might be overly conservative when the error distribution is discrete. The bootstrap forecasts are showing both the small biases and small random errors.

Column (4) is a bit bigger than column (5) again in Table VI, except for $t=93$. This table can be interpreted in a similar way. The bootstrap forecasts through the maximum likelihood estimates are subject to small biases and small chance errors.

Similar bootstrap forecasting experimental results are shown from the second seed (see Tables XV and XVI, Appendix C). The conclusions from these are consistent with those drawn from Tables V and VI above.

We should emphasize again, as we did at the end of the previous chapter, that y_1 and y_2 have been held fixed and the fitted residuals have not been inflated. This may account for the tendency for column (6) to be smaller than column (3).

CHAPTER VI

QUALITY OF THE BOOTSTRAP ESTIMATES

Chapter Introduction

One of our objectives in subsequent sections of this chapter is to examine the quality of the bootstrap estimates of standard error by means of an extensive simulation study using moderate sample sizes, based on the maximum likelihood estimation procedure. Ansley and Newbold (1980) recommend that if a researcher was to choose one estimation method the choice should probably be the maximum likelihood. We felt that, because the conventional method of standard error estimation uses asymptotic results, the bootstrap method might outperform it in a small sample setting. This was the primary motivation for this simulation study.

In Chapter IV, we performed a bootstrap simulation experiment to show how, in the context of the second-order autoregressive model, the conventional formulae for asymptotic standard errors overestimate the actual variability of the coefficient estimates. The rationale underlying the bootstrap methodology was:

- i. We wish to have an estimate of the precision of $\hat{\phi}_i$, $i=1,2$,
- ii. We would like to use $\sigma_F(\phi_i)$, $i=1,2$, where F is the true probability distribution function,
- iii. We don't know F , so instead we estimate it with $F = \hat{F}_n$, the empirical distribution. In other words, we use the same basic

method--a simple substitution of \hat{F}_n for the unknown true distribution F .

Basically, the variance of the error terms is unknown, so it has to be replaced by some estimate from the data. This replacement is acceptable if the sample size is large enough, but how large is sufficiently large? Ansley and Newbold (1980) considered sample sizes of 50 and 100 in their studies. Box and Jenkins (1976) suggested that at least 50 observations are desirable for fitting simple ARMA models. We select a sample size of $n=52$ throughout the investigations in the simulation trials and bootstrap experiment.

Our bootstrap method in Chapter IV was based on resampling the original observations in a suitable way, where all the ingredients of the simulating equation were known:

- i. The first two observations y_1 and y_2 were held fixed,
- ii. The error terms ε_t were independent with common empirical distribution of the residuals $\hat{\varepsilon}_t$,
- iii. The parameters were fixed at their estimated values.

Two problems with the bootstrap estimates of standard error obtained in Chapter IV are:

- i. y_1 and y_2 were held fixed when generating the bootstrap pseudo-random samples,
- ii. The residuals $\hat{\varepsilon}_t$ tend to be smaller than the true error terms ε_t .

An easy fix for (i) is to generate observations of normal random variables, with appropriate means and variances for Y_1 and Y_2 . For (ii), we can inflate the residuals to compensate for the deflation due to the effect of fitting, using the inflation factor $\sqrt{52/(52-3)}$. These fixes were performed in this simulation study.

The following is a brief outline of the steps in the simulation study:

- i. Generate 100 time series, each of length 52, from a specified AR(2) process with normal errors,
- ii. For each generated series, fit coefficients using maximum likelihood, and obtain standard errors by
 - a. The conventional method based on large sample theory,
 - b. Bootstrapping.
- iii. Calculate the "true" observed variability in $\hat{\mu}$, $\hat{\phi}_1$, and $\hat{\phi}_2$ over the 100 trials,
- iv. Compare the conventional method and the bootstrap method of standard error assessment both with each other and with the "true" variability.

Simulation Procedure for the Trial and Bootstrap Experiment

The series analyzed in the 100 trial experiment are generated from the following second-order autoregressive model

$$\begin{aligned}
 y_t &= \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
 &= 26.5477 + 0.6478245 y_{t-1} - 0.407965 y_{t-2} + \varepsilon_t
 \end{aligned}
 \tag{VI-1}$$

where ε_t is a normal random vector with mean zero and variance 4.83772 (see Figure 3). These parameter values were chosen as representative of the least squares and maximum likelihood estimates resulting from the analysis of the viscosity data first presented in Chapter IV. The trial experiment serves to generate the observed data. The simulation experiment involves a nested iteration: at the

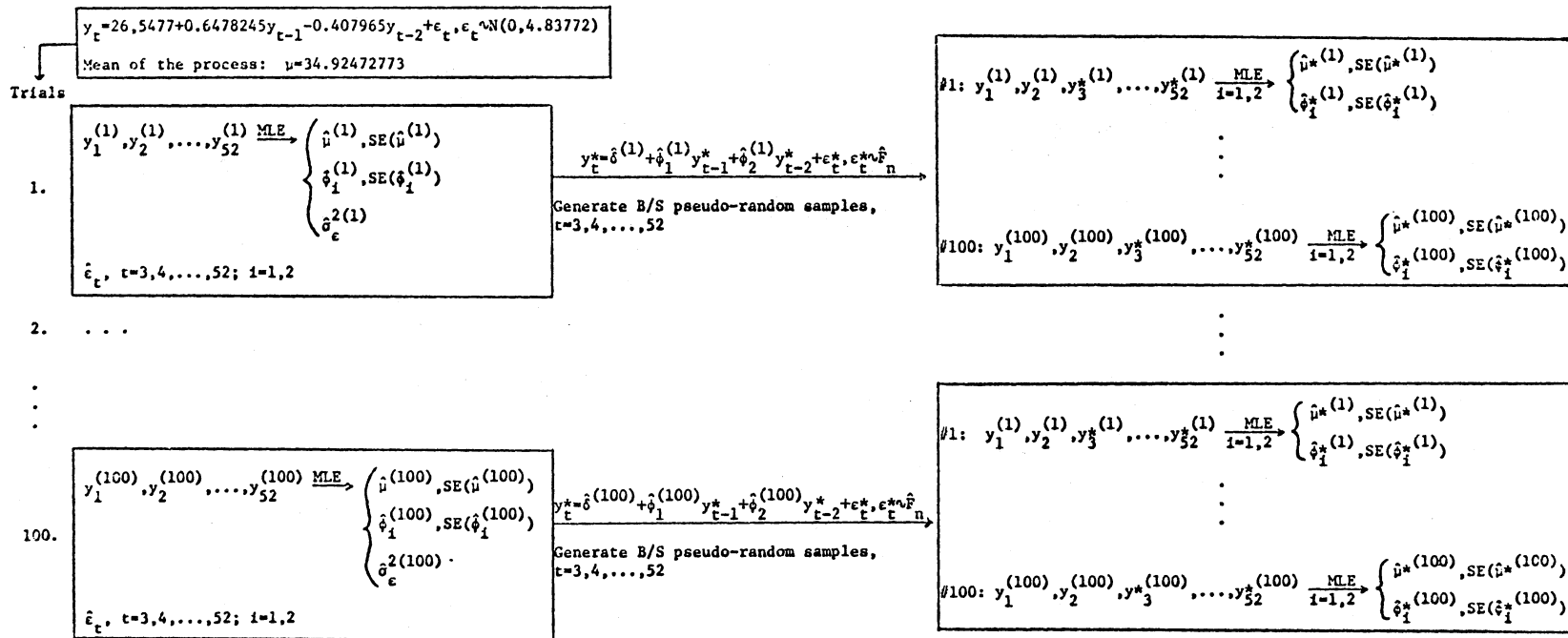


Figure 3. Generating 100 Trials and 100 B/S Pseudo-Random Samples for Each Trial

trial experiment observed data sets are built up and, for each of these, a bootstrap experiment generates 100 pseudo-random samples for a bootstrap estimate of the standard errors. The sample size is 52 throughout.

The trial experiment: In the first step of the simulation experiment the first two observations y_1 and y_2 are generated in the following way:

$$i. \quad Y_1 \sim N(\mu, \sigma_y^2) \quad (VI-2)$$

$$ii. \quad Y_2 \sim N\{\mu + \rho(y_1 - \mu), \sigma_y^2(1 - \rho^2)\} \quad (VI-3)$$

Here $\mu = \frac{\delta}{1 - \phi_1 - \phi_2}$ is the unconditional mean, and

$$\sigma_y^2 = \left(\frac{1 - \phi_2}{1 + \phi_2} \right) \frac{\sigma_\varepsilon^2}{[(1 - \phi_2)^2 - \phi_1^2]} \quad (VI-4)$$

is the unconditional variance of the process y_t . Similarly, $\mu + \rho(y_1 - \mu)$ and $\sigma_y^2(1 - \rho^2)$, with $\rho = \frac{\phi_1}{1 - \phi_2}$ are the conditional mean and variance respectively of Y_2 given $Y_1 = y_1$. Of course ϕ_1 , ϕ_2 , and σ_ε^2 have the values given in the original model (VI-1). We emphasize that the variable Y_2 is generated from the conditional normal distribution of Y_2 given $Y_1 = y_1$. See p. 56 of Box and Jenkins (1976) for more details. Therefore, in each trial the first two observed values are random instead of being fixed.

The observed data can now be built up iteratively from the model (VI-1) for $t=3,4,\dots,52$. This procedure was repeated 100 times. On each repetition, a new set of observed data was generated,

$$y_1^{(j)}, y_2^{(j)}, \dots, y_{52}^{(j)}, \quad j=1,2,\dots,100 \text{ trials.}$$

The parameter estimates are obtained through the maximum likelihood estimation procedure from each of these data sets,

$$\hat{\mu}_{MLE}^{(j)}, \hat{\phi}_{i,MLE}^{(j)} \quad i=1,2; j=1,2,\dots,100 \quad (\text{VI-5})$$

Also the conventional estimates of the standard errors,

$$SE_j(\hat{\mu}^{(j)}), SE_j(\hat{\phi}_i^{(j)}), \quad i=1,2; j=1,2,\dots,100 \quad (\text{VI-6})$$

the residuals $\hat{\epsilon}_t$ and the variance estimates of the error terms $\hat{\sigma}_\epsilon^2$, are obtained and kept for the bootstrap simulation experiment.

The bootstrap experiment: On each pass in the trial experiment generate the first two bootstrap samples y_1 and y_2 through normal distributions (VI-2) and (VI-3), but using the fitted values $\hat{\mu}$, $\hat{\phi}_1$, $\hat{\phi}_2$, and $\hat{\sigma}_\epsilon^2$ in place of μ , ϕ_1 , ϕ_2 , and σ_ϵ^2 . Note that the generating parameters differ from one pass to the next. At each pass generate ϵ_t^* for $t=3,4,\dots,52$ as independent draws from the inflated residuals $\hat{\epsilon}_t$, $t=3,4,\dots,52$. Construct then a starred bootstrap data set with the resampled residuals as in Figure 1, Chapter IV:

$$y_t^* = \hat{\delta}^{(j)} + \hat{\phi}_1^{(j)} y_{t-1}^* + \hat{\phi}_2^{(j)} y_{t-2}^* + \epsilon_t^*, \quad \begin{matrix} t=3,4,\dots,52; \\ j=1,2,\dots,100 \end{matrix} \quad (\text{VI-7})$$

Obtain the starred parameter estimates and their standard errors by the maximum likelihood estimation procedure. Repeat this procedure 100 times for each of the original trials. Therefore 100 sets of the bootstrap pseudo-random samples, and 100 sets of starred parameter estimates and their standard errors are generated for each trial.

The next step of the simulation experiment involves the computations and storage of the appropriate results for interpretation. Table VII summarizes the computations for the 100 replications of the

TABLE VII
COMPUTATIONS FOR THE TRIAL AND BOOTSTRAP EXPERI-
MENT (THERE ARE 100 TRIALS AND 100 B/S
SAMPLES FOR EACH TRIAL)

Parameters		MLE		(4) True SE	Bootstrap		
(0)	(1)	(2) Estimate Mean	(3) Conv. SE		(5) B/S Sample Mean	(6) B/S SD	(7) RMS Conv. SE
μ	34.92472773	$\bar{\mu} = \frac{1}{100} \sum_{j=1}^{100} \hat{\mu}^{(j)}$	$\hat{\sigma}_{\hat{\mu}} = \left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\mu}^{(j)}) \right]^{\frac{1}{2}}$	$\sigma_{\hat{\mu}} = \left[\frac{1}{100-1} \sum_{j=1}^{100} (\hat{\mu}^{(j)} - \bar{\mu})^2 \right]^{\frac{1}{2}}$	$\bar{\mu}^{(*)} = \frac{1}{100} \sum_{j=1}^{100} \hat{\mu}^{*(j)}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SD_j^2(\hat{\mu}^{*(j)}) \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\mu}^{*(j)}) \right]^{\frac{1}{2}}$
ϕ_1	0.6478245	$\bar{\phi}_1 = \frac{1}{100} \sum_{j=1}^{100} \hat{\phi}_1^{(j)}$	$\hat{\sigma}_{\hat{\phi}_1} = \left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\phi}_1^{(j)}) \right]^{\frac{1}{2}}$	$\sigma_{\hat{\phi}_1} = \left[\frac{1}{100-1} \sum_{j=1}^{100} (\hat{\phi}_1^{(j)} - \bar{\phi}_1)^2 \right]^{\frac{1}{2}}$	$\bar{\phi}_1^{(*)} = \frac{1}{100} \sum_{j=1}^{100} \hat{\phi}_1^{*(j)}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SD_j^2(\hat{\phi}_1^{*(j)}) \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\phi}_1^{*(j)}) \right]^{\frac{1}{2}}$
ϕ_2	-0.407965	$\bar{\phi}_2 = \frac{1}{100} \sum_{j=1}^{100} \hat{\phi}_2^{(j)}$	$\hat{\sigma}_{\hat{\phi}_2} = \left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\phi}_2^{(j)}) \right]^{\frac{1}{2}}$	$\sigma_{\hat{\phi}_2} = \left[\frac{1}{100-1} \sum_{j=1}^{100} (\hat{\phi}_2^{(j)} - \bar{\phi}_2)^2 \right]^{\frac{1}{2}}$	$\bar{\phi}_2^{(*)} = \frac{1}{100} \sum_{j=1}^{100} \hat{\phi}_2^{*(j)}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SD_j^2(\hat{\phi}_2^{*(j)}) \right]^{\frac{1}{2}}$	$\left[\frac{1}{100} \sum_{j=1}^{100} SE_j^2(\hat{\phi}_2^{*(j)}) \right]^{\frac{1}{2}}$

trial experiment and the 100 bootstrap pseudo-random samples for each of these trials just described above. The analyses of the computations obtained are given in the next section. The entries in columns (2), (3), and (4) are from the 100 sets of the observed data. Actually those are from the statistics (VI-5) and (VI-6). The entries in columns (5), (6), and (7) are from the 10,000 bootstrap pseudo-random samples. The following are the details of computations for the last three columns in Table VII. For example, take the 100 sets of bootstrap pseudo-random samples from the first trial. The following statistics can be obtained from the bootstrap parameter estimates:

$$i. \quad \bar{\hat{\mu}}^*(1) = \frac{1}{100} \sum_{b=1}^{100} \hat{\mu}^*(b), \quad \bar{\hat{\phi}}_i^*(1) = \frac{1}{100} \sum_{b=1}^{100} \hat{\phi}_i^*(b)$$

$$ii. \quad SD_1(\hat{\mu}^*) = \left[\frac{1}{100-1} \sum_{b=1}^{100} (\hat{\mu}^*(b) - \bar{\hat{\mu}}^*)^2 \right]^{1/2},$$

$$SD_1(\hat{\phi}_i^*) = \left[\frac{1}{100-1} \sum_{b=1}^{100} (\hat{\phi}_i^*(b) - \bar{\hat{\phi}}_i^*)^2 \right]^{1/2},$$

$$iii. \quad SE_1(\hat{\mu}^*) = \left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2(\hat{\mu}^*) \right]^{1/2},$$

$$SE_1(\hat{\phi}_i^*) = \left[\frac{1}{100} \sum_{b=1}^{100} SE_b^2(\hat{\phi}_i^*) \right]^{1/2}$$

where $i=1,2$; $b=1,2,\dots,100$ bootstrap samples. On each pass of the trial experiment, a new set of the above statistics are generated. The analyses of those statistics are the entries in columns (5), (6), and (7) in Table VIII. As before, "SE" denotes conventional standard error computation and "SD" denotes bootstrap standard error.

TABLE VIII

TRIAL AND BOOTSTRAP EXPERIMENT (THERE ARE 100 TRIALS AND
100 B/S SAMPLES FOR EACH TRIAL. ESTIMATION
IS BY MAXIMUM LIKELIHOOD.)

Parameters		MLE			Bootstrap					Ratios								
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)						
		Este	Mean	Conv.	SE	True	SE	B/S	SMean	B/S	SD	RMS	C.SE	(3)/(4)	(6)/(4)	(7)/(4)	(2)/(1)	(5)/(1)
μ	34.92472773	34.9174	0.3997	0.4262	34.9098	0.4031	0.4005	0.94	0.95	0.94	1.00	1.00						
ϕ_1	0.6478245	0.6086	0.1299	0.1254	0.5839	0.1298	0.1282	1.04	1.03	1.02	0.94	0.90						
ϕ_2	-0.407965	-0.4138	0.1304	0.1296	-0.4228	0.1236	0.1285	1.01	0.95	0.99	1.01	1.04						

Results of the Simulation Experiment

The numerical results of the simulation experiment are reported in Table VIII. The entries in this table correspond to those of Table VII. We compare the performance of conventional maximum likelihood estimates of standard errors to the performance of a bootstrap procedure in the setting of 100 generated sets of data from the model (VI-1) along with 100 bootstrap pseudo-random samples at each trial.

The parameters μ , ϕ_1 , and ϕ_2 and their true values are listed in columns (0) and (1) of Table VIII. The 100 trials serve to generate the observed data. The estimates of the parameters from these data are obtained by the maximum likelihood procedure; the results are summarized in columns (2) and (3). Column (2) shows the sample means of the maximum likelihood estimates over the 100 trials. These means are close to the true values in column (1) (see column (1)). The sample mean for ϕ_1 is 6% off the true value of ϕ_1 . Column (3) shows the typical variability indicated by the conventional maximum likelihood approach, as applied to the 100 observed sets of data. Column (4) shows for each parameter the sample standard deviation for the estimates from the 100 observed sets of data. These values represent the "true" variability of the maximum likelihood estimates for parameters μ , ϕ_1 , and ϕ_2 , in the simulation world of the trial experiment. These provide a baseline for the comparison of the conventional and bootstrap estimates.

Columns (5) to (7) contain the results for the bootstrap experiment. For each pass in the trial experiment we obtained 100 sets of bootstrap samples as we mentioned, and we computed the means, bootstrap standard deviations, and root-mean-square conventional

standard errors of the coefficient estimates thus obtained. Column (5) shows the sample means in the simulation world of the bootstrap experiment, i.e., mean of the parameter estimates from 10,000 bootstrap pseudo-random samples, under the empirical error distribution \hat{F}_n ; the entries in this column are close to those in columns (1) and (2). The bootstrap sample mean for ϕ_1 is 10% off the true value of ϕ_1 (see column (12)). Column (6) gives the bootstrap estimates of variability in the parameter estimates, i.e., the variability of the maximum likelihood estimates for parameters μ , ϕ_1 , and ϕ_2 in the simulation world of the bootstrap experiment, under the empirical error distribution \hat{F}_n . Column (7) shows the variability indicated by the conventional maximum likelihood approach, as applied to the 10,000 typical starred data sets, under the empirical error distribution \hat{F}_n . The results do not confirm those of Chapter IV: the asymptotic standard errors in columns (3) and (7) are close to the true standard errors in column (4), and the bootstrap estimates of the standard errors in column (6) do not provide an improvement. These interpretations are pursued in columns (8) to (10), which contain the ratios of the entries in columns (3), (6), and (7), respectively, to those in column (4). These ratios are close to one. Some entries are a bit less than one and some are a bit larger than one. On the whole, both the conventional maximum likelihood and bootstrap estimates of the standard errors appear to be performing quite well. To choose either one is not appropriate since both look like good approximations to the true standard errors, and a more complete simulation study would be necessary before firm conclusions could be drawn.

Evaluation of the Random Number Generators

One important aspect of our simulation experiments is the adequacy of the random number generator used to construct pseudo-random samples. Some questions relevant to our study would be the following:

- i. Was the normal error generator (the 50 ε_t 's for each of the 100 trials) working satisfactorily?
- ii. Was the normal generator working satisfactorily for the first two observations in each trial?
- iii. Was the normal generator working satisfactorily for the first two observations in the bootstrap samples?
- iv. Was the uniform generator working satisfactorily?

The tests of normality test whether or not the observations fit, or are consistent with, the assumption that they are from a specific normal density. We used a confidence interval method for tests on the mean, and χ^2 test for the tests on the variance. For (i), histogram plot of the empirical distribution of the 5,000 observed values exhibited symmetric bell-shape, and the confidence interval and χ^2 tests of normality for true mean and variance were passed. Similarly, for (ii), histogram of the 100 observations showed symmetric bell-shape, and the test of normality was passed. For (iii), we need to perform the test of normality 100 times for 100 different normal densities which are from 100 trials. Ninety-eight times passed for tests on the mean, and 99 times passed for tests on the variance. In addition to these tests on the parameters of the distribution, we also performed a Kolmogorov-Smirnov test for normality of the

distribution. This passed in all cases. For (iv), the uniform generator generated numbers from the uniform distribution on the interval (0,1) satisfactorily, based on a χ^2 goodness-of-fit test.

CHAPTER VII

BOOTSTRAPPING AN ECONOMETRIC MODEL

Chapter Introduction

This chapter is mainly concerned with presenting the bootstrap in the context of econometric equations describing the unemployment rate and individual income tax. Each model is fitted using conventional least squares. In contrast with the results for the AR(2) model in Chapter IV, the conventional estimate of the standard errors appears to do well when applied to a particular finite sample.

Let us assume that there exists a linear relationship between an endogenous variable Y and p exogenous variables $X_{1t}, X_{2t}, \dots, X_{pt}$ and an error term ε_t . If we have a sample of n observations on Y and the X 's we can write

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_p X_{pt} + \varepsilon_t, \quad t=1,2,\dots,n,$$

where $\varepsilon_t \sim (0, \sigma^2)$, and X_{it} , $i=1,2,\dots,p$, are considered fixed. In practice, the standard errors are used to develop the form of the regression equation by testing the hypothesis that one of regression coefficients is zero. Considering the i^{th} coefficient, the null hypothesis to be tested is

$$H_0: \beta_i = 0, \quad i=0,1,2,\dots,p.$$

Thus the standard errors are used to decide whether to enter or remove variables in the regression equation. Also, the standard errors are used to find confidence intervals for the regression coefficients. Since these tests are important to the implications of the regression model, the more reliable standard error estimates for the coefficients should be preferred.

When the variance-covariance matrix of the error terms is unknown it must be estimated from the data. Then the conventional asymptotic formula for calculating the variance-covariance matrix of the regression coefficients cannot be fully dependable. The principal part of the bootstrap is the development and implementation of a computer-based methodology which checks the accuracy of these asymptotic standard errors of the regression coefficients for any actual use of the model.

The model we are going to study is

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim (0, \sigma^2 \underline{I})$$

The distributional form of the disturbances is unspecified. The expected value of each ε_t is zero, the ε_t are uncorrelated, and the ε_t have a common unknown variance σ^2 . Furthermore, autoregressive structures are permitted (see the Oklahoma unemployment rate model). The coefficients are estimated by the conventional least squares. We obtain the estimated variance-covariance matrix $\hat{\sigma}^2 (\underline{X}'\underline{X})^{-1}$ for the estimated coefficients by substituting $\hat{\sigma}^2$ for σ^2 in the formula for least squares coefficient variance-covariance.

In this chapter, we compare the performance of conventional asymptotic estimates of standard error for regression coefficients to the performance of the bootstrap in the setting of Oklahoma

regional regression equations. The examples studied here are taken from real econometric models, the equations for the Oklahoma unemployment rate and individual income tax. These equations are designed to forecast the rate of unemployment and individual income tax for the state of Oklahoma based on the value of variables representing the level of national activity as well as local conditions. These equations are representative of much current practice in econometric modeling. So, this statistical methodology should be interesting to economists who are planning to build and analyze similar models.

The idea is the same as in Chapter IV. We resample the original observations in a suitable way to construct "bootstrap pseudo-random samples" on which the estimator of interest is exercised. Now then let us consider a linear model,

$$\begin{array}{ccccc} \underline{y} & = & X & \underline{\beta} & + & \underline{\varepsilon} \\ nx1 & & nxp & px1 & & nx1 \end{array}$$

where \underline{y} is an observable $nx1$ random vector,

X is an $n \times p$ matrix of known constants,

$\underline{\beta}$ is a $px1$ vector of unknown constants, and

$\underline{\varepsilon}$ is an unobservable $nx1$ random vector, $\underline{\varepsilon} \sim (0, \sigma^2 I)$.

Usually the variance of the disturbance terms is unknown and must be estimated from the data. Therefore, we should not adopt the usual formula,

$$\hat{\sigma}^2 = \text{Residual Sums of Squares} / (n-p)$$

for calculating the variance of the regression coefficients straightforwardly. If a poor estimate of σ^2 is used to replace σ^2 in the

formula $\text{Var-Cov}(\hat{\beta}) = \sigma^2 (XX)^{-1}$, the resulting variance-covariance estimates will be unreliable. It would be helpful to have more reliable estimates of the variance-covariance of the $\hat{\beta}_i$, $i=1,2,\dots,p$.

Data are available for $t=1,2,\dots,n$. The coefficient $\hat{\beta}$ is estimated by some well-defined statistical procedure, like least squares. Coming now to the bootstrap, when $\hat{\beta}$ is computed, residuals are defined:

$$\hat{\varepsilon} = y - X\hat{\beta}.$$

Let \hat{F}_n be the empirical distribution function of the residuals, assigning mass $\frac{1}{n}$ to each of $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$. Next let us set the coefficients of linear model at $\hat{\beta}$ with common empirical error distribution \hat{F}_n . The exogenous X's are kept fixed. We can then generate "the bootstrap pseudo-data" using this model. These pseudo-data will be denoted by stars:

$$y_1^*, y_2^*, \dots, y_n^*$$

For all $t=1,2,\dots,n$,

$$y_t^* = X_t \hat{\beta} + \varepsilon_t^*,$$

where the ε_t^* are independent with the common distribution function \hat{F}_n .

Now using the previous least squares estimation procedure, compute the least squares estimates from the pseudo-data, $\hat{\beta}^*$. We can compute the distribution of the pseudo-errors $\hat{\beta}^* - \hat{\beta}$ to approximate the distribution of the real errors $\hat{\beta} - \beta$ by the Monte Carlo simulation method.

Conventional Least Squares Estimates

Let us consider the basic linear regression model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}, E(\underline{\varepsilon}) = \underline{0}, \text{Var-Cov}(\underline{\varepsilon}) = \sigma^2 \underline{I}$$

The sum of squares of deviations of the observations from their expected values is then

$$\underline{\varepsilon}'\underline{\varepsilon} = (\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})$$

which, according to the conventional least squares method, is to be minimized by choosing that value of $\underline{\beta}$ which minimizes $\underline{\varepsilon}'\underline{\varepsilon}$. The least squares estimator after minimizing the sum of squares of the residuals is

$$\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}.$$

The least squares estimator is unbiased

$$E(\hat{\underline{\beta}}) = \underline{\beta},$$

and its variance-covariance matrix is

$$\text{Var-Cov}(\hat{\underline{\beta}}) = \sigma^2(\underline{X}'\underline{X})^{-1}.$$

When σ^2 is unknown, we use $\hat{\sigma}^2(\underline{X}'\underline{X})^{-1}$ with σ^2 replaced by some estimate $\hat{\sigma}^2$. When we are dealing with finite samples, the validity of our analysis depends on whether $\hat{\sigma}^2$ is a good estimate for σ^2 or not. If $\hat{\sigma}^2$ is a poor estimate for σ^2 , the standard errors estimated from the variance-covariance matrix $\hat{\sigma}^2(\underline{X}'\underline{X})^{-1}$ may prove to be grossly incorrect. We often use the least squares estimators even when there

are doubts as to the validity of the estimate of σ^2 , as in the case of small sample situations. In such cases, the bootstrap standard errors are more reliable estimates of the standard errors, and give useful diagnostics (see Freedman and Peters (1983a)).

Bootstrapping Oklahoma Unemployment Rate Model

The main objective of this section is to illustrate the bootstrap methodology for measuring the precision of parameter estimates in the model of Oklahoma unemployment rate. The main finding is that for the least squares estimates, the standard errors are dependable when applied to this particular finite sample. The unemployment model fitting is designed to develop proper plans or courses of action in the future through accurate forecasts for regional policy makers, both on the state and local level. The Business and Economic Research Center at Oklahoma State University established the following econometric equation:

$$y_t = \beta_0 + \beta_1 x1_t + \beta_2 y_{t-1} + \beta_3 x2_t + \beta_4 x3_t + \varepsilon_t$$

$$t = 1959, 1960, \dots, 1982,$$

where y_t is the Oklahoma unemployment rate (%),

$x1_t$ is the U.S. unemployment rate (%),

$x2_t$ is the Oklahoma annual personal income (\$1,000,000), and

$x3_t$ is the Oklahoma manufacturing sector's nominal wages (\$1,000).

The observed values of our variables are in Table IX.

The assumptions on the stochastic disturbance terms ε_t are as follows:

TABLE IX
DATA FOR THE OKLAHOMA UNEMPLOYMENT RATE EQUATION

year	t	y_t	$x1_t$	$x2_t$	$x3_t$	y_{t-1}
1958	0	4.5	6.8	3958.38	4.594	.
1959	1	3.7	5.5	4129.92	4.746	4.5
1960	2	4.0	5.5	4343.67	4.812	3.7
1961	3	4.8	6.7	4481.46	4.927	4.0
1962	4	4.2	5.5	4673.20	5.124	4.8
1963	5	4.2	5.7	4843.19	5.340	4.2
1964	6	3.8	5.2	5190.87	5.621	4.2
1965	7	3.5	4.5	5594.37	5.778	3.8
1966	8	2.9	3.8	5990.86	6.030	3.5
1967	9	2.8	3.8	6537.74	6.292	2.9
1968	10	2.9	3.6	7149.12	6.766	2.8
1969	11	2.7	3.5	7819.75	6.972	2.9
1970	12	3.9	4.9	8565.21	7.228	2.7
1971	13	3.7	5.9	9157.91	7.453	3.9
1972	14	3.9	5.6	10024.00	7.859	3.7
1973	15	3.0	4.9	11541.70	8.345	3.9
1974	16	4.3	5.6	12947.50	9.198	3.0
1975	17	7.1	8.5	14394.30	9.965	4.3
1976	18	5.6	7.7	16075.60	11.023	7.1
1977	19	5.0	7.0	18073.70	11.983	5.6
1978	20	3.9	6.0	20841.40	12.944	5.0
1979	21	3.4	5.8	24179.40	14.384	3.9
1980	22	4.8	7.1	27906.80	16.428	3.4
1981	23	3.6	7.6	32919.20	18.192	4.8
1982	24	5.7	9.7	36119.40	19.492	3.6

- i. $E(\varepsilon_t) = 0$ for all t .
- ii. The ε_t are stochastically independent of the $x1_t$, $x2_t$, and $x3_t$.
- iii. The components of the vector $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{24})$ are independent and identically distributed.

To get started on the bootstrap method, let $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ be the least squares estimates reported in Table X.

Consider the residual errors,

$$\hat{\varepsilon}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x1_t - \hat{\beta}_2 y_{t-1} - \hat{\beta}_3 x2_t - \hat{\beta}_4 x3_t.$$

Let $\hat{\varepsilon}_t$ be the 24-vector $(\hat{\varepsilon}_{59}, \hat{\varepsilon}_{60}, \dots, \hat{\varepsilon}_{82})$ of residual errors. Now simulate the unemployment rate equation putting mass $\frac{1}{24}$ on each of the $\{\hat{\varepsilon}_t: t = 1959, 1960, \dots, 1982\}$. Note that the fitting period runs from 1959 to 1982: a year of data is lost due to the lag term.

- i. y_{1958} is held fixed.
- ii. The exogenous variables x_1 , x_2 , and x_3 are held fixed.
- iii. The parameters are set at their least squares estimates, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_4$.
- iv. The disturbance terms are independent with common distribution \hat{F}_n .

More specifically, let $\{\varepsilon_t^*: t = 1959, 1960, \dots, 1982\}$ be the results of 24 independent draws made at random from the 24-vector $\{\hat{\varepsilon}_t: t = 1959, 1960, \dots, 1982\}$. The bootstrap procedure is similar to the one we used in Chapter IV. The bootstrap pseudo-random samples can be collected iteratively:

$$y_t^* = \hat{\beta}_0 + \hat{\beta}_1 x1_t + \hat{\beta}_2 y_{t-1}^* + \hat{\beta}_3 x2_t + \hat{\beta}_4 x3_t + \varepsilon_t^*$$

TABLE X

BOOTSTRAP RESULT FOR THE UNEMPLOYMENT RATE EQUATION
(THERE ARE 200 BOOTSTRAP REPLICATIONS. ESTIMATION
IS BY LEAST SQUARES, USING THE FIRST SEED.)

Col. Parm.	Least Squares Estimates		Bootstrap		
	(1) Estimate	(2) Conv. SE	(3) B/S Mean	(4) B/S SD	(5) RMS Conv. SE
β_0	-4.494942	0.891737	-4.55921587	0.78847126	0.788907
β_1	0.969444	0.064317	0.97285149	0.05386287	0.055103
β_2	-0.206437	0.072266	-0.21653335	0.06461229	0.060774
β_3	-0.000742365	0.0001246082	-0.00075394	0.00011077	0.000106
β_4	1.452783	0.262546	1.47777931	0.23299500	0.229194

These are artificial data for the unemployment rate equation. Now then we can obtain the bootstrap parameter estimates $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, $\hat{\beta}_3^*$, and $\hat{\beta}_4^*$ from the above artificial data using the previous least squares estimation procedure. These steps are repeated, independently 200 times (see Figure 4).

Now let us look at Table X. The validity of the conventional standard errors shown in this table is open to serious question, because $\hat{\sigma}^2$ may not be an accurate estimate of σ^2 due to the relatively small sample size and possibly specification errors in the model. Columns (3) and (4) show for each parameter in the original model the sample mean and sample standard deviation for 200 starred estimates. These standard deviations

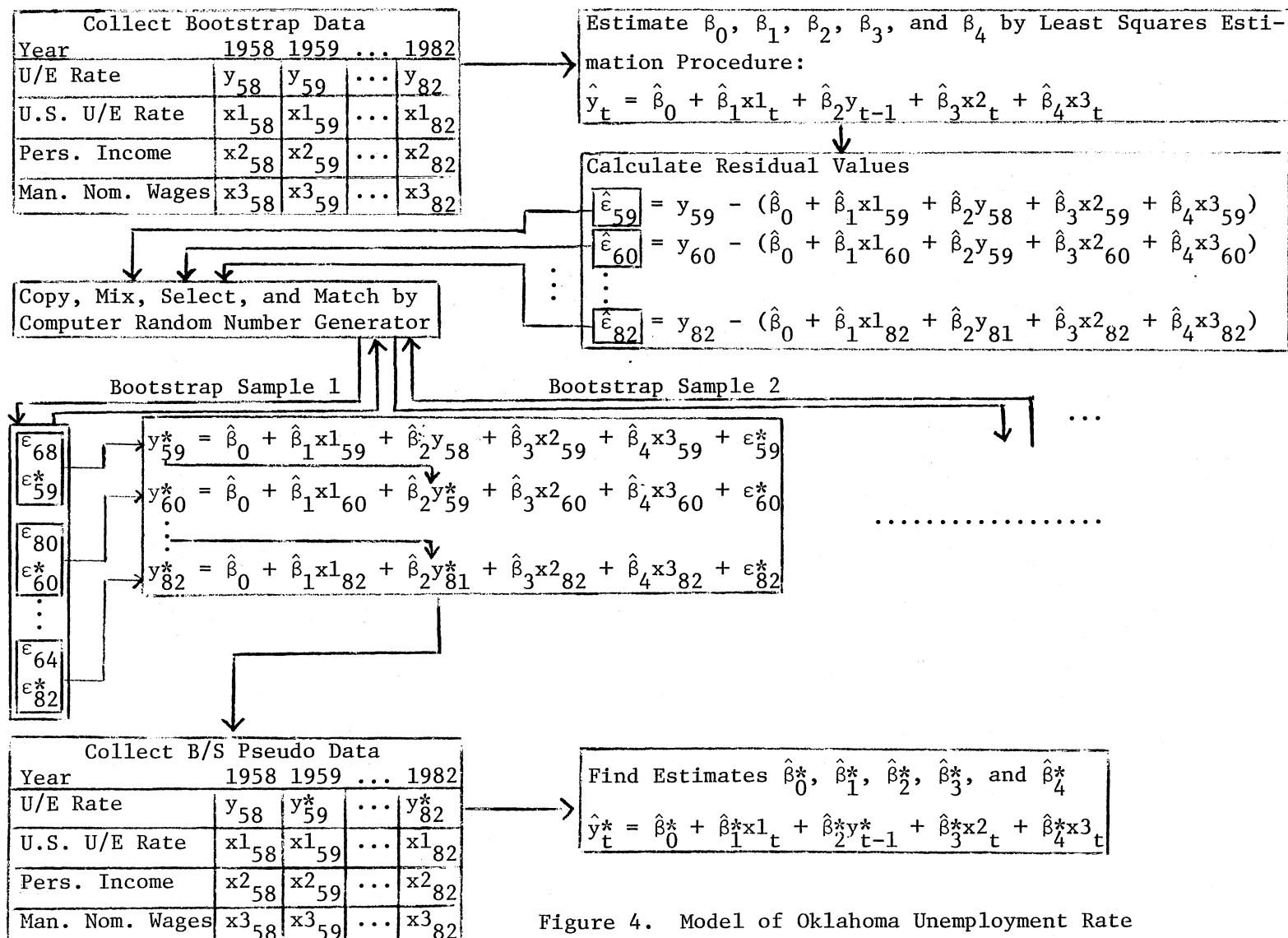


Figure 4. Model of Oklahoma Unemployment Rate

$$\sum_{b=1}^{200} \left[(\hat{\beta}_i^*(b) - \hat{\beta}_i^{*(\cdot)})^2 / (200-1) \right]^{1/2}, \quad i=0,1,2,3,4, \text{ and}$$

$$\hat{\beta}_i^{*(\cdot)} = \sum_{b=1}^{200} \hat{\beta}_i^*(b) / 200$$

are the bootstrap estimates of variability in the parameter estimates.

Comparing columns (1) and (3) does not indicate the presence of any appreciable small sample bias in the least squares coefficient estimates. For example, in the bootstrap simulation world, β_2 was set at the estimated value -0.206437. However, the 200 bootstrap coefficients $\hat{\beta}_i^{*(b)}$, $b=1,2,\dots,200$, has a sample average of -0.21653335. With 200 bootstrap replications the standard error of the mean is

$$0.06461229/\sqrt{200} = 0.0045687788.$$

The null hypothesis here is that the average of the starred estimates is equal to the least squares estimate:

$$H_0: E_{\hat{F}_n}(\hat{\beta}_i^*) - \hat{\beta}_i = 0, \quad i=0,1,2,3,4.$$

So the calculated t-value would be:

$$t = (\hat{\beta}_i^{*(\cdot)} - \hat{\beta}_i) / SE(\hat{\beta}_i^{*(\cdot)})$$

$$= (\hat{\beta}_i^{*(\cdot)} - \hat{\beta}_i) / \{SE(\hat{\beta}_i^*) / \sqrt{200}\}, \quad i=0,1,2,3,4.$$

The t-values for the biases in the least squares estimates are:

- i. For β_0 : $t = \{-4.55921587 - (-4.494942)\} / (0.78847126 / \sqrt{200}) = -1.15,$
- ii. For β_1 : $t = (0.97285149 - 0.969444) / (0.05386287 / \sqrt{200}) = 0.89,$
- iii. For β_2 : $t = \{-0.21653335 - (-0.206437)\} / 0.0045687788 = -2.21,$

iv. For β_3 : $t = \{-0.00075394 - (-0.000742365)\} / (0.00011077 / \sqrt{200}) = -1.48,$

v. For β_4 : $t = (1.47777931 - 1.452783) / (0.23299500 / \sqrt{200}) = 1.52.$

Small sample bias is statistically significant in only one of the least squares estimates. The estimated coefficients $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_3$, and $\hat{\beta}_4$ are not significant. On the whole, the least squares coefficient estimates are performing very well in the Oklahoma unemployment rate equation.

Column (5) in Table X shows the root mean square of the conventional variances computed by applying the formula $\hat{\sigma}^2(X'X)^{-1}$ to the starred data set. For example, in the simulation world of the bootstrap experiment, the real variability of the least squares estimate for β_2 is 0.06461229 from column (4). But the typical variability by the conventional formula is 0.060774. In the bootstrap starred data set, the conventional formula $\hat{\sigma}^2(X'X)^{-1}$ is off by 6.3%. The other RMS Conv. SEs can be understood in a similar way. Some are a bit large, some a bit small compared to the entries in column (4). The other RMS Con. SEs are off by 0.05% \sim 6.3%. On the whole, the conventional least squares formulae seem to perform very well.

For the above bootstrap results, we used a particular seed for the random number generator. We obtained similar bootstrap experiment results from another seed as shown in Table XVII (Appendix C). In column (6) of this table, the t-values are shown for checking the biases in the conventional least squares estimates. None of the estimated coefficients show significant bias.

Bootstrapping Oklahoma Individual

Income Tax Model

The second econometric model to which we applied the bootstrap methodology is the Oklahoma individual income tax equation developed by the same Business and Economic Research Center (see Table XI for the data):

$$y_t = \beta_0 + \beta_1 x1_t + \beta_2 x2_t + \beta_3 D1_t + \beta_4 D2_t + \varepsilon_t, \quad t = 1962, 1963, \dots, 1982,$$

where y_t is the Oklahoma individual income tax (\$1,000,000,

$x1_t$ is the Oklahoma fiscal personal income (\$1,000,000),

$x2_t$ is the value of Oklahoma oil and gas production (\$1,000,000),

$D1_t$ and $D2_t$ are dummy variables.

Dummy variables were used to account for a change in the level of y_t due to changes in the oil industry in Oklahoma during this period.

Note here that we do not have a lag term as an explanatory variable. The bootstrap procedure was repeated 200 times. The bootstrap results are shown in Table XII. We believe that column (4) and (5) show that column (2) is quite dependable. The conventional standard errors from the simulation world of the bootstrap in column (5) are off 1.3% - 10.3% compared to the real variability in column (4). None of the estimated coefficients from the conventional least squares are biased (see column (6)). On the whole, the conventional least squares coefficient estimates are performing very well.

The bootstrap results from the second seed are shown in Table XVIII (Appendix C). Comparing columns (4) and (5), some entries in column (5)

TABLE XI
DATA FOR THE OKLAHOMA INDIVIDUAL INCOME TAX EQUATION

Year	t	y_t	$x1_t$	$x2_t$	$D1_t$	$D2_t$
1962	1	26.025	4577.3	653.54	0	0
1963	2	17.974	4758.2	669.55	0	0
1964	3	21.652	5017.0	714.81	0	0
1965	4	26.739	5392.6	726.21	0	0
1966	5	32.293	5792.6	753.51	0	0
1967	6	32.403	6264.3	860.54	0	0
1968	7	40.917	6843.4	878.02	0	0
1969	8	48.253	7479.6	902.83	0	0
1970	9	50.594	8192.5	957.53	0	0
1971	10	59.531	8861.5	971.04	0	0
1972	11	91.621	9590.9	1006.62	1	0
1973	12	104.721	10782.8	984.16	1	0
1974	13	120.743	12244.6	1391.31	0	0
1975	14	151.723	13670.9	1815.69	0	0
1976	15	180.294	15234.9	2143.90	0	0
1977	16	206.541	17074.6	2677.70	0	0
1978	17	255.342	19457.5	3085.08	0	0
1979	18	318.726	22510.4	3450.14	0	0
1980	19	365.342	26043.1	5732.41	0	1
1981	20	483.365	30413.0	8000.14	0	1
1982	21	617.187	34519.3	10241.00	0	1

TABLE XII

BOOTSTRAP RESULT FOR THE INDIVIDUAL INCOME TAX EQUATION (THERE ARE 200 BOOTSTRAP REPLICATIONS. ESTIMATION IS BY LEAST SQUARES, USING THE FIRST SEED.)

Col. Parm.	Least Square Estimates		Bootstrap			Bias
	(1) Estimate	(2) Conv. SE	(3) B/S Mean	(4) B/S SD	(5) RMS Conv. SE	(6) t
β_0	-60.424068	4.184160	-60.41803617	3.96298057	3.701018826	0.02
β_1	0.010569	0.0007081285	0.01056024	0.00067256	0.0006248927	-0.18
β_2	0.036638	0.003396381	0.03665554	0.00304411	0.0030037766	0.08
β_3	14.463899	5.887318	14.52648928	5.06593522	5.20751457	0.17
β_4	-64.224287	12.716744	-64.0514208	10.0859674	11.24835258	0.24

are slightly higher than their counterparts in column (4), some are slightly lower. All the interpretations are similar.

CHAPTER VIII

SUMMARY AND FURTHER CONSIDERATIONS

The bootstrap is a methodology for estimating standard errors. The idea is to use a Monte Carlo simulation experiment based on a nonparametric estimate of the error distribution. The main objective of this dissertation was to demonstrate the use of the bootstrap to attach standard errors to coefficient estimates and multi-period forecasts in a second-order autoregressive model fitted by least squares and maximum likelihood estimation. A secondary objective of this article was to present the bootstrap in the context of two econometric equations describing the unemployment rate and individual income tax in the state of Oklahoma.

As it turns out, the conventional asymptotic formulae (both the least squares and maximum likelihood estimates) for estimating standard errors appear to overestimate the true standard errors. But there are two problems:

- i. The first two observations y_1 and y_2 have been fixed, and
- ii. The residuals have not been inflated.

After these two factors are considered in the trial and bootstrap experiment, both the conventional maximum likelihood and bootstrap estimates of the standard errors appear to be performing quite well. Now we need to ask in what situations the bootstrap will work most of the time, and how much we can generalize it. At present, there

does not seem to be a good rule of thumb for deciding when the conventional asymptotic formulae will give acceptable results. Developing such a rule might be a good focus for future research. In the present example, the conventional asymptotic methods for estimating standard errors of multi-period forecast seem satisfactory. The bootstrap forecasts through both least squares and maximum likelihood estimates show small biases and small random errors. The application of the bootstrap method to the Oklahoma econometric equations reveals that the conventional estimates of standard error for least squares coefficients perform very well.

Besides what we have done so far, the following developments would be good projects for further study.

- i. Perform the bootstrap for various ARMA type models (AR(1), MA(1), MA(2), and ARMA(1,1), etc.) to check the accuracy of the asymptotics.
- ii. Perform the bootstrap at different sample sizes to see if the results derived from various sample sizes are consistent or not.
- iii. The bootstrap reported here uses error vectors whose distribution is discrete. Perform the bootstrap in which the error vectors are generated as pseudo-normal vectors.
- iv. Provide some improvements for the bootstrap standard errors.
- v. Do a similar experiment to assess the quality of the bootstrap using least squares estimation.
- vi. Change the number of bootstrap replications and see what happens.
- vii. Develop multi-period forecasting errors via the bootstrap for econometric equations.

viii. Compare the performance of conventional asymptotic estimates of standard error to the performance of a bootstrap procedure in various types of time series models when the parameters are near the stationarity and invertibility boundaries.

A SELECTED BIBLIOGRAPHY

- Adams, F. G., Brooking, C. G., and Glickman, N. J. 1975. "On the Specification and Simulation of a Regional Econometric Model: A Model of Mississippi", Review of Economics and Statistics, 57, pp. 286-298.
- Ansley, C. F. 1979. "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process," Biometrika, 66, 1, pp. 59-65.
- Ansley, C. F. and Newbold, P. 1980. "Finite Sample Properties of Estimators for Autoregressive Moving Average Models", Journal of Econometrics, 13, pp. 159-183.
- Beran, R. 1982. "Estimated Sampling Distributions: The Bootstrap and Competitors," The Annals of Statistics, 10, 1, pp. 212-225.
- Beran, R. 1984. "Bootstrap Methods in Statistics," Jahresbericht der Deutschen Mathematiker-Vereinigung, 86, pp. 14-30.
- Bickel, P. J. and Freedman, D. A. 1983. "Bootstrapping Regression Models with Many Parameters," A Festschrift for Erich Lehman, pp. 22-48, p. Bickel, K. Doksum, and J. L. Hodges editors. Belmont, California (Wadsworth).
- Bickel, P. J. and Freedman, D. A. 1981, "Some Asymptotic Theory for the Bootstrap," Annals of Statistics, 9, 6, pp. 1196-1217.
- Bowerman, B. L. and O'Connell, R. T. 1979. Forecasting and Time Series, California: Duxbury Press.
- Box, G. E. P. and Jenkins, G. M. 1976. Time Series Analysis Forecasting and Control, San Francisco: Holden-Day, Inc.
- Chang, S. 1979. "An Econometric Forecasting Model Based on Regional Economic Information System Data: the Case of Mobile, Alabama," Journal of Regional Science, 19, 4, pp. 437-447.
- Copas, J. B. 1966. "Monte Carlo Results for Estimation in a Stable Markov Time Series," Journal of the Royal Statistical Society, Series A, pp. 110-116.
- Dent, W. and Min, A. S. 1978. "A Monte Carlo Study of Autoregressive Integrated Moving Average Processes," J. Econometrics, 7, pp. 23-55.

- Diaconis, P. and Efron, B. 1983. "Computer-Intensive Methods in Statistics," Scientific American, May, pp. 116-130.
- Duobinis, S. F. 1981. "An Econometric Model of the Chicago Standard Metropolitan Statistical Area," Journal of Regional Science, 21, 3, pp. 293-319.
- Efron, B. 1979a. "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7, pp. 1-26.
- Efron, B. 1979b. "Computers and the Theory of Statistics: Thinking and Unthinkable," SIAM Review, 21, pp. 460-480.
- Efron, B. 1981a. "Censored Data and the Bootstrap," Journal of the American Statistical Association, 76, pp. 312-319.
- Efron, B. 1981b. "Nonparametric Standard Errors and Confidence Intervals," Canadian Journal of Statistics, 9, pp. 139-192.
- Efron, B. 1982. "The Jackknife, the Bootstrap, and Other Resampling Plans," Society for Industrial and Applied Mathematics, Monograph #38, The Conference Board of the Mathematical Sciences, supported by the Nat. SC. Found.
- Efron, B. and Gong, G. 1983. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," The American Statistician, Feb., 37, 1, pp. 36-47.
- Fair, R. C. 1979. "An Analysis of the Accuracy of Four Macroeconomic Models," Journal of Political Economy, 87, 41, pp. 701-718.
- Fisher, R. A. 1915. "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," Biometrika, 10, pp. 507-521.
- Fishman, G. S. and Moore, L. R. 1982. "A Statistical Evaluation of Multiplicative Congruential Random Number Generators With Modulus $2^{31}-1$," Journal of the American Statistical Association, 77, 377, pp. 129-136.
- Freedman, D. A. 1981. "Bootstrapping Regression Models," The Annals of Statistics, 9, 6, pp. 1218-1228.
- Freedman, D. A. and Peters, S.C. 1983a. "Bootstrapping a Regression Equation: Some Empirical Results," Technical Report, Department of Statistics, University of California, Berkeley.
- Freedman, D. A. and Peters, S. C. 1983b. "Using the Bootstrap to Evaluate Forecasting Equations," Technical Report, Department of Statistics, University of California, Berkeley.

- Freedman, D. A. and Peters, S. C. 1984. "Bootstrapping an Econometric Model: Some Empirical Results," Journal of Business and Economic Statistics, 2, 2, pp. 150-158.
- Gong, G. 1982. "Cross-Validation, the Jackknife, and the Bootstrap: Excess Error Estimation in Forward Logistic Regression," Technical Report, Division of Biostatistics, Stanford University, Stanford, California.
- Granger, C. W. J. and Newbold, P. 1977. Forecasting Economic Time Series, New York: Academic Press.
- Graybill, F. A. 1976. Theory and Application of the Linear Model, Massachusetts: Duxbury Press.
- Hurwicz, L. 1950. "Least Squares Bias in Time Series," In T. Koopmans, ed., Statistical Inference in Dynamic Economic Models, pp. 365-383. New York: John Wiley & Sons., Inc.
- Ireland, T. C. 1978. "An Econometric Model for the State of Oklahoma," (Ph.D. dissertation, Oklahoma State University).
- Johnston, J. 1972. Econometric Methods, New York: McGraw-Hill.
- Latham, W. R. 1979. "Regional Econometric Models' Specification and Simulation of a Quarterly Alternative for Small Regions," Journal of Regional Science, 19, 1, pp. 1-13.
- Lewis, P. A. W., Goodman, A. S., and Miller, J. M. 1969. "A Pseudo-random Number Generator for the System/360," IBM Systems Journal, 8, pp. 136-146.
- Liew, C. K. 1974. "A Bridge Model for Oklahoma," Oklahoma Business Bulletin, 42, 1, pp. 16-17.
- Liew, C. K. and Kahng, D. K. 1971. "The Oklahoma Econometric Model I," Oklahoma Business Bulletin, 39, 7, pp. 7-24.
- Marquardt, D. W. 1963. "An Algorithm for Least Squares Estimation of Nonlinear Parameters," Journal of the Society of Industrial and Applied Mathematics, 11, 2(June), pp. 431-441.
- Mikhail, W. M. 1975. "A Comparative Monte Carlo Study of the Properties of Econometric Estimators," Journal of the American Statistical Association, March, 70, 349, pp. 94-104.
- Miller, R. G. 1974. "The Jackknife--a Review," Biometrika, 61, 1, pp. 1-15.
- Nelson, C. R. 1973. Applied Time Series Analysis, San Francisco: Holden-Day Inc.
- Office of Business and Economic Research, College of Business Administration, Oklahoma State University 1984, Oklahoma Economic Outlook.

- Quenouille, M. 1949. "Approximate Tests of Correlation in Time Series," J. Roy. Statist. Soc. Ser. B, 11, pp. 18-84.
- Rao, C. R. 1973. Linear Statistical Inference and Its Applications, New York: John Wiley & Sons.
- Salem, S. A., and Kline, S. 1978. "Estimation in the Second Order Autoregressive Models," American Statistical Association, Proceedings of Statistical Education Section, pp. 652-654.
- Salem, S. A., Kline, S. and French Laboratories. 1978. "Estimation in the Second Order Autoregressive Models," American Statistical Association, Proceedings of Business and Economic Statistics Section, pp. 652-654.
- Singh, K. 1981. "On the Asymptotic Accuracy of Efron's Bootstrap," The Annals of Statistics, 9, 6, pp. 1187-1195.
- Stephens, M. A. 1974. "EDF Statistics for Goodness of Fit and Some Comparisons," Journal of the American Statistical Association, 69, 347, pp. 730-737.
- Theil, H. 1978. Introduction to Econometrics, N.J.: Prentice-Hall.
- Theil, H., Rosalsky, M. C., and Finke, R. 1983. "A Comparison of Normal and Discrete Bootstraps for Standard Errors in Equation Systems," Technical Report, University of Florida Graduate School of Business Administration.
- Tukey, J. 1958. "Bias and Confidence in not quite Large Samples," abstract, Ann. Math. Statist., 29, p. 614.

APPENDIXES

APPENDIX A

COMPUTATIONAL DETAILS

All computations reported here were performed on the IBM 3081 computer at Oklahoma State University. The bootstrap experiments were conducted using the ARIMA procedure of SAS/ETS, the SAS Econometrics and Time Series Library. The ARIMA procedure analyzes and forecasts time series data.

Uniformly distributed pseudo-random numbers were obtained from one of the SAS functions, RANUNI. This random number generator returns numbers generated from the uniform distribution on the interval (0,1) using a prime modulus multiplicative generator. This generator has period $2^{31} - 1 = 2.1 \times 10^9$ and multiplier 397204094. See Fishman and Moore (1982) for more details. The technique requires the initialization of a random number stream with seeds. The first seed (7035209) was used for the bootstrap experiments. The results from the first seed were replicated in Appendix C using the second seed (4659537). Normally distributed pseudo-random numbers were obtained from RANNOR SAS function, which generates normal variates with assigned mean and variance.

APPENDIX B

A SIMPLE MATHEMATICAL EXPLANATION

In a simpler context, the bias in the least squares estimator will be shown in this Appendix.

Let us consider the simplest autoregressive model, i.e., the first-order autoregressive model:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \quad t=1,2,\dots,n \quad (\text{B-1})$$

where each ε_t is assumed to be generated by a white noise process, so that $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, and $E(\varepsilon_t \varepsilon_{t-s}) = 0$ for $s \neq 0$. The current value of the time series, y_t , is expressed as a linear function of the previous value of the series and a random shock ε_t . Let z_t and z_{t-1} be deviations from the means of the time series: $z_t = y_t - \bar{y}_t$, and $z_{t-1} = y_{t-1} - \bar{y}_{t-1}$. Then, by the least squares estimation, we obtain for the estimate of ϕ_1 :

$$\hat{\phi}_1 = \frac{\sum_{t=2}^n z_t z_{t-1}}{\sum_{t=2}^n z_{t-1}^2} \quad (\text{B-2})$$

This can be written as:

$$\hat{\phi}_1 = \frac{\sum_{t=2}^n u_t z_t}{\sum_{t=2}^n z_t^2} \quad (\text{B-3})$$

where

$$u_t = \frac{z_{t-1}}{\sum_{t=2}^n z_{t-1}^2}.$$

Substituting into equation (B-1),

$$\begin{aligned}\hat{\phi}_1 &= \sum_{t=2}^n u_t (\delta + \phi_1 y_{t-1} + \varepsilon_t) \\ &= \delta \sum u_t + \phi_1 \sum u_t y_{t-1} + \sum u_t \varepsilon_t.\end{aligned}$$

Since $\sum u_t = 0$ and $\sum u_t y_{t-1} = 1$,

$$\hat{\phi}_1 = \phi_1 + \sum u_t \varepsilon_t \quad (\text{B-4})$$

Therefore, the expected value of this equation is:

$$E(\hat{\phi}_1) = \phi_1 + E(\sum u_t \varepsilon_t) \quad (\text{B-5})$$

Note that, following the definition of the weights u_t , we have:

$$\sum u_t \varepsilon_t = \frac{\sum z_{t-1} \varepsilon_t}{\sum z_{t-1}^2}.$$

In the usual regression model the regressor is assumed to be fixed in repeated sampling, however, in our autoregressive model, the regressor y_{t-1} cannot be considered a fixed variable. According to our model specification, y_t is dependent on y_{t-1} and ε_t . The variable z_{t-1} includes y_t , and y_t is not independent of ε_t . Hence,

$$E(\sum u_t \varepsilon_t) = E\left\{ \frac{\sum z_{t-1} \varepsilon_t}{\sum z_{t-1}^2} \right\} \neq \frac{\sum z_{t-1}}{\sum z_{y-1}^2} E(\varepsilon_t).$$

Therefore,

$$E(\sum u_t \varepsilon_t) \neq 0.$$

Then, from (B-5),

$$E(\hat{\phi}_1) = \phi_1 + E(\sum u_t \varepsilon_t) \neq \phi_1,$$

i.e., the least squares estimator is biased.

APPENDIX C

TABLES

TABLE XIII

BOOTSTRAP RESULT FOR THE AR(2) MODEL (THERE ARE 100
BOOTSTRAP PSEUDO-RANDOM SAMPLES. ESTIMATION
IS BY LEAST SQUARES, USING THE SECOND SEED.)

Least Squares Estimates			Bootstrap			
Col.	(1)	(2)	(3)	(4)	(5)	(6)
Parm.	Estimate	Conv. SE	B/S Mean	B/S SD RMS	Conv. SE	t
μ	34.9039	0.2978	34.9369	0.2749	0.3077	1.20
ϕ_1	0.613551	0.0971	0.646206	0.0836	0.0958	3.91
ϕ_2	-0.383048	0.0975	-0.400100	0.0786	0.0961	-2.23

TABLE XIV

BOOTSTRAP RESULT FOR THE AR(2) MODEL (THERE ARE 100
BOOKSTRAP PSEUDO-RANDOM SAMPLES. ESTIMATION
IS BY MAXIMUM LIKELIHOOD, USING THE
SECOND SEED.)

Maximum Likelihood Estimates			Bootstrap			
Col.	(1)	(2)	(3)	(4)	(5)	(6)
Parm.	Estimate	Conv. SE	B/S Mean	B/S SD RMS	Conv. SE	t
μ	34.9461	0.2962	34.9994	0.2804	0.3107	1.90
ϕ_1	0.682098	0.0983	0.776515	0.0906	0.0937	10.42
ϕ_2	-0.432882	0.0944	-0.500111	0.0904	0.0898	-7.44

TABLE XV

BOOTSTRAP FORECAST EXPERIMENT FOR MODEL (V-7) (THERE ARE 100 BOOTSTRAP
REPLICATIONS. ESTIMATION IS BY LEAST SQUARES,
USING THE SECOND SEED.)

	(1)	Least Squares Estimates		Bootstrap		
		(2)	(3)	(4)	(5)	(6)
t=n+h n=85, h=1,2,...,12	Actuals y_t , t=86,87,...,97	Forecasts	Standard Error of Forecasts	Sample Mean Actuals	Sample Mean Forecasts	SD of Forecast Error
86	33.9252	33.9950	2.2189	34.8583	34.7257	1.9829
87	36.1036	34.9416	2.6417	35.0249	34.8738	2.1117
88	36.7351	35.2622	2.6417	35.3198	34.9898	2.1796
89	35.4576	35.0786	2.7057	35.4195	35.0803	1.8236
90	37.5924	34.8278	2.7325	35.1118	35.0358	1.7142
91	34.4895	34.7414	2.7325	34.8997	34.7864	1.9482
92	39.1692	34.7892	2.7369	35.1730	34.7730	2.0222
93	35.8242	34.8557	2.7388	34.8394	35.0531	1.8245
94	32.3875	34.8789	2.7388	34.7769	34.6927	1.9007
95	31.2846	34.8665	2.7391	34.9128	34.7956	1.9106
96	33.5516	34.8489	2.7392	35.0979	34.9185	1.9925
97	35.6008	34.8426	2.7392	35.1995	34.9514	1.8227

TABLE XVI

BOOTSTRAP FORECAST EXPERIMENT FOR MODEL (V-7) (THERE ARE 100 BOOTSTRAP
 REPLICATIONS. ESTIMATION IS BY MAXIMUM LIKELIHOOD,
 USING THE SECOND SEED.)

t=n+h n=85, h=1,2,...,12	(1) Actuals y_t , t=86,87,...,97	Least Squares Estimates		Bootstrap		
		(2) Forecasts	(3) Standard Error of Forecasts	(4) Sample Mean Actuals	(5) Sample Mean Forecasts	(6) SD of Forecast Error
86	33.9252	33.9342	2.1342	35.0772	34.9745	2.1290
87	36.1036	34.9657	2.6363	35.0057	34.8882	2.0267
88	36.7351	35.4082	2.6387	35.0701	34.9441	1.7427
89	35.4576	35.2399	2.7184	35.0326	35.0461	1.9513
90	37.5924	34.9081	2.7690	35.2209	34.9439	1.8716
91	34.4895	34.7473	2.7699	35.2350	35.0616	1.8349
92	39.1692	34.7880	2.7769	35.2788	34.9316	2.0138
93	35.8242	34.8938	2.7826	35.2775	34.9721	2.1799
94	32.3875	34.9512	2.7828	35.2995	34.9729	1.7158
95	31.2846	34.9426	2.7834	35.1927	35.0207	1.9139
96	33.5576	34.9092	2.7840	35.1776	34.9360	2.0842
97	35.6008	34.8891	2.7841	35.2000	34.9721	1.9054

TABLE XVII

BOOTSTRAP RESULT FOR THE UNEMPLOYMENT RATE EQUATION (THERE ARE 200 BOOTSTRAP REPLICATIONS. ESTIMATION IS BY LEAST SQUARES, USING THE SECOND SEED.)

Least Squares Estimates			Bootstrap			Bias
Col.	(1)	(2)	(3)	(4)	(5)	(6)
Parm. \	Estimate	Conv. SE	B/S Mean	B/S SD	RMS Conv. SE	t
β_0	-4.494942	0.891737	-4.53387683	0.86055392	0.806478	-0.64
β_1	-0.969444	0.064317	0.97573241	0.06137876	0.056459	1.45
β_2	-0.206437	0.072266	-0.21198438	0.06557038	0.062019	-1.20
β_3	-0.000742365	0.0001246082	-0.00074636	0.00011299	0.000109	0.50
β_4	1.452783	0.262546	1.46017850	0.23796198	0.234256	0.44

TABLE XVIII

BOOTSTRAP RESULT FOR THE INDIVIDUAL INCOME TAX EQUATION (THERE ARE 200 BOOTSTRAP REPLICATIONS. ESTIMATION IS BY LEAST SQUARES, USING THE SECOND SEED.)

Least Squares Estimates			Bootstrap			Bias
Col.	(1)	(2)	(3)	(4)	(5)	(6)
Parm. \	Estimate	Conv. SE	B/S Mean	B/S SD	RMS Conv. SE	t
β_0	-60.424068	4.184160	-60.54791842	3.69964531	3.57953176	-0.47
β_1	0.010569	0.0007081285	0.01056471	0.00059428	0.000604191	-0.10
β_2	0.036638	0.003396381	0.03668221	0.00272731	0.00290643	0.23
β_3	14.463899	5.887318	14.90243472	4.76189917	5.036576164	1.30
β_4	-64.224287	12.716744	-63.95681600	10.63527216	10.87912178	0.36

VITA ✓

Mun Shig Son

Candidate for the Degree of
Doctor of Philosophy

Thesis: BOOTSTRAPPING A TIME SERIES MODEL

Major Field: Statistics

Biographical:

Personal Data: Born in Hwanggan, Yeongdong, Choongbuk, Korea,
February 5, 1950, the son of Young Hee Son and Jung Poon Park.

Education: Graduated from Dae Jun High School, Dae Jun, Korea, in
February, 1969; received Bachelor of Economics Degree from
Sung Kyun Kwan University, Seoul, Korea, in February, 1975;
received Master of Science Degree in Statistics from Oklahoma
State University in December, 1982; received Master of Science
Degree in Economics from Oklahoma State University in May,
1984; completed requirements for the Doctor of Philosophy
Degree at Oklahoma State University in July, 1984.

Professional Experience: Economic Research Assistant, the Central
Bank of Korea, Seoul, Korea, 1975-1978; Graduate Teaching
Associate, Statistics and Mathematics Departments, Oklahoma
State University, 1980-1984.

Professional Organizations: American Statistical Association,
Korean Scientists and Engineers Association in America, Mu
Sigma Rho National Statistical Honor Fraternity.