BOOTSTRAP TECHNOLOGY AND APPLICATIONS

by

Christian Léger and Dimitris N. Politis
Université de Montréal Purdue University

Joseph P. Romano
Stanford University

Technical Report #91-54

Department of Statistics
Purdue University

September 1991

◇

Bootstrap Technology and Applications

Christian Léger
Département d'informatique et Recherche Opérationnelle
Université de Montréal

Dimitris N. Politis
Department of Statistics
Purdue University

Joseph P. Romano
Department of Statistics, Stanford University

Abstract. Bootstrap resampling methods have emerged as powerful tools for constructing inferential procedures in modern statistical data analysis. While these methods depend upon the availability of fast, inexpensive computing, they offer the potential for highly accurate methods of inference. Moreover, they can even eliminate the need to impose a convenient statistical model that does not have a strong scientific basis. In this article, we review some bootstrap methods, emphasizing applications through illustrations with some real data. Special attention is given to regression, problems with dependent data, and choosing tuning parameters for optimal performance.

Key Words: Bootstrap; Confidence limits; Prediction; Regression; Resampling; Time Series; Tuning Parameters.

1. Introduction

The past decade or so has been revolutionary for the incorporation of computers into statistical practice. Bootstrap resampling methods have emerged as powerful tools for constructing inferential procedures in modern statistical data analysis. The inferential component of a statistical analysis usually consists of constructing a confidence region, attaching a standard error to an estimate, carrying out a test of a hypothesis, constructing a prediction region, or selecting a regression equation. Often, it is implicitly required to approximate or estimate the sampling distribution of some statistic. The bootstrap approach, as initiated by Efron (1979), avoids having to derive formulas via difficult analytical arguments by taking advantage of fast computers. Computer intensive methods have and will continue to have a profound influence throughout science, as the availability of fast, inexpensive computing has enhanced our abilities to make valid statistical inferences about the world without the need for using unrealistic or unverifiable assumptions. Some of the power and limitations of bootstrap methods have become mathematically and practically apparent in some situations. However, the full potential of computer-intensive methods as applied to complex inferential problems has yet to be realized.

The bootstrap should not be viewed as a replacement for mathematics, for only with a sound theoretical foundation can resampling methods be applied safely in practice. Much theoretical work has been devoted to justifying the mathematical validity of the bootstrap, though much more remains. On the applied side, resampling methods are becoming increasingly popular. Indeed, the bootstrap offers the potential for highly accurate methods of inference, and can even eliminate the need to assume or impose a convenient statistical model that does not have a strong scientific basis. At the same time, the basic idea motivating the bootstrap approach is conceptually simple, resampling methods are applicable quite generally, and their implementation is usually automatic (though recent perturbations of basic bootstrap techniques have been suggested that may not be as automatic as we would like).

The goal of this paper is to acquaint the statistical practitioner with the bootstrap as a means toward solving complex inferential problems. It is impossible to discuss all bootstrap methods and their potential applications. However, we intend to provide a basic expository account of the fundamental principles behind bootstrap resampling, while addressing some current issues arising in the application of bootstrap methods. An excellent introduction to the bootstrap may be found in Efron and Tibshirani (1986). More recent and more technical expository accounts may be found in Hinkley (1988) and DiCiccio and Romano (1988). Our intention here is to provide an up-to-date discussion of bootstrap technology emphasizing modern issues and applications, such as variable selection, problems with dependent data, and determining an optimal replacement policy in a reliability study.

In Section 2, the bootstrap reampling approach is presented. A motivating example, involving constructing a confidence interval for a reliability coefficient, is presented. The use of the bootstrap to construct prediction regions is also discussed. Section 3 discusses the use of the bootstrap in regression analysis, with emphasis on the use of robust procedures, choice of regression model, and selection of variables. In Section 4, some resampling schemes are presented to deal with dependent data. The emphasis is on the construction of valid inferential procedures for stationary time series

data, and some illustrations with real data are given. Section 5 goes beyond the traditional application of applying the bootstrap to construct confidence regions or estimating a standard error of an estimator by actually using the bootstrap itself to construct an "optimal" estimator. Examples provided include: choosing an optimal trimming proportion to estimate location, combining estimators in a calibration problem, and determining a nonparametric age replacement policy in a reliability study. Other applications to nonparametric regression and density estimation are discussed as well.

2. The Bootstrap Paradigm

2.1 The Bootstrap Approximation

The heuristics motivating the bootstrap are the following. Data is drawn from an unknown stochastic model. A plausible probability model is fitted to the data and one proceeds as if the fitted model were true. More specifically, data x is assumed to be generated by a probability mechanism P. Interest lies in some quantity (assumed real-valued) $R(x, P)$ depending on the data x and possibly on P as well.

For example, interest might focus on some parameter $\theta(P)$. An estimator $\hat{\theta}$ is suggested, and an estimated standard error is desired, or possibly an estimate of bias of $\hat{\theta}$ as an estimate of $\theta(P)$. A more ambitious goal is to construct a confidence region for $\theta(P)$. Clearly, some knowledge of the distribution of $R(x, P) = \hat{\theta} - \theta(P)$ is desired.

In general, the sample distribution of $R(x, P)$ under P is required. Based on the data x, P is estimated by some probability mechanism $\hat{P}$. The bootstrap approximation to the distribution of $R(x, P)$ under P is the distribution (conditional on x) of $R(x^*, \hat{P})$ when x^* is generated from $\hat{P}$. Define $J(P)$ to be the distribution of $R(x, P)$ under P. Then, in a compact functional notation, $J(P)$ is estimated by $J(\hat{P})$. Moreover, the bootstrap estimates of the quantiles, variance, and mean of $J(P)$ are the corresponding quantiles, variance, and mean of $J(\hat{P})$.

Usually, a stochastic approximation to $J(\hat{P})$ is necessary, and can be simulated by the following Monte Carlo "resampling" algorithm. Conditional on the observed data x, let $x_1^*, \ldots, x_B^*$ be i.i.d. samples according to the estimated model $\hat{P}$. For each simulated data set x_i^*, calculate $R(x_i^*, \theta(\hat{P}))$. The empirical distribution of these B values serves as an approximation to the bootstrap distribution $J(\hat{P})$.

2.2. An Application

To fix ideas, consider the following application, where it is desired to construct a confidence bound for a reliability coefficient. An i.i.d. sample $X_1, \ldots, X_m$ is generated from an unknown distribution F and an independent i.i.d. sample $Y_1, \ldots, Y_n$ is assumed to be generated from an unknown distribution function G. Here, $x = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$. The parameter of interest is $\theta(F, G) = P_{F,G}(X > Y)$. For example, Y might represent the strength of some structural component which is produced in mass, i.e. the stress at which the component will fail. The component is installed at an assembly line and exposed to a stress X. The event $(X > Y)$ represents failure of the component, and $1 - \theta(F, G)$ is the reliability coefficient. Let $\hat{F}_m$ be the empirical distribution of the X_i and let $\hat{G}_n$ be the empirical distribution of the Y_j. The usual Mann-Whitney statistic $\hat{\theta}$ to be used, as suggested by Birnbaum (1956), is equal to $\theta(\hat{F}_m, \hat{G}_n)$ and is the proportion of pairs (X_i, Y_j) such that $X_i > Y_j$. It is desired to find an upper confidence bound for $\theta(F, G)$ by approximating the distribution of $\hat{\theta}$, or by approximating the distribution of $R(x, \theta(F, G)) = \theta(\hat{F}_m, \hat{G}_n) - \theta(F, G)$. In this setting, the model P is completely determined by F and G; a bootstrap approximation consists in replacing F and G by $\hat{F}_m$ and $\hat{G}_n$, respectively. In other works, a resampled data x_i^* is obtained by taking a sample of size m from $\hat{F}$ and size n from $\hat{G}_n$.

We will consider the following data set, taken from Basu (1981). Here, $m = n = 15$; the X's are 0.0352, 0.0397, 0.0677, 0.0233, 0.0873, 0.1156, 0.0286, 0.0200, 0.0797, 0.0072, 0.0245, 0.0251, 0.0469, 0.0838, 0.0796; the Y's are 1.7700, 0.9457, 1.8985, 2.6121, 1.0929, 0.0362, 1.0615, 2.3895, 0.0982, 0.7971, 0.8316, 3.2304, 0.4373, 2.5648, 0.6377. For this data, $\theta(\hat{F}_m, \hat{G}_n) = 0.04$. A simple exact formula for the variance of this estimator is:

$$\sigma^2(F,G) = [\theta + (1-m-n)\theta^2 + (m-1)\theta_X + (n-1)\theta_Y]/(mn),$$

where $\theta = \theta(F,G)$, $\theta_X = P_{F,G}(Y_1 < min(X_1, X_2))$, and $\theta_Y = P_{F,G}(X_1 > max(Y_1, Y_2))$. In this case, $\sigma^2(F,G)$ is just the variance of $J(F,G)$, equal to the sampling distribution of $\theta(\hat{F}_m, \hat{G}_n)$ under the true distributions F and G. Hence, a bootstrap estimate of variance of $\theta(\hat{F}_m, \hat{G}_n)$ can be obtained explicitly as $\hat{\sigma}^2 = \sigma^2(\hat{F}_m, \hat{G}_n)$. This amounts to replacing θ by $\theta(\hat{F}_m, \hat{G}_n)$ in (2.1), and replacing θ_X by $\hat{\theta}_X$ given by

$$\hat{\theta}_X = m^{-2}n^{-1}\sum_{i=1}^{m}\sum_{k=1}^{m}\sum_{j=1}^{n} 1(X_i > Y_j, X_k > Y_j).$$

Also, θ_Y is analogously replaced by its empirical estimate. For the data set considered, $\hat{\sigma}^2 = 0.0014$.

In this example, a normal approximation to the distribution of $\hat{\theta}$ may be questioned, particularly because, in reliability problems, $\theta(F,G)$ may be very small, resulting in a skewed distribution for the true sampling distribution of $\hat{\theta}$. In Figure 2.1, the bootstrap approximation to the distribution of $\hat{\theta}$ is displayed. It was constructed by resampling $B = 2000$ data sets and recomputing the Mann-Whitney statistic for each of these simulated data sets. The 0.95 quantile of this bootstrap histogram is 0.111. Hence, Efron's simple percentile method upper confidence bound of nominal level 0.95 is 0.111.

Alternatively, one may approximate the distribution of $\hat{\theta} - \theta$ rather than $\hat{\theta}$ by calculating $\hat{\theta}_i^* - \hat{\theta}$, where $\hat{\theta}_i^*$ is the Mann-Whitney statistic obtained from the ith resampled data set x_i^*. But, the 0.05 quantile of the bootstrap distribution of $\hat{\theta}_i^* - \hat{\theta}$ is simply the 0.05 quantile of the bootstrap histogram in Figure 2.1 minus $\hat{\theta}$, and is given by -0.04. The approximation $P_{F,G}(\hat{\theta} - \theta(F,G) \geq -0.04) \approx .95$ is then "inverted" to yield 0.08 as an approximate upper confidence bound for θ.

Of course, any approximate method based on asymptotics should be questioned when dealing with small sample sizes. Many refinements of bootstrap methods exist. The bootstrap-t method, for example, approximates the distribution of $R(x,P) = [\hat{\theta} - \theta]/\hat{\sigma}$, by its distribution under $\hat{P}$. In the above example, with $B = 2000$, 413 of the simulated data sets resulted in a Mann-Whitney statistic equal to 0; hence, the estimated standard error for these data sets was 0 as well. This results in division by 0 for the studentized statistic, further exemplifying the difficult nature of this particular problem. For some simulation results concerning the performance of bootstrap methods in small samples in the context of the application considered, see Constantine, et. al. (1989).

2.3. General Considerations

Much theoretical work has been done in comparing bootstrap methods with the goal of obtaining higher order accuracy. Traditional asymptotic and simple bootstrap methods used to construct one-sided confidence limits typically have coverage error of order $n^{-1/2}$, where n is the sample size in a one-sample problem. A similar result holds for two-sample problems, when the sizes of the two samples are comparable. Efron (1987) has refined the simple percentile method, and his so-called accelerated bias-corrected percentile method has one-sided coverage error of order n^{-1}. Hall (1988) has advocated the bootstrap-t method. Iterative bootstrap procedures, such as Beran (1987), Loh (1987), Hall and Martin (1988), and Martin (1990) have been proposed which offer the potential for extremely low coverage error. Basically, one starts with any reasonable confidence bound (constructed by a simple large sample theory approach or even the bootstrap itself). One then uses the bootstrap to estimate coverage of the procedure and corrects the procedure in such a way that its estimated coverage is equal to the nominal level. Some of these methods are reviewed in DiCiccio and Romano (1988).

The thrust of the bootstrap approach is to approximate the probability mechanism generating the data P by some estimate $\hat{P}$. Traditionally, in a nonparametric one-sample problem, the estimate $\hat{P}$ is the empirical distribution of the data, and the bootstrap amounts to "resampling" the original data. However, the scope of the bootstrap is much larger. All that is needed is an appropriate means of estimating P. For example, the bootstrap can be applied to parametric models, in which case data is simulated from a distribution in the assumed family. Even in nonparametric problems, choices other than the empirical distribution exist. Indeed, in some problems, it is benefical to smooth the empirical distribution; see Silverman and Young (1987) and Hall, DiCiccio, and Romano (1989). In nonparametric problems with more complicated data structures, the empirical distribution may be totally inappropriate. For example, Efron and Tibshirani (1986) consider a one-sample problem with censored data, and an appropriate estimate of the parent distribution is given by the Kaplan-Meier estimate. In the next sections, appropriate bootstrap algorithms are discussed for regression problems and problems with dependent data.

Bootstrap methods can also be applied to other modes of inference other than confidence regions for a parameter. For example, based on data $X_1, \ldots, X_n$, it is desired to construct a prediction interval for future observations. In particular, approximate a constant q so that the interval $\bar{X} \pm qS$ contains at least k out of m future observations with specified confidence $1 - \alpha$, where $\bar{X}$ is the sample mean and S^2 is the usual unbiased estimate of variance. For applications of this problem, see Odeh (1990), who specifically assumes that the underlying population is Gaussian. The bootstrap approach may be applied here without parametric assumptions. Implicitly, we seek an approximation to the distribution of the kth largest value of $|Y_i - \bar{X}|/S$, where Y_i represents the ith of m future observations. One possible resampling scheme is the following. Let $\hat{F}_n$ be the empirical distribution of the data. Let $X_1^*, \ldots, X_n^*$ be an i.i.d. (bootstrap) sample from $\hat{F}_n$, with mean $\bar{X}_n^*$ and sample standard deviation S^*. Independent of the X_i^*, let $Y_1^*, \ldots, Y_m^*$ be i.i.d. according to $\hat{F}_n$, and let C be the k largest value of $|Y_i^* - \bar{X}^*|/S^*$. Repeat this resampling of $n + m$ observations B times, each time recalculating C to obtain $C_1, \ldots, C_B$. Let $\hat{q}$ be the upper α quantile of the C's; the resulting prediction interval is $\bar{X} \pm \hat{q}S$. The bootstrap avoids the need for

tables for each parent distribution, and each value of n, m and k.

To obtain some insight into how well the bootstrap is performing in finite samples, some simulations were done. The following estimates of coverage at a nominal level of 0.90 were all based on 2000 simulations with $B = 1000$ bootstrap samples and $n = 10$. With normal data: $m = 20$, $k = 10$, the estimated coverage was 0.901; $m = 10$, $k = 6$, the estimated coverage was 0.896; $m = 5$, $k = 5$, the estimated coverage was 0.859. With double exponential data: $m = 20$, $k = 10$, the estimated coverage was 0.925; $m = 10$, $k = 6$, the estimated coverage was 0.897; $m = 5$, $k = 5$, the estimated coverage was 0.809. In short, if k/m is not too close to 1, the actual level of the bootstrap procedure is pretty close to the nominal level. However, if k/m is near 1, the bootstrap does not fare as well, because we are trying to estimate more extreme quantiles. Of course, the parametric methods do not do well then either unless the data is nearly normal. For a discussion of the bootstrap in prediction, see Bai and Olshen (1988) and Beran (1990).

The bootstrap may also be applied to constructing tests of hypotheses. In this case, one may invert a bootstrap confidence region. Alternatively, the distribution of a test statistic may be simulated directly, as done in Beran (1986) and Romano (1988a). In this case, one must approximate the probability model generating the data by a model obeying the constraints specified by the null hypothesis even when the null hypothesis is false, because critical values are estimated under the null hypothesis. In some problems, bootstrap tests are closely related to randomization and permutation tests; see Romano (1989). In the reliability application, exact confidence bounds exist based on inverting permutation tests if it is assumed F and G differ only in location. Without such an assumption, permutations tests and their associated confidence regions are not of the nominal level even in large samples, unlike the bootstrap; see Romano (1990).

3. Bootstrapping Linear Regression Models

3.1. Introduction and Example

Linear regression is one of the tools most often used by statisticians. Despite the simplicity of the model, there remain many complex interesting problems. New estimators have been introduced to compete with the classical least squares estimator. Among them are robust estimators which are not as influenced by outliers. Such estimators are now widely available in computing environments such as the S language. Unfortunately, the lack of good, simple inference tools based on these estimators complicates the life of applied statisticians who might otherwise use them. The bootstrap paradigm provides a simple method to make inferential statements that can be used with numerous estimators. In addition, there is hope the bootstrap can address complicated issues arising in selection of variables and model selection in general.

Consider the fuel-consumption data of Weisberg (1985 p.35–36, 126). The data consists of four explanatory variables on each of the 50 states of the United States. Note that our analysis includes Alaska and Hawaii which are often excluded. The response is Fuel, the fuel consumption in gallons per person, while the explanatory variables are Tax, the tax on fuel in cents per gallon, Dlic, the percentage of the population with a driver's license, Inc, the average income in thousands of dollars, and Road, the length of federal-aid primary highways in thousands of miles.

The classical least squares analysis is summarized in Table 3.1. Note that the coefficients of the variables Tax and Road are not significantly different from 0. An analysis of influence reveals that Hawaii, with a Cook's distance of 1.47, is the only influential state. After removing Hawaii from the analysis, the variable Tax becomes significantly different from 0. The question becomes: What should we do with Hawaii and the variable Tax?

The results of a robust analysis are also included in Table 3.1. The estimation procedure is the default estimator of the function *rreg* of the language S. It consists of a converged Huber estimate with constant 1.345 followed by two iterations of the Bisquare with constant 1.4. The coefficients are roughly the same as those of the least squares analysis after removing Hawaii, also included in Table 3.1. This may suggest that, after all, it is a good idea to remove Hawaii from the data set, but how can we assess if we should leave the variable Tax in the model?

One could use the asymptotic variance of the estimator, based on the influence curve, to provide a confidence interval for the coefficients. But what is the influence curve of a converged Huber estimate followed by two iterations of the Bisquare? The bootstrap algorithm provides a simple solution that can be used easily with numerous regression estimators. In our problem, a bootstrap 95% confidence interval for the parameter of the variable Tax is $[-43.3, -4.9]$ which suggests that the variable Tax should be left in the model. We shall return to this example after we describe the bootstrap algorithm in regression with fixed explanatory variables.

The model is the following: $y = X\beta + \epsilon$, where y is the vector of n observations, X is the fixed $n \times p$ full rank matrix of independent variables, β is the vector of unknown parameters of length p, and ϵ is a sample of size n of independently and identically distributed random variables with distribution function F, expectation 0 and finite variance. For technical reasons, we shall assume that a constant is included in the model.

In order to apply the bootstrap, a model must be fitted to the data, including an estimate of

the distribution F of the unobserved errors ϵ. Let $\hat{\beta} = \hat{\beta}(y)$ be an estimate of β computed from the observations y. We have already seen two specific estimators, the least squares and a converged Huber followed by two iterations of the Bisquare. Least absolute deviation will also be considered later. We are interested in finding estimates of bias and variance of the estimator $\hat{\beta}$ in estimating β. We are also interested in constructing confidence intervals for β and prediction intervals for a future observation with explanatory variables x_a. The problem of simultaneous confidence bands can also be considered.

The bootstrap is accomplished as follows. Let $\hat{e}$ be the vector of centered residuals, i.e., $\hat{e} = (X\hat{\beta} - y)$–mean $(X\hat{\beta} - y)$. Note that the ordinary residuals are already centered when the estimator is least squares and a constant term is included in the model. Let $\hat{F}$ be the empirical distribution function of the centered residuals. A bootstrap sample of observations is given by: $y^* = X\hat{\beta} + e^*$, where e^* is a random sample of size n from $\hat{F}$, i.e., a simple random sample with replacement from the residuals $\hat{e}$. From this bootstrap sample we can compute a bootstrap estimate of β, namely $\hat{\beta}^* = \hat{\beta}(y^*)$. Repeating this operation a large number of times, we can construct the distribution of $\hat{\beta}^* - \hat{\beta}$ which is the bootstrap estimate of the distribution of $\hat{\beta} - \beta$. From this, we immediately obtain the bootstrap estimates of bias and variance of $\hat{\beta}$, namely $E_*\hat{\beta}^* - \beta$ and $Var_*\hat{\beta}^*$ respectively where a starred subscript refers to the bootstrap distribution, i.e., conditional on the observed y.

Let w^*_α be the α-quantile of the bootstrap distribution of $\hat{\beta}^*_j - \hat{\beta}_j$. Then a confidence interval for β_j is given by

$$[\hat{\beta}^*_j - w^*_{1-\alpha}, \hat{\beta}^* - w^*_\alpha].$$

Note that this confidence interval is constructed as in the previous section. These are the bootstrap intervals that are computed in Table 3.1 for the fuel data using our robust estimator.

Now suppose that Wisconsin policy makers wished to reduce the fuel consumption in the state by raising the tax on a gallon of fuel from 7 cents to 20 cents, the actual tax on a gallon of gas as of January 1, 1988 in the state of Wisconsin according to the *World Almanac and Book of Facts 1989.* Thus, the goal is to construct a prediction interval for the new set of values of the independent variables. For simplicity's sake, the other three predictors will be left at their 1974 value. A bootstrap prediction interval is constructed as follows.

We want to estimate the distribution of the prediction error $D_a = \tilde{Y}_a - \hat{y}_a$ for the prediction based on x_a, a vector of explanatory variables, where $\tilde{y}_a = x_a\beta + \epsilon_a$ is a future observation, and $\hat{y}_a = x_a\hat{\beta}$ is its prediction based on the independent vector y. Note that ϵ_a is distributed according to F and independent from $\hat{\beta}$. The bootstrap estimation of that distribution consists of replacing $\tilde{y}_a$ by $\tilde{y}^*_a = x_a\hat{\beta} + \epsilon^*_a$ and $\hat{y}_a$ by $\hat{y}^*_a = x_a\hat{\beta}^*$ where ϵ^*_a is distributed according to $\hat{F}$ independently from $\hat{\beta}^*$ which is computed from the bootstrap data. So letting $D^*_a = \tilde{y}_a - \hat{y}^*_a$ and repeating a large number of times, we can approximate the bootstrap distribution of D^*_a by its empirical distribution. If we let u^*_α be the α-quantile of the bootstrap distribution of D^*_a, then a bootstrap prediction interval is given by

$$[x_a\hat{\beta} - u^*_{1-\alpha}, x_a\hat{\beta} - u^*_\alpha].$$

Changing the value of the variable Tax for Wisconsin from 7 cents to 20 cents, the vector of explanatory variables becomes $x' = (1, 20, 54.5, 4.207, 6.58)$. A bootstrap prediction interval based on our

robust estimator is given by $[-82.4, 517.3]$. This interval covers the actual per-capita consumption of 504 thousands gallons per person for the year 1987 in the state of Wisconsin, also given in the almanac. A bootstrap prediction interval based on the least squares estimator using all 50 states is given by $[45.4, 867.3]$. This interval also covers the 1987 value of 504, but is much wider. Note that this interval compares favorably with the classical prediction interval given by

$$\left[x_a\hat{\beta} - t(1-\alpha; n-p)\hat{\sigma}\sqrt{1 + x_a'(X'X)^{-1}x_a}, x_a\hat{\beta} - t(\alpha; n-p)\hat{\sigma}\sqrt{1 + x_a'(X'X)^{-1}x_a}\right],$$

where $\hat{\sigma}^2$ is the usual estimator of the variance of F based on the residual sum of squares and $t(\alpha; n)$ is the α-quantile of a t-distribution with n degrees of freedom. For this example, the classical interval is given by $[45.4, 853.3]$.

It was argued before that Hawaii was exerting too much influence on the least squares estimates and should probably be dropped from the analysis. If we do that, the bootstrap prediction interval based on least squares and the classical interval are given by $[-274.3, 436.7]$ and $[-315.3, 427.1]$, respectively. Note that neither interval covers the 1987 value despite the fact that they are almost as wide as the previous intervals based on least squares and all states and still much larger than the robust interval.

Simultaneous confidence bands are sometimes needed and Hall and Pittelkow (1990) describe a bootstrap method to construct such bands. They are based on the least squares estimator, where classical solutions exist, but they have the advantage of being asymptotically correct for fairly general distributions on the errors and to allow for a wide variety of shapes.

Classical least squares estimation and its normal-based inference augmented by residual plots and influence measures remain the basic regression tools. Our analysis of the fuel data showed us that Hawaii has a large influence on the estimated least squares coefficients and consequently on the variables that should be deleted from the model. Specifically, including Hawaii in the analysis excludes the variable Tax. The absence or presence of this variable may be very important for Public Policy makers. It may be possible to rationalize removing Hawaii from the analysis, for instance, on the grounds that it is not one of the contiguous states, in which case Alaska should also be removed.

On the other hand, a robust regression estimator can reduce the weight of influential observations. The bootstrap can then be used to make valid approximate inferences based on such a robust estimator. Such applications of the bootstrap can now be done routinely and should be done more often. Some argue that least squares estimation accompanied by some deletions of influential observations leads to a robust estimation. Nevertheless, the fact remains that inference made after deleting influential observations does not take into account such deletions and can therefore be misleading. Our classical and bootstrap prediction intervals for Wisconsin based on the least squares estimator without Hawaii is such an example. Bootstrap inference based on a robust estimator does not suffer from that problem.

3.2 Theory and Enhancements

Many papers have been devoted to the asymptotic and small sample performance of bootstrap estimators in regression. We will begin by discussing the least squares estimator.

Among the key papers in the area are Freedman (1981) and Bickel and Freedman (1983). The first paper treated the case of a fixed number of parameters whereas the second one considered an increasing number of parameters. If p is fixed, $n^{-1}(X(n)'X(n)) \to V$ where $X(n)$ is the design matrix with n observations and V is positive definite, and if F, the distribution of the errors, has mean 0 and finite variance σ^2, the bootstrap is valid. That means that the coverage of bootstrap confidence intervals for parameters converges to the nominal level. Note that these conditions are sufficient for the classical normal-theory based intervals to be asymptotically valid. In that paper, the problem of random explanatory variables has also been treated. We refer the interested reader to that paper, but we want to point out that the resampling unit is not a residual, but rather a vector (Y_i, x_i) where x_i is the vector of explanatory variables for the ith observation. When p is not fixed, Bickel and Freedman (1983) showed that $p/n \to 0$ is sufficient for bootstrapping (normalized) contrasts, whereas $p^2/n \to 0$ is sufficient when bootstrapping the whole p-dimensional distribution.

Hall (1989) has shown that bootstrap confidence intervals have unusually good properties if the parameter of interest is a slope, i.e., all parameters but the constant. The critical points of the interval are second order correct and the coverage error of both one-sided and two-sided intervals is $O(n^{-1})$. This is in contrast to most statistical problems where they are first order correct and the coverage error of one-sided intervals is $O(n^{-1/2})$ and $O(n^{-1})$ for two-sided intervals.

Hall (1989) also discusses studentizing. This consists in bootstrapping the distribution of $(\hat{\beta} - \beta)/\hat{\sigma}$ where $\hat{\sigma}^2$ is the maximum likelihood estimate of variance of F. So instead of simply computing $\hat{\beta}^*$ from the bootstrap data, an estimate of variance is also computed, namely $(\hat{\sigma}^*)^2 = n^{-1}\text{RSS}^*$, where RSS* is the residual sum of squares of the bootstrap data. The intervals, called percentile-t intervals, are also unusually good for slope estimates; they are second-order correct and have coverage errors which are $O(n^{-3/2})$ for one-sided intervals and $O(n^{-2})$ for two-sided intervals.

These results establish the legitimacy of the bootstrap for large samples. In small samples, some problems may arise. In particular, there may be a problem of underfitting due to the fact that the residuals are smaller than the actual errors. One possible solution is to rescale the residuals by multiplying them by $\sqrt{n/(n-p)}$ as suggested by Bickel and Freedman (1983). Weber (1984) and Stine (1985) suggest using the centered studentized residuals, i.e., $\hat{e}_i(1 - h_{ii}^{-1/2}) - \tilde{\mu}$ where $\tilde{\mu} = 1/n \sum e_j(1 - h_{jj})^{-1/2}$, and h_{ii} is the ith diagonal element of the hat matrix $X(X'X)^{-1}X'$. The residuals of this asymptotically valid solution all have the same variance. Some simulation results for bootstrap prediction intervals based on the least squares estimator are presented in Stine (1985).

The application of the bootstrap is very much model dependent. Therefore, if certain crucial hypotheses of the model are not met, the bootstrap may obviously fail. One such assumption is the homogeneity of the variance of the errors. Wu (1986) shows that in the presence of variance heteroscedasticity, the bootstrap, which relies on the exchangeability of the errors, is inconsistent in estimating the variance of $\hat{\beta}$. Wu's paper and its discussion provide many possible solutions to the problem, including some based on bootstrap ideas.

Theoretical justification for bootstrap procedures based on estimators other than least squares is much more scarce. On the other hand, a number of small sample simulations have been performed. Shorack (1982) shows that the bootstrap applied to contrasts is asymptotically valid when using robust M-estimates. The regularity conditions are moment conditions and smoothness conditions

on the ψ function. Simulations for bootstrap estimates using the least absolute deviations estimator are reported in Dielman and Pfaffenberger (1988), Stangenhaus (1987), Morey and Schenck (1984), and also for bootstrap estimates using the least maximum absolute residual in Hand (1986).

The bootstrap can also be used in other regression contexts. For instance, Delaney and Chatterjee (1986) proposed a bootstrap method for choosing the ridge parameter in a ridge regression, although their bootstrap unit is a vector of observations (Y_i, x_i) rather than a residual. Freedman (1984) and Freedman and Peters (1984) have applied the bootstrap in more complex regression models, such as a dynamic linear model estimated by two-stage least squares estimates. They show that it is asymptotically valid, like classical methods, but that the bootstrap solution sometimes outperforms the classical methods. Efron and Tibshirani (1986) have also applied the bootstrap to Cox's proportional hazards model and to projection pursuit.

4. Resampling Schemes for Dependent Data.

4.1. The bootstrap applied to time series.

The problem of nonparametrically estimating the variance and/or the sampling distribution of statistics based on data $X_1, X_2, \ldots, X_N$ is considerably more difficult in the case the data are not independent. Recall that a sequence of random variables $\{X_n, n \in \mathbf{Z}\}$ is called (strictly) stationary if, for all $n \in \mathbf{N}$, the joint distribution of $(X_k, X_{k+1}, \ldots, X_{k+n})$ does not depend on k; a stationary sequence is called m-dependent if the set of random variables $\{X_n, n = -1, -2, \ldots\}$ is independent of $\{X_n, n = m, m+1, \ldots\}$. In this setting, independence can be thought of as 0-dependence.

Consider the simple case where the $X_1, \ldots, X_N$ are observations from a univariate m-dependent stationary sequence, and the statistic of interest is the sample mean $\bar{X}_N$. Then, $E\bar{X}_N = \mu$, where $\mu = EX_1$, and (due to stationarity) if $m \leq N$,

$$\sigma_N^2 \equiv Var(\sqrt{N}\bar{X}_N) = Var(X_1) + 2\sum_{i=1}^{m}(1 - \frac{i}{N})Cov(X_1, X_{1+i}).$$

It is easy to show that $\bar{X}_N$ is asymptotically normal, so to obtain confidence intervals for μ we would just need a consistent estimate of $\sigma_\infty^2 \equiv \lim_{N\to\infty} \sigma_N^2 = Var(X_1) + 2\sum_{i=1}^{m} Cov(X_1, X_{1+i})$.

Looking at the classical (i.i.d.) bootstrap estimate of variance of $\sqrt{N}\bar{X}_N$, it is apparent that it converges in probability to $Var(X_1)$, and therefore it is inconsistent for σ_∞^2. The reason of course is that the classical bootstrap resampling scheme is valid for *independent* data, and is insensitive to their 'time-order'. In other words, the classical bootstrap procedure utilizes the data only through the empirical distribution that they define. This empirical distribution is an approximation to the first-marginal distribution of the sequence $\{X_n\}$, and hence, based on it, we cannot expect to estimate a parameter like σ_∞^2 which depends on the $(m+1)$th-marginal distribution of the m-dependent observations.

Recently, Künsch(1989) and Liu and Singh(1988) have each independently proposed a block-resampling scheme that takes care of this problem by working with empirical estimates of the b-dimensional marginal distribution of the sequence $\{X_n\}$. This method (hereafter termed the 'moving blocks' method), can be described as follows. Define $\mathcal{B}_i$ to be the block of b consecutive observations starting from X_i, that is $\mathcal{B}_i = (X_i, \ldots, X_{i+b-1})$, where $i = 1, \ldots, q$ and $q = N - b + 1$. Sampling with replacement from the set $\{\mathcal{B}_1, \ldots, \mathcal{B}_q\}$, defines a (conditional on the original data) probability measure P^* which is used in the 'moving blocks' bootstrap procedure. If k is an integer such that $kb \sim N$, then letting $\xi_1, \ldots, \xi_k$ be drawn i.i.d. from P^*, it is seen that each ξ_i is a block of b observations $(\xi_{i,1}, \ldots, \xi_{i,b})$. If all $l = kb$ of the $\xi_{i,j}$'s are concatenated in one long vector denoted by $Y_1, \ldots, Y_l$, then the 'moving blocks' bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$ is the variance of $\sqrt{l}\bar{Y}_l$ under P^*, and the 'moving blocks' bootstrap estimate of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ is $P^*\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$, where $\bar{Y}_l = \frac{1}{l}\sum_{i=1}^{l} Y_i$. As a final step, confidence intervals for μ can be obtained either by means of the Central Limit Theorem using the 'moving blocks' bootstrap estimate of variance, or by approximating the quantiles of the distribution $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ by the corresponding quantiles of $P^*\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$. If P^* probabilities turn out to be cumbersome to analytically calculate, one can always resort to Monte Carlo, i.e. drawing a large

number of samples $\xi_1^{(j)}, \ldots, \xi_k^{(j)}$ i.i.d. from P^*, where $j = 1, \ldots, B$, and evaluating the required probabilities or quantiles empirically from the Monte Carlo set of the B *resamples*. It is obvious that taking $b = 1$ makes the 'moving blocks' bootstrap coincide with the classical (i.i.d.) bootstrap of Efron(1979).

It can be shown (cf. Lahiri(1990)) that a slightly modified 'moving blocks' bootstrap estimate of sampling distribution turns out to be more accurate than the normal approximation, under some regularity conditions, resulting in more accurate confidence intervals for μ. The modification amounts to approximating the quantiles of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ by the corresponding quantiles of $P^*\{\sqrt{l}(\bar{Y}_l - E^*\bar{Y}_l) \leq x\}$, where $E^*\bar{Y}_l$ denotes the expected value of $\bar{Y}_l$ under the P^* probability.

To illustrate the 'moving blocks' method, consider the following numerical example taken from Politis and Romano(1989). A sample $Y_1, \ldots, Y_{100}$ was generated from the moving average model: $Y_t = Z_t + Z_{t-1} + Z_{t-2}$, where the Z_t's are i.i.d. N(0,1). By observing that $\sum_{i=1}^{100} Y_i \simeq 3\sum_{i=1}^{100} Z_i$, it is immediate that $Var(\frac{1}{10}\sum_{i=1}^{100} Y_i) \simeq 9$. A plot of the 'moving blocks' bootstrap estimate of the variance of $\frac{1}{10}\sum_{i=1}^{100} Y_i$ as a function of the block size b is shown in Figure 4.1. As expected, the classical (i.i.d.) bootstrap (that corresponds to the choice $b = 1$) underestimates the variance, yielding an estimate about 3. The 'moving blocks' bootstrap with b near the value 10 seems to give the most accurate estimate, while taking a greater b *worsens* the approximation, and for $b \geq 20$ the 'moving blocks' correction is totally lost.

In Politis and Romano(1989,1990), the 'blocks of blocks' resampling scheme was introduced, in order to address the problem of setting confidence intervals for parameters associated with the whole (infinite-dimensional) distribution of the $X_1, X_2, \ldots$ observations. A prime example of such a parameter is the spectral density function of the $\{X_n\}$ sequence, evaluated at a point. As a by-product, the 'blocks of blocks' method also provides more accurate confidence intervals for parameters associated with a finite-dimensional distribution of the observations, as compared to confidence intervals obtained by the normal approximation. Examples of such parameters include the autocovariance $Cov(X_0, X_s)$ and the autocorrelation $Cov(X_0, X_s)/Var(X_0)$ at lag s. The 'blocks of blocks' scheme is a generalization of the 'moving blocks' method, and the two coincide if the parameter under consideration is the mean EX_1.

The 'blocks of blocks' resampling scheme will be discussed in the next section, together with a general description of the method and its properties. Also discussed is a related resampling scheme, the 'stationary bootstrap' (cf. Politis and Romano(1991)), that has the additional property that the pseudo-sequence obtained by resampling is stationary, as is the original sequence. In addition, some interesting applications and examples of the aforementioned resampling schemes will be presented, where again the objective is to construct accurate estimates of sampling distribution for statistics that are smooth functions of sample-mean type estimators.

Although our main emphasis is in nonparametric estimation, let us mention here that, analogously to the independent case, a parametric bootstrap can be formulated for time series models as well. For example, if it is assumed that the time series is Gaussian, with mean zero and autocovariance sequence R (unknown), bootstrap replications can be constructed by simulating Gaussian time series with mean zero and estimated autocovariance $\hat{R}$, (cf. Ramos(1988)).

For completeness, let us also describe how the classical (i.i.d.) bootstrap can be successfully modified to apply to a certain class of time series models. These models are characterized by the

fact that they are generated by a 'white' noise sequence (also called the 'residuals'), that is just a sequence of i.i.d. random variables. Although this noise sequence is not directly observable (as is in the classical bootstrap setting), its underlying distribution can typically be estimated consistently. To fix ideas, suppose $\{Y_n, n \in \mathbf{Z}\}$ is a stationary autoregressive process of order p, i.e. $Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + e_n$, where $\mathbf{a} = (a_1, \ldots, a_p)$ is a p-dimensional unknown parameter, and the residuals e_n are i.i.d. with (unknown) distribution F, such that $Ee_n = 0$ and $Ee_n^2 = 1$. After observing a finite sample, the problem at hand is to establish confidence intervals for the parameters $a_1, \ldots, a_p$, or for some other related quantity, e.g. $h(a_1, \ldots, a_p)$, where $h(\cdot)$ is a smooth function. The bootstrap algorithm in this situation would be the following (cf. Freedman(1984), Freedman and Peters(1986), Efron and Tibshirani(1986), Swanepoel and van Wyk(1986), Bose(1988)). Suppose that a sample $Y_{1-p}, \ldots, Y_N$ is observed and least squares estimates $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_p)$ are calculated based on the $Y_1, \ldots, Y_N$ observations. Then, estimates of the residuals are formed by letting $\hat{e}_n = Y_n - \sum_{i=1}^{p} \hat{a}_i Y_{n-i}$, for $n = 1, \ldots, N$. Define the centered estimated residuals as $\tilde{e}_n = \hat{e}_n - \frac{1}{N}\sum_{n=1}^{N} \hat{e}_n$, for $n = 1, \ldots, N$, and let $\tilde{F}_N(\cdot)$ denote the distribution function that puts mass $1/N$ at each $\tilde{e}_n$. Now let $e_1^*, \ldots, e_N^*$ be an i.i.d. sample with distribution $\tilde{F}_N$. Finally generate the resampled sequence Y_n^* as follows: Let $Y_n^* = Y_n$, for $n = 1-p, \ldots, 0$, and let $Y_n^* = \sum_{i=1}^{p} \hat{a}_i Y_{n-i}^* + e_n^*$, for $n = 1, \ldots, N$. Based on the $Y_1^*, \ldots, Y_N^*$ resample, new least squares estimates $\mathbf{a}^* = (a_1^*, \ldots, a_p^*)$ can be computed, and the bootstrap estimate of the (multivariate) sampling distribution of $\sqrt{N}(\hat{\mathbf{a}} - \mathbf{a})$ is the distribution (conditional on the data $Y_{1-p}, \ldots, Y_N$) of $\sqrt{N}(\mathbf{a}^* - \hat{\mathbf{a}})$. It has been shown (cf. Bose(1988)) that the bootstrap approximation to the sampling distribution of $\hat{\mathbf{a}}$ is quite accurate; indeed, under some conditions, it is more accurate than the asymptotic normal distribution for $\hat{\mathbf{a}}$.

The above bootstrap algorithm can be easily generalized to mixed autoregressive moving average models, in which the residuals can be estimated in a similar way (cf. Box and Jenkins (1970)). It can also be generalized to nonlinear autoregressions, e.g. of the type $Y_n = f(Y_{n-1}, \ldots, Y_{n-p}) + e_n$, provided the function f can be estimated consistently. Examples (with $p = 1$) of assumed models for the function f include a smooth parametric model, e.g. $f(x) = a_1 cos(x - a_2)$, where a_1, a_2 are the unknown parameters, or a threshold model (cf. Tong(1983)), e.g. $f(x) = a_1 x$, if $x > a$, and $f(x) = a_2 x$, if $x \leq a$, where a, a_1, a_2 are the unknown parameters. Alternatively, f could be estimated nonparametrically, using methods of nonparametric regression such as kernel smoothing and multivariate splines (cf. Friedman(1991), Lewis and Stevens(1990)). Finally, it should be mentioned that in Franke and Härdle (1987) the ideas related to bootstrapping a nonparametric regression (cf. Härdle and Bowman (1988)) are applied in the setting of kernel smoothed spectral estimation.

4.2. The 'blocks of blocks' general resampling scheme and its properties

A general resampling procedure for stationary time series, termed the 'blocks of blocks' resampling scheme (cf. Politis and Romano (1989, 1990)), will now be outlined, and some examples of its implementation will be presented. The 'blocks of blocks' scheme is a nonparametric procedure that yields confidence intervals of asymptotically correct coverage for a parameter associated with the infinite-dimensional joint distribution of the terms in the time series.

Let $X_1, \ldots, X_N$ be observations from the (strictly) stationary multivariate time series $\{X_n, n \in \mathbf{Z}\}$, where X_1 takes values in $\mathbf{R}^d$. The time series $\{X_n, n \in \mathbf{Z}\}$ is assumed to have a weak depen-

dence structure. With an α-mixing assumption there is associated a certain notion of asymptotic independence of the 'future' of X_k with the 'past' of X_0. Examples where $\alpha_X(k) = O(k^{-\lambda})$, for some $\lambda > 0$, include (but are not limited to) a Gaussian time series with a sufficiently smooth spectral density, or a time series generated by an ARMA (autoregressive moving average) model with innovations that have a general absolutely continuous distribution. In particular, Gaussian ARMA processes with bounded spectral density have the α-mixing coefficients decreasing geometrically, and a moving average (MA) process of order p will have $\alpha_X(k) = 0, \forall k > p$.

As a first step, set up the estimation problem in the following manner. Suppose $\mu \in \mathbf{R}^D$ is a parameter of the joint distribution of the sequence $\{X_n, \in \mathbf{Z}\}$. For each $N = 1, 2, \ldots$ let $B_{i,M,L}$ be the block of M consecutive observations starting from $(i-1)L + 1$, i.e., the subseries $X_{(i-1)L+1}, \ldots, X_{(i-1)L+M}$, where M, L are integer functions of N. Define $T_{i,M,L} = \phi_M(B_{i,M,L})$, where $\phi_M : \mathbf{R}^{dM} \rightarrow \mathbf{R}^D$ is some function. So for fixed N, the $T_{i,M,L}$ for $i \in \mathbf{Z}$ constitute a strictly stationary sequence. In practice, a segment $X_1, \ldots, X_N$ from the time series $\{X_n\}$ would be observed, which would permit us to compute $T_{i,M,L}$ for $i = 1, \ldots, Q$ only, where $Q = [\frac{N-M}{L}] + 1$ and $[\cdot]$ is the integer part function. Now, define the general linear statistic:

$$\bar{T}_N = \frac{1}{Q} \sum_{i=1}^{Q} T_{i,M,L}$$

Under broad regularity conditions $\bar{T}_N$ is a consistent estimator of μ. Loosely stated, these regularity conditions consist of a weak dependence structure (allowing the variance of $\bar{T}_N$ to tend to zero as $N \rightarrow \infty$), and a condition of unbiasedness or asymptotic unbiasedness of $T_{1,M,L}$, i.e., $ET_{1,M,L} = \mu$, or $ET_{1,M,L} \rightarrow \mu$ as $M \rightarrow \infty$.

Some examples of time series statistics that can fit in this framework are the following. For the examples assume X_n is univariate, that is $d = 1$.

(I) The sample mean : $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$. Just take $M = L = 1$ and ϕ_M to be the identity function.

(II) The (unbiased) sample autocovariance at lag s: $\frac{1}{N-s} \sum_{i=1}^{N-s} X_i X_{i+s}$. Take $L = 1, M = s+1$ and $\phi_M(x_1, \ldots, x_M) = x_1 x_M$.

(III) The lag-window spectral density estimator: take

$$\phi_M(B_{i,M,L}) = \frac{1}{2\pi M} \left| \sum_{t=L(i-1)+1}^{L(i-1)+M} W_t^{(M)} X_t e^{-jtw} \right|^2$$

i.e., $T_{i,M,L}(w)$ is the periodogram of block $B_{i,M,L}$ of data 'tapered' by the function $W_t^{(M)}$, and evaluated at the point $w \in [0, 2\pi]$. (Note that the symbol j denotes the unit of imaginary numbers $\sqrt{-1}$, in order to avoid confusion with i, the block count.)

Note that in example (I), μ is just EX_1, i.e. it is a parameter of the m-dimensional marginal distribution of sequence $\{X_n, n \in \mathbf{Z}\}$, with m=1. Similarly, in example (II), $\mu = EX_0 X_s$ is a parameter of the m-dimensional marginal, with m=s+1, and in example (III), μ is the spectral density evaluated at the point w, i.e. a parameter of the whole (infinite-dimensional) joint distribution of $\{X_n, n \in \mathbf{Z}\}$.

With the objective of setting confidence intervals for μ, the 'blocks of blocks' bootstrap procedure goes as follows. Define $\mathcal{B}_{j,b}$ to be the block of b consecutive $T_{i,M,L}$'s starting from $T_{j,M,L}$; that is, let $\mathcal{B}_{j,b} = (T_{j,M,L}, \ldots, T_{j-1+b,M,L})$. Note that there are $q = Q - b + 1$ such $\mathcal{B}_{j,b}, j = 1, \ldots, q$. Sampling with replacement from the set $\{\mathcal{B}_{1,b}, \ldots, \mathcal{B}_{q,b}\}$ defines (conditionally on the original observations $X_1, \ldots, X_N$) a probability measure denoted by P^*, which is used in the 'blocks of blocks' bootstrap procedure. Let $Y_1, \ldots, Y_k$ be i.i.d. samples from P^*, where k is of the same asymptotic order as Q/b, (for instance, let $k = [Q/b] + 1$). Obviously, each Y_i is a block of size b which we denote as $Y_i = (y_{i1}, \ldots, y_{ib})$. Let us concatenate the y_{ij} in one long vector of size $l = kb$ denoted by $T_1^*, \ldots, T_l^*$, where $T_i^* = y_{rv}$, for $r = [i/b], v = i - br$. Now both $P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\}$ and $P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x\}$ constitute 'blocks of blocks' bootstrap estimates of $P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x\}$, where $\bar{T}_l^* = \frac{1}{l}\sum_{i=1}^{l} T_i^*$, and the variance of $\sqrt{l}\bar{T}_l^*$ under the P^* probability constitutes the 'blocks of blocks' bootstrap estimate of the variance of $\sqrt{Q}\bar{T}_N$.

As can be easily checked, the 'moving blocks' technique is a special case of the 'blocks of blocks' resampling scheme, as applied to the sample mean example (I). Under mixing and moment conditions, consistency of the 'blocks of blocks' bootstrap estimate of sampling distribution was proved in Politis and Romano(1989,1990) in the general case (where m might be infinite). Some direct limit statements are:

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x(Var^*(\sqrt{l}\bar{T}_l^*))^{1/2}\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \leq x\sigma_\infty\}| \to 0 \qquad (4.1)$$

in probability and

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \leq x\}| \to 0 \qquad (4.2)$$

in probability. where E^* denotes expectation under the P^* probability. If in addition $ET_{1,M,L} = \mu + o(Q^{-1/2})$, then μ can be substituted instead of $E\bar{T}_N$ in equations (4.1) and (4.2)

To actually compute confidence intervals using the (4.1) or (4.2) approximations, one has to compute (or approximate by Monte Carlo) the corresponding quantiles of $P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x(Var^*(\sqrt{l}\bar{T}_l^*))^{1/2}\}$ or $P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\}$. Finding a constant C^* that satisfies $P^*\{E^*\bar{T}_l \leq \bar{T}_l^* + C^*\} = 1 - \epsilon$, would then immediately imply that $P\{E\bar{T}_N \leq \bar{T}_N + C^*\} \simeq 1 - \epsilon$, and, also that $P\{\mu \leq \bar{T}_N + C^*\} \simeq 1 - \epsilon$, the latter provided $ET_{1,M,L} = \mu + o(Q^{-1/2})$, that is, if the bias of the estimator $T_{1,M,L}$ is of smaller order than its standard deviation.

To fix ideas, suppose we are looking for a 95% equal tailed confidence interval for μ, under the assumption $ET_{1,M,L} = \mu + o(Q^{-1/2})$. Other types of confidence intervals (e.g. symmetric around μ or smallest length) can be treated in the same way. Starting with the approximation $.975 = P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x^*\} \simeq P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x^*\}$, note that $l^{-1/2}x^* + E^*\bar{T}_l^*$ is just the .975 quantile of the bootstrap distribution $P^*\{\bar{T}_l^* \leq x\}$. Similarly, define y^* such that $l^{-1/2}y^* + E^*\bar{T}_l^*$ is the .025 quantile of $P^*\{\bar{T}_l^* \leq x\}$. Then, the 95% equal tailed confidence interval for μ is $[\bar{T}_N - x^*/\sqrt{Q}, \bar{T}_N - y^*/\sqrt{Q}]$. This, in the terminology of Hall(1988), is a 'hybrid' bootstrap confidence interval based on the approximation (4.2). Using the approximation (4.1) would lead to a 'bootstrap-t' confidence interval.

An important observation is that $Var^*(\sqrt{l}\bar{T}_l^*) = \frac{b}{q}\sum_{i=1}^{q}(\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L} - E^*\bar{T}_l^*)^2$, and $E^*\bar{T}_l^* = \frac{1}{q}\sum_{i=1}^{q}\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L}$, both of which can be computed without resampling. The

variance estimate $Var^*(\sqrt{l}\bar{T}_l^*)$ is asymptotically equivalent to the estimate $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = \frac{b}{q}\sum_{i=1}^{q}(\frac{1}{b}\sum_{j=i}^{i-1+b}T_{j,M,L} - \bar{T}_N)^2$, which is referred to as the 'blocks of blocks' jackknife estimate of variance.

Let us break for a moment to illustrate the practical implementation of the block-resampling scheme just presented by means of an example. Consider the Canadian Lynx data (annual number of lynx trappings in the Mackenzie River for the period 1821 to 1934), that are available to all users of the S statistical language (cf. Becker, Chambers, and Wilks(1988)). It is important to note that the Lynx data have been shown (cf. Subba Rao and Gabr (1980)) to be non-linear (and non-Gaussian), and hence would not succumb to usual linear parametric modeling, such as fitting ARMA (autoregressive moving average) models. Suppose that we are interested in obtaining a 95% confidence interval for the mean of the annual number of lynx trappings, i.e. EX_1, where the data $X_1, \ldots, X_{114}$ are pictured in Figure 4.2. Note that $\bar{X}_{114} = \frac{1}{114}\sum_{i=1}^{114}X_i = 1538.018$. The first step is to compute a variance estimate for $\sqrt{114}\bar{X}_{114}$. Using the estimator $\hat{V}_{JACK}$, (the bootstrap estimate $Var^*(\sqrt{l}\bar{T}_l^*)$ is practically indistinguishable from $\hat{V}_{JACK}$ for all $b = 1, 2, \ldots, 40$), involves choosing the design parameter b appropriately. Different choices of b lead to quite different estimates as shown in Figure 4.3.

Now, it is known in this case that to have a variance estimator with asymptotically smallest Mean Squared Error (M.S.E.) it is necessary that $b^3 \sim \frac{3N}{4\sigma_\infty^4}(\sum_{s=-\infty}^{\infty}|s|R(s))^2$, where $\sigma_\infty^2 = \lim_{N\to\infty} Var(\sqrt{N}\bar{X}_N)$, and $R(s) = Cov(X_1, X_{1+s})$ (cf. Künsch(1989)). This is hardly a new result, since, for the particular case of the sample mean, $\hat{V}_{JACK}$ is essentially a nonparametric estimate of the spectral density at zero, smoothed by Bartlett's kernel. A rough estimate of $\sum_{s=-\infty}^{\infty}|s|R(s)$ based on the sample autocovariances can be obtained. However, σ_∞^2 is unknown. In fact, this is exactly what $\hat{V}_{JACK}$ asymptotically estimates. Looking at the variability of $\hat{V}_{JACK}$ for different choices of b, indicates that this route for choosing b is really a vicious circle. Therefore, the choice of b requires an educated guess based on studying the data more deeply. In the case of the lynx data, looking at the estimated sample autocovariance sequence $\hat{R}(s) = \frac{1}{114}\sum_{i=1}^{114-|s|}X_iX_{i+|s|}$, or at its unbiased version $\frac{114}{114-|s|}\hat{R}(s)$, is particularly enlightening. In Figure 4.4, a plot of $\hat{R}(s)/\hat{R}(0)$ is presented, i.e. the sample autocorrelation sequence. There is a clear indication of cyclical behavior with a period of about ten years (cf. Cambell and Walker (1977) and the references therein). Bearing in mind that $Var(\sqrt{N}\bar{X}_N) = R(0) + 2\sum_{i=1}^{N-1}(1 - \frac{i}{N})R(i)$, it can be seen that chosing $b = b_0$ effectively retains only the first b_0 autocovariances in the summation. But by the almost periodicity of $R(s)$, one expects that there will be a lot of cancellations among the terms in the above sum. Hence, it would be advisable to let b be of the order of two or three cycles, in order to allow for the cancellations to take place. We would opt to take $b = 25$ in this case, leading to a variance estimate of 2,873,828.

Returning to the problem of determining b via minimum M.S.E. considerations, it is quite interesting that solving for b in the equation $b^3\hat{V}_{JACK}^2 = (3N/4)\tilde{R}^2$, where $\tilde{R} = 2\sum_{s=1}^{40}s(1 - s/40)\hat{R}(s) = -30,533,255$ is an estimate of $\sum_{s=-\infty}^{\infty}|s|R(s)$ using Bartlett's kernel, and $\hat{V}_{JACK}$ (depend on b) and is obtained from Figure 4.3, leads to $b = 22$ and an estimated variance of 2,853,373 (taken from Figure 4.3 with $b = 22$). However, it should be noted that due to the asymptotic nature of this way of determining b (and due to the high variance of $\tilde{R}$) we would not *a priori* trust this result if it were not supported by some other reasoning.

Now that a variance estimate for $\bar{X}_N$ is available (the variance estimate is 2,873,828/114 = 25209, and the standard deviation estimate is 158.8), a 95% confidence interval for EX_1 using the Central Limit Theorem would be [1226.8, 1849.3]. Note that this confidence interval could be inaccurate because the distribution of the lynx data seems to be excessively skewed (see Figure 4.5).

A traditional way out of this difficulty is the use of transformations. In the literature (cf. Subba Rao and Gabr(1980), Cambell and Walker(1977)) it is suggested to use the logarithmic transformation data. However, this leads to a confidence interval for $E \log X_1$, and does not transform to a confidence interval for EX_1. Alternatively, one can use a transformation on $\bar{X}_N$ and the δ-method (cf. Miller(1986)). Using the logarithmic transformation on $\bar{X}_N$ (and the asymptotic normal distribution of $\log \bar{X}_N$) leads to the [1256.3, 1882.9] 95% confidence interval for EX_1, which is markedly different from the previous interval.

The question then is: which of the two confidence intervals is better? The bootstrap solves this dilemma, because it automatically captures the skewness without the need for transformations. In Figures 4.6 and 4.7, histograms of the 'moving blocks' bootstrap distribution of $\bar{X}_N$ are presented for two bootstrap simulations. The choice of design parameters for Figure 4.6 was $k = 20$, $b = 25$, and the number of bootstrap replications was $B = 500$; for Figure 4.7, $k = 5$, $b = 25$, and $B = 100$. The 95% bootstrap (hybrid) confidence intervals for EX_1 were [1233.37, 1826.07] (Figure 4.6), and [1221.03, 1862.62] (Figure 4.7). It is apparent that the bootstrap confidence intervals are quite close to the confidence interval derived by the Central Limit Theorem, not surprisingly since the bootstrap histograms do not appear to be too skewed. The histogram of Figure 4.6 has coefficient of skewness 0.333, compared to 1.33 of the empirical distribution of the lynx data.

It is interesting to note that if μ is a parameter of the m-dimensional marginal distribution of sequence $\{X_n\}$, with m finite, then M could be taken to be a *fixed* constant equal to m, and L can be taken equal to one in the 'blocks of blocks' procedure. In this case, and under some additional regularity conditions (including that $ET_{1,M,L} = \mu$, and that $\alpha(k)$ has an exponential decay), it has been proved (Lahiri(1990), Politis and Romano(1990)) that the approximation provided by equation (4.1) is more than first-order accurate. This fact establishes that the bootstrap approximation (4.1) is preferable to the normal approximation provided by a Central Limit Theorem for $\bar{T}_N$, especially if there is significant skewness in the distribution of the $T_{i,M,L}$'s.

The reason that (4.1) provides a more accurate approximation than (4.2) is that $E^*\bar{T}_l^* = \bar{T}_N + O_p(b/Q)$, i.e. the distribution of $\bar{T}_l^*$ under P^* possesses a random bias of significant order.

A way to dispense with this difficulty is to 'wrap' the $T_{i,M,L}$'s around in a 'circle', that is, to define (for $i > Q$) $T_{i,M,L} \equiv T_{i_Q,M,L}$, where $i_Q = i(\mathrm{mod} Q)$, and $T_{0,M,L} \equiv T_{Q,M,L}$. This idea is incorporated in the following general formulation which, for reasons to become apparent, will be called the *stationary* (blocks of blocks) resampling scheme.

With the $T_{i,M,L}$'s defined for all i as above, define the $\mathcal{B}_{j,b}$ as previously, but note that now for any integer b there are Q such $\mathcal{B}_{j,b}, j = 1, \ldots, Q$. Let p be a number in [0,1]. Independent of $X_1, \ldots, X_N$, let $L_1, L_2, \ldots$ be a sequence of independent and identically distributed random variables having the geometric distribution, so that the probability of the event $\{L_i = n\}$ is $(1-p)^{n-1}p$ for $n \in \mathbf{N}$. Independent of the X_i and the L_i, let $I_1, I_2, \ldots$ be a sequence of independent and identically distributed variables which have the discrete uniform distribution on $\{1, \ldots, N\}$. Now,

a new pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $\mathcal{B}_{I_1,L_1}, \mathcal{B}_{I_2,L_2}, \ldots$. The first L_1 observations in the pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ are determined by the first resampled block $\mathcal{B}_{I_1,L_1}$, the next L_2 observations in the pseudo-sequence are the observations in the second resampled block $\mathcal{B}_{I_2,L_2}$, and so on. The process is stopped once Q observations in the pseudo-sequence have been generated. It is not hard to see that the pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ is *stationary*, conditionally on the original data.

This method of resampling and generating $T_1^\star, \ldots, T_Q^\star$ defines a (conditional on the original data $X_1, \ldots, X_N$) probability measure $P^\star$. If we define $\bar{T}_Q^\star = \frac{1}{Q}\sum_{i=1}^{Q} T_i^\star$, then it is easy to see that $E^\star \bar{T}_Q^\star = \bar{T}_N$, where $E^\star$ denotes expectation under the $P^\star$ probability. Hence, the stationary blocks of blocks bootstrap estimate of $P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x\}$ is $P^\star\{\sqrt{Q}(\bar{T}_Q^\star - \bar{T}_N) \leq x\}$.

Under mixing and moment conditions, it can be shown (cf. Politis and Romano (1991)) that the stationary bootstrap estimates of variance and of distribution are consistent. Considering the Lynx data again, a problem equivalent to determining b in the 'moving blocks' or 'blocks of blocks' bootstrap presents itself, namely choosing the design parameter p. It can be shown (cf. Politis and Romano(1991)) that the stationary bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$ can be analytically calculated by $\hat{V}_{St.B.} = \hat{R}(0) + 2\sum_{i=1}^{N-1} h_N(i)\hat{R}(i)$, where $h_N(i) = (1 - \frac{i}{N})(1-p)^i + \frac{i}{N}(1-p)^{N-i}$. Note that the stationary bootstrap resamples blocks of random length. Since the average length of these blocks is $1/p$, it is expected that the quantity $1/p$ should play a similar role as the parameter b in the moving blocks method. In Figure 4.8, the $\hat{V}_{St.B.}$ variance estimates are pictured for different values of p between 0 and 1/2. Choosing $p = 0.05$ corresponds to a variance estimate of 2,335,502.

It should be stressed that all the abovementioned block-resampling methods give valid results for statistics that are smooth functions of sample means, since 'the bootstrap commutes with smooth functions' (cf. Bickel and Freedman (1981)). In addition, the moving blocks and the stationary bootstrap methods have been shown to work with statistics that are representable by smooth functionals, i.e. that can only be approximated by sample means (cf. Künsch(1989), Liu and Singh(1988), Politis and Romano(1991)), as is, for example, a trimmed-mean.

It should also be pointed out that the 'blocks of blocks' method is asymptotically correct in the general case where $\mu \in \mathbf{R}^D$, in which case $\bar{T}_N$ and the $T_{i,M,L}$'s are multivariate. Denote $\mu^{(n)}, \bar{T}_N^{(n)}$, etc. to be the nth coordinate of $\mu, \bar{T}_N$, and so forth. Then, for example, we can get approximations to the distribution of $\max_{n=1,2,\ldots,D}\{\sqrt{Q}|\bar{T}_N^{(n)} - \mu^{(n)}|\}$, which allows for the possibility of constructing simultaneous confidence intervals for all coordinates of μ (cf. Politis and Romano (1990)). In particular, in Politis, Romano, and Lai (1990), the case where μ is the spectral or cross-spectral density function sampled at a grid of points was studied, with the objective of setting uniform confidence bands.

As our final example, consider the important case where the parameter of interest is the autocorrelation coefficient at lag s, i.e. the parameter $\rho(s) = R(s)/R(0)$, where $R(s) = EX_0X_s$ and for simplicity it is assumed that $EX_0 = 0$. In that case, the linear statistic $\bar{T}_N$ is $(s+1)$-dimensional, with $\bar{T}_N^{(n)} = \frac{1}{N-s}\sum_{i=1}^{N-s} X_iX_{i+n-1}$, and $L = 1$, $M = s+1$ and $\phi_M^{(n)}(x_1, \ldots, x_M) = x_1x_n$, for $n = 1, \ldots, s+1$. It is easy to see that $\bar{T}_N^{(n)}$, for $n = s+1$, is just the sample autocovariance $\hat{R}(s)$ at lag s. By the 'blocks of blocks' resampling scheme applied to the linear statistic $\bar{T}_N$, accurate confidence intervals for the autocovariances can be obtained, as well as variance estimates for the sample autocovariances. Considering the complicated form of the asymptotic variance of the sample

autocovariances (that involves estimates of the fourth order cumulants, cf. Anderson(1971)), the advantage of using an automatic procedure like the bootstrap is apparent.

Now the estimator $\hat{\rho}(s) = \bar{T}_N^{(s+1)}/\bar{T}_N^{(1)}$ is a smooth function of the linear statistic $\bar{T}_N$, and its statistical properties can be analyzed via the 'blocks of blocks' bootstrap. The usefulness of considering the $(s+1)$-dimensional statistic $\bar{T}_N$ lies in that we can instantly obtain *simultaneous* confidence intervals (confidence band) for $\rho(k), k = 1, \ldots, s$ (and for $R(0)$), that are *not* available by classical methods (cf. Priestley(1981)). An obvious use of such confidence bands is in testing hypotheses regarding the covariance structure. The way this can be done is as follows. For concreteness, assume that we are looking for a 95% confidence band for $\rho(k), k = 1, \ldots, s$. That is, we are looking for two sequences $c_1(k), c_2(k)$ such that $P\{\forall k \in \{1, \ldots, s\} : \hat{\rho}(k) - c_1(k) \leq \rho(k) \leq \hat{\rho}(k) + c_2(k)\} = 0.95$. To start with, apply Fisher's z transformation to approximately stabilize the variance of the estimates at different lags, i.e. let $\zeta(k) = \frac{1}{2}\log\frac{1+\rho(k)}{1-\rho(k)}$, and $\hat{\zeta}(k) = \frac{1}{2}\log\frac{1+\hat{\rho}(k)}{1-\hat{\rho}(k)}$, for $k = 1, \ldots, s$. Then, by the 'blocks of blocks' bootstrap, obtain an approximation to the distribution of the 'maximum modulus' $\sqrt{N}\max_{k=1,\ldots,s}|\hat{\zeta}(k) - \zeta(k)|$. This immediately leads to a uniform width (i.e. $c_1(k) = c_1$ and $c_2(k) = c_2, k = 1, \ldots, s$) and symmetric (i.e. $c_1(k) = c_2(k)$) confidence band for $\zeta(k), k = 1, \ldots, s$, and can be translated to a confidence band (of non-uniform width) for $\rho(k), k = 1, \ldots, s$. Alternatively, we can get a (non-symmetric in general) equal-tailed uniform width confidence band for $\zeta(k), k = 1, \ldots, s$, by finding bootstrap approximations to x and y such that $P\{\sqrt{N}\max_{k=1,\ldots,s}(\hat{\zeta}(k) - \zeta(k)) \leq x\} = 0.975$, and $P\{\sqrt{N}\min_{k=1,\ldots,s}(\hat{\zeta}(k) - \zeta(k)) \leq y\} = 0.025$.

As an illustration, a time series $Y_1, \ldots, Y_{200}$ was generated according to the model

$$X_t - 1.352X_{t-1} + 1.338X_{t-2} - 0.662X_{t-3} + 0.240X_{t-4} = Z_t - 0.2Z_{t-1} + 0.04Z_{t-2}$$

and $Y_t = X_t|X_t|$, where the Z_t's are independent normal N(0,1) random variables. A plot of the $Y_1, \ldots, Y_{200}$ data set is presented in Figure 4.9, and Figure 4.10 contains a histogram of the empirical distribution of the data that reflects the lack of normality of the Y_t's.

In Figures 4.11 and 4.12, 95% equal-tailed confidence bands are set for $\zeta(k)$ and $\rho(k), k = 1, \ldots, 10$, via the 'blocks of blocks' bootstrap, with design parameters, $b = 15, k = 20$, and number of bootstrap replications $B = 500$. In both figures, the middle curve can be considered to be the 'true' values of $\zeta(k)$ and $\rho(k)$. These are really estimates of $\zeta(k)$ and $\rho(k)$ obtained by generating a Y_t time series of length 20,000. From this extra long stretch of the Y_t time series, one hundred approximately independent series of length 200 were extracted in order to get an empirical estimate of the distribution of the 'maximum modulus', to be compared to its bootstrap approximation.

In Figures 4.13 and 4.14, histograms of the bootstrap and the empirical estimates of the distribution of $\sqrt{N}\max_{k=1,\ldots,s}|\hat{\zeta}(k) - \zeta(k)|$ are pictured. Maybe due to the small number (one hundred) of available series, the empirical distribution has a shorter right tail than its bootstrap counterpart. This implies that the bootstrap confidence band would be more conservative (i.e. wider) than a confidence band based on this empirical distribution, provided the latter was somehow available.

5. Bootstrap and Adaptive Estimation.

5.1. Introduction

So far, the bootstrap has been used to estimate characteristics of the distribution of a statistic, such as its bias and its variance, or to construct confidence or prediction intervals for a parameter or future observation. In this section, we will use the bootstrap to *choose* an estimator among a family of estimators. Moreover, the possibility of constructing confidence intervals based on an estimator with an adaptively chosen parameter constitutes one of the principal strengths of the bootstrap over cross-validation as a method for choosing a tuning parameter.

Suppose that interest focuses on estimating a parameter $\theta(F)$, where F is the distribution of i.i.d. observations $X_1, X_2, \ldots, X_n$. Let $\{T_{n,\beta} : \beta \in I\}$ be a class of estimators for $\theta(F)$ indexed by the tuning parameter β. The goal consists of choosing $\hat{\beta}$ in such a way that $T_{n,\hat{\beta}}$ is best among the class. Therefore, we implicitly or explicitly have in mind a loss function and we want to select the value of β such that the expected loss, or risk, is minimized. For instance, the loss might be squared error loss, and so we want to choose the estimator which minimizes mean squared error within the family. In order to do that, a bootstrap estimate of risk is computed for each value of β, or a grid of such values, and β is taken as the value corresponding to the smallest bootstrap estimate of the risk. Léger and Romano (1990a) give numerous examples and general theorems on the validity of the bootstrap in the selection of a tuning parameter. In the following subsections, we will look at a number of examples, including some which are directly relevant to the practitioners of the physical, chemical, and engineering sciences.

5.2 Trimmed means

Léger and Romano (1990b) considered the problem of using the bootstrap to adaptively choose the trimming proportion in an adaptive trimmed mean. The (symmetric) α-trimmed mean is defined as follows:

$$T_{n,\alpha} = (n - 2[n\alpha])^{-1} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)},$$

where $\alpha \in [0, 1/2)$, $[\cdot]$ is the greatest-integer function, and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the order statistics. The choice of α is made as follows. For each value of α which trims an integer number of observations, compute a bootstrap estimate of variance. This computation is done via Monte Carlo. Then $\hat{\alpha}$ is the value of α corresponding to the smallest bootstrap estimate of variance. The estimate of location is $T_{n,\hat{\alpha}}$. Léger and Romano (1990b) show that this estimator is asymptotically normal with an asymptotic variance equal to the smallest among all trimmed means. In other words, the bootstrap adaptive trimmed mean does asymptotically as well as the best trimmed mean for that particular distribution.

It is also possible to construct a bootstrap confidence interval for the location parameter based on a bootstrap adaptive trimmed mean. This involves a double bootstrap: one level of bootstrapping to choose the trimming proportion and another level to construct the confidence interval. This bootstrap confidence interval, which reflects the adaptive nature of the estimator by rechoosing the trimming proportion for each bootstrap sample, outperforms a classical interval based on asymptotic normality and an estimate of the variance of the adaptive estimator. Small

sample simulations showed that the bootstrap adaptive trimmed means adapted themselves rather well to the different distributions, even for samples of size as small as 10.

5.3. Weighted Combination of Two Estimators: The Calibration Example

Léger and Romano (1990a) considered the problem of choosing the optimal weight in a weighted combination of two estimators. As an example of this problem, we use the calibration problem where Srivastava and Singh (1989) have recently introduced a weighted combination of the classical and inverse calibration estimators. Assume a simple linear regression experiment:

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n, \tag{5.1}$$

where the x's are fixed known values and the ϵ's are i.i.d. mean 0 variance σ^2 random variables. In this experiment, the exact x values are very expensive or difficult to obtain. Hence, the goal is to estimate an unknown value x_{new} which generated a new observed value y_{new}.

The classical estimator $\hat{x}_{new}$ is $(y_{new} - \hat{\alpha})/\hat{\beta}$ where $\hat{\alpha}$ and $\hat{\beta}$ are the usual least squares estimators of α and β. Krutchkoff (1967) introduced the inverse calibration estimator based on the relationship $X_i = \gamma + \delta y_i + \nu_i$, where $\gamma = -\alpha/\beta$, $\delta = 1/\beta$, and $\nu_i = -\epsilon_i/\beta$. Using least squares regression of the x's on the y's (even though the x's are not random and the y's are), compute $\hat{\gamma}$ and $\hat{\delta}$ and obtain the inverse calibration estimator $\tilde{x}_{new} = \hat{\gamma} + \hat{\delta} y_{new}$.

Much controversy persists on the "right" estimator for this problem. See Srivastava and Singh (1989) and the references there for details. The major difficulty is that the classical estimator does not have finite moments whereas the inverse calibration estimator is inconsistent as $n \to \infty$. Moreover, it is argued that a large sample theory is unrealistic since the calibration sample sizes are rarely large in practice. On the other hand, some argue that the fact that the classical estimator doesn't have finite sample moments (due to the presence of $\hat{\beta}$ in the denominator) should not really be a problem since, in practice, calibration would never be used in cases where $\hat{\beta}$ is close to 0. Srivastava and Singh (1989) used small-disturbance asymptotic theory to analyze the two estimators and introduced the following family of estimators:

$$x_\lambda = \lambda \tilde{x}_{new} + (1 - \lambda)\hat{x}_{new}, \qquad \lambda \in [0, 1].$$

Small-disturbance asymptotic theory assumes that the variance of te ϵ's in (5.1) goes to 0.

The bootstrap will be used to select λ adaptively on the basis of the regression data. To do this, we will construct bootstrap samples $y_1^*, \ldots, y_n^*$ from $\hat{\alpha}$ and $\hat{\beta}$ and the residuals, as described in the section on regression. Note that $x_1, \ldots, x_n$ remain fixed. From the bootstrap samples, we obtain $\hat{\alpha}^*$, $\hat{\beta}^*$, $\hat{\gamma}^*$, and $\hat{\delta}^*$ in order to compute $\hat{x}_{new}^*$ and $\tilde{x}_{new}^*$. Note that we also keep y_{new} fixed in these computations. Hence we are only taking into account the randomness in the calibration line and not the randomness in y_{new}. For each value of λ, we compute x_λ^* in order to compute bootstrap estimates of bias and variance. Choose $\hat{\lambda}$ as the value of λ corresponding to the smallest bootstrap estimate of mean squared error.

Srivastava and Singh (1989) carried their analysis on a real-world example taken from Dunsmore (1968). The two estimators are $\hat{x}_{new} = 8.8352$, whereas $\tilde{x}_{new} = 8.7985$. Note that this last value differs from the published value in Srivastava and Singh (1989). Figure 1 shows the

estimated mean squared error as a function of λ. The estimated mean squared error of the classical estimator is smaller than that of the inverse calibration estimator and is minimized by $\hat{\lambda} = 0.12$ and $x_{\hat{\lambda}} = 8.8308$. The estimated bias is minimized by taking $\lambda = 0.08$. Interestingly, this value is close to the value that minimizes bias according to the small-disturbance asymptotic theory, namely $1/(n-2) = 1/9$ in this example. On the other hand, the estimated values of mean squared error are 4 times larger than those suggested by the small-disturbance asymptotic theory. Obviously, the difference between $x_{\hat{\lambda}}$ and $\hat{x}_{new}$ is unlikely to be of importance in this data set. In fact, the difference between $\hat{x}_{new}$ and $\tilde{x}_{new}$ was quite small to begin with.

5.4. Nonparametric Age Replacement Policy

An important problem in reliability is determining a replacement policy. One of the older policies is the age replacement policy where replacements of a unit occur at failure or at age x_0, whichever occurs first. Barlow and Proschan (1965) discuss it in detail and give an extensive bibliography. The problem consists of determining the age x_0 in view of the fact that the cost c_1 incurred in replacing a failed unit is larger than the cost c_2 of making a preventive replacement.

The optimal x_0 is a function of F, the lifetime distribution of the unit. It is known from renewal theory that the average cost per time unit, over an infinite horizon, for the policy with preventive replacement at time x is given by

$$A(x, F) = \frac{c_1 F(x) + c_2[1 - F(x)]}{\int_0^x [1 - F(u)]du}, \tag{5.2}$$

provided that $F(0) = 0$, F has a finite mean and has an increasing failure rate. The optimal age is then the value $x_0(F)$ minimizing $A(x, F)$. Ingram and Schaeffer (1976) considered parametric models whereas Arunkumar (1972) considered a nonparametric approach. In this approach, he simply replaced F by $\hat{F}_n$ in (5.2) where $\hat{F}_n$ is a sample size n from F and computed $x_0(\hat{F}_n)$, the value that minimizes $A(x, \hat{F}_n)$. He showed that $x_0(\hat{F}_n)$ and $A(x_0(\hat{F}_n), \hat{F}_n)$, are consistent for $x_0(F)$ and $A(x_0(F), F)$, respectively. Moreover he shows that the asymptotic distribution of $x_0(\hat{F}_n)$ is related to the solution of the heat equation.

Léger and Cléroux (1990) generalize Arunkumar's results by obtaining the asymptotic distribution of $A(x_0(\hat{F}_n), \hat{F}_n)$ and by constructing bootstrap confidence intervals for the actual cost $A(x_0(\hat{F}_n), F)$ of using the policy $x_0(\hat{F}_n)$, Arunkumar's estimate of the optimal policy.

Since $n^{1/2}[A(x_0(\hat{F}_n), \hat{F}_n) - A(x_0(\hat{F}_n), F)]$ is asymptotically normal with an asymptotic variance $\sigma^2(x_0(F), F)$ an explicit function of F, one may construct an approximate confidence interval using the normal quantiles and $\sigma(x_0(\hat{F}_n), \hat{F}_n)$ as an estimate of standard deviation. See Léger and Cléroux (1990) for the explicit form of $\sigma^2(x, F)$. Better still, one can compute a bootstrap-t interval as follows. First, compute $x_0(\hat{F}_n)$, the value of x which minimizes $A(x, \hat{F}_n)$, $A(x_0(\hat{F}_n), \hat{F}_n)$, and $\sigma(x_0(\hat{F}_n), \hat{F}_n)$. Then for each bootstrap sample with empirical distribution function $\hat{G}_n$, compute $[A(x_0(\hat{G}_n), \hat{G}_n) - A(x_0(\hat{G}_n), \hat{F}_n)]/\sigma(x_0(\hat{F}_n), \hat{F}_n)$. Repeating this operation B times and letting $\eta_{.05}$ and $\eta_{.95}$ be the 5% and 95% percentiles, we obtain the following bootstrap-t interval

$$[A(x_0(\hat{F}_n), \hat{F}_n) - \sigma(x_0(\hat{F}_n), \hat{F}_n)\eta_{.95}, A(x_0(\hat{F}_n), \hat{F}_n) - \sigma(x_0(\hat{F}_n), \hat{F}_n)\eta_{.05}].$$

Léger and Cléroux (1990) show that the bootstrap intervals and the classical intervals based on the asymptotic normality all have the claimed coverage probability asymptotically. Even though,

it was not shown that the bootstrap-t intervals of this problem are second-order correct, small sample simulations indicate that it is likely to be true. For instance, with a sample of size 20 from a Gamma distribution with mean 9080 and standard deviation 3027, and costs $c_1 = 1100$ and $c_2 = 100$, the estimated coverage of the classical, bootstrap pivotal, and bootstrap-t one-sided intervals were 0.9068, 0.9150, and 0.9453 for a nominal probability of 0.95. Similar results were obtained for the 5% level and for Weibull and truncated normal distributions. These results are based on 5000 simulations and 1000 bootstrap samples for each simulation.

Arunkumar's estimator has a major drawback, namely the need for a sample of size n from the distribution F of the lifetime of a unit. In reliability problems, we may not have the luxury to wait until n items eventually fail, especially in systems where we may have to wait until an item fails before studying another one. Moreover, such an experiment may be costly. Hence, it may be desirable to use a sequential policy. Bather (1977) introduced two such sequential policies and Léger and Cléroux (1990) have shown how the bootstrap can be used to construct confidence intervals based on these sequential estimators. Many other policies from reliability theory could use the bootstrap paradigm to obtain variance estimates or confidence intervals.

5.5 Nonparametric Regression and Density Estimation

In the last few years, researchers have paid a lot of attention to the problems of nonparametric regression and nonparametric density estimation. Most of the efforts have centered on an automatic choice of the bandwidth. The bootstrap has recently been applied to these problems with success, but the solutions demonstrate that care must be taken to effectively use the bootstrap in complex problems. Since the two problems are very similar from a theoretical point of view, we will illustrate the key points of the use of the bootstrap in nonparametric density estimation.

Let $X_1, X_2, \ldots, X_n$ be a sample size n of i.i.d observations from density f. The kernel density estimator is given by $\hat{f}_\alpha(x) = (n\alpha)^{-1} \sum_{i=1}^{n} K[(x - X_i)/\alpha]$, where K is a kernel function and $\alpha \equiv \alpha_n$ is the bandwidth parameter. The choice of bandwidth is crucial: a large bandwidth reduces variance while increasing bias whereas a small bandwidth increases variance and reduces bias; see Silverman (1986) and Devroye (1987). The goal of the estimation procedure is usually to minimize a criterion such as the expected loss at a point, $E\, l(\hat{f}_\alpha(x_0), f(x_0))$, or the expected integrated loss, $E \int l(\hat{f}_\alpha(x), f(x))dx$, where l is a loss function. Usually, researchers use squared error loss for its mathematical tractability. Nevertheless, the bootstrap can be used with other loss functions, including the L_1 loss.

The bandwidth can be chosen adaptively by estimating these criteria, using the bootstrap, by taking the bandwidth corresponding to the smallest such estimate. As an illustration, we use the criterion of expected integrated loss. The bootstrap estimate of the criterion is $\int l(\hat{f}^*_\alpha(x), \hat{g}(x))dx$, where $\hat{f}^*_\alpha(x)$ is the kernel estimate of an i.i.d. bootstrap sample from the distribution with density $\hat{g}$, an estimate of f. Different bootstrap procedures have been introduced corresponding to different choices of $\hat{g}$. Note that the role of $\hat{g}$ is two-fold: as a replacement for f in computing the loss and as the distribution from which bootstrap samples will be generated.

Noting that the criterion to be estimated is a functional of the distribution function F, the bootstrap idea suggests replacing F by $\hat{F}_n$, the empirical distribution function. Unfortunately, this doesn't work since $\hat{F}_n$ doesn't have a density. In the case of squared error loss, this problem could be

bypassed by estimating the variance and the bias by the bootstrap method. But $E_* \hat{f}^*_\alpha(x) = \hat{f}_\alpha(x)$, due to the linearity in the data, leading to a bias estimate of 0. Hence another estimate of F must be considered. The next obvious choice is to use the distribution with density $\hat{f}_\alpha(x)$, which is a smoothed version of the empirical distribution function; Taylor (1989) used this choice. Taylor (1989) noted that his bootstrap estimate of the squared error loss criterion also has a non-negligible bias component. He therefore explicitly corrected the bias using a leave-one-out idea. His resulting criterion is a modified bootstrap which combines bootstrap and cross-validation ideas. This type of bias in the application of the bootstrap had already been reported by Romano (1988b) in the context of bootstrapping kernel density estimates of the mode. His solution to the bias problem is to use an oversmooth density to generate the bootstrap data. Léger and Romano (1990a) obtain similar results in the context of choosing an optimal bandwidth. They use $\hat{f}_\beta$ as the density $\hat{g}$. The bootstrap is shown to work asymptotically provided that $\alpha_n = cn^{-1/5}$, i.e., α converges at the optimal rate for estimating f, whereas $n\beta_n^5/log(n) \to \infty$. Hence β must be larger than α and so the distribution of the bootstrap observations must be oversmooth. Faraway and Jhun (1990) also considered using a bandwidth different from α for the density $\hat{g}$. Specifically, they use the bandwidth selected by cross-validation. Their simulations indicate that a bootstrap choice of the bandwidth leads to a better density estimate than a cross-validated choice. Hall (1990) introduced yet another bootstrap method in nonparametric problems, including density estimation. His method is based on sampling from the empirical distribution function $\hat{F}_n$. He succeeds by using a smaller bootstrap sample size $n_1 < n$ and arguing that a smaller bandwidth should be used in the bootstrap estimate to go along with the smaller bootstrap sample size. So, he tries to find the best bandwidth on the scale of n_1 and then makes the adjustment to the scale of n whereas Léger and Romano (1990a) work directly on the scale of n.

The same type of problems are found in the use of the bootstrap in nonparametric regression. First, residuals are computed from one fit and centered about its mean. Then bootstrap observations are constructed by adding i.i.d. observations from the empirical distribution function of the centered residuals to an estimate of the smooth curve m. Different bootstrap methods are obtained by using different estimates of m to construct the bootstrap observations. Härdle and Bowman (1988) used the same bandwidth in the construction of the bootstrap observations as is used for the estimation step. As in density estimation, this leads to a non-negligible bias that was explicitly estimated and corrected. Faraway (1990) uses a different bandwidth to construct the bootstrap observations, chosen by cross-validation. Also, Hall (1990) applies his scheme of reducing the bootstrap sample size and the corresponding bandwidth to nonparametric regression.

All authors agree on the great potential of the bootstrap in nonparametric problems that require the choice of a bandwidth parameter. This is in part due to the fact that the bootstrap can easily be applied to losses other than squared error, but most importantly, to the possibility of constructing nonparametric confidence intervals and bands. But as this discussion has demonstrated, there is still a lot of work that needs to be done to carefully apply the bootstrap in these and other complex problems.

References

Anderson, T.W. (1971). *The Statistical Analysis of Time Series*, John Wiley.

Arunkumar, S. (1972). Nonparametric age replacement policy, *Sankhyā*, Series A, **34**, 251-256.

Bai, C. and Olshen, R. (1988). Discussion of Hall's Comparison of Bootstrap Confidence Intervals. *Ann. Statist.*, **16**, 953-956.

Barlow, R.E., and Proschan, F. (1965). *Mathematical Theory of Reliability*, New York: John Wiley.

Basu, A. P. (1981). The estimation of $P(X < Y)$ for distributions useful in life testing. *Naval Logistics Quarterly*, **28**, 383-392.

Bather, J. A. (1977). On the sequential construction of an optimal age replacement policy, *Bulletin of the International Statistical Institute*, **47**, 253-266.

Becker, R.A., Chambers, J.M., Wilks, A.R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole.

Beran, R. (1986). Simulated power functions. *Ann. Statist.*, **14**, 151-173.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, **74**, 457-468.

Beran, R. (1990). Calibrating prediction regions. *J. Amer. Statist. Ass.* **85**, 715-723.

Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, 1196-1217.

Bickel, P.J. and Freedman, D.A. (1983). Bootstrapping regression models with many parameters, in *A festschrift for Erich Lehmann*, eds. P.J. Bickel, K. Doksum, J.L. Hodges, Belmont: Wadsworth, pp. 28-48.

Birnbaum, Z. (1956). On a use of the Mann-Whitney statistic. In *Proceeding of the Third Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1)*: University of California Press, 13-17.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions, *Ann. Statist.*, **16**, 1709-1722.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Brillinger, D.R. (1975). *Time Series, Data Analysis and Theory*, Holt, Rinehart and Winston, New York.

Cambell, M.J. and Walker, A.M. (1977). A survey of statistical work on the MacKenzie River series of annual Canadian lynx trappings for the years 1821 - 1934 and a new analysis, *J. Roy. Statist. Soc. (A)*, **140**, 411-431.

Constantine, K., Karson, M., and Tse, S. (1989). Bootstrapping estimators of $P(X < Y)$ in the gamma case. *J. Statist. Comp. Sim.*, **33**, 217-231.

Delaney, N.J. and Chatterjee, S. (1986). Use of the bootstrap and cross-validation in ridge regression, *Journal of Business & Economic Statistics*, 4, 255-262.

Devroye, L. (1987). *A course in density estimation*, Boston: Birkhäuser.

DiCiccio, T. and Romano, J. (1988). A review of bootstrap confidence intervals (with discussion). *J. Roy. Statist. Soc, B*, **50**, 338-370.

Dielman, T.E. and Pfaffenberger, R.C. (1988). Bootstrapping in least absolute regression: An application to hypothesis testing, *Communications in Statistics - Simulation and Computation*,

17, 843-856.

Dunsmore, I.R. (1968). A bayesian approach to calibration, *Journal of the Royal Statistical Society, Ser. B*, **30**, 396-405.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, **7**, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*, SIAM NSF-CBMS, Monograph 38.

Efron, B. (1987). Better bootstrap confidence intervals, *J. Amer. Statist. Assoc.*, **82**, 171-185.

Efron, B. and Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statist. Sci.*, **1**, 54-77.

Faraway, J.J. (1990). Bootstrap selection of bandwidth and confidence bands for nonparametric regression, *Journal of Statistical Computation and Simulation*, **37**, 37-44.

Faraway, J.J. and Jhun, M. (1990). Boostrap choice of bandwidth for density estimation, *Journal of the American Statistical Association*, **85**, 1119-1122.

Franke, J. and Härdle, W. (1987). On bootstrapping kernel spectral estimates, Preprint (to appear in *Ann. Statist.*).

Freedman, D.A. (1981), Bootstrapping regression models, *Ann. Statist.*, **9**, 1218-1228.

Freedman, D.A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models, *Annals of Statistics*, **12**, 827-842.

Freedman, D.A. and Peters, S.C. (1984). Bootstrapping a regression equation: Some empirical results, *Journal of the American Statistical Association*, **79**, 97-106.

Friedman, J. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.*, **19**, 1-141.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *Ann. Statist.*, **16**, 927-953.

Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems, *Probability Theory and Related Fields*, **81**, 247-273.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis*, **32**, 177-203.

Hall, P., DiCiccio, T. and Romano, J. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17**, 692-704.

Hall, P. and Martin, M. (1988). On bootstrap resampling and iteration. *Biometrika*, **75**, 661-672.

Hall, P. and Pittelkow, Y.E. (1990). Simultaneous bootstrap confidence bands in regression, *Journal of Statistical Computation and Simulation*, **37**, 99-113.

Hand, M.L. (1986). Bootstrap resampling procedures for regression models under non-least-squares estimation criteria, in *Proceedings of the Statistical Computing Section of the American Statistical Association*, 172-176.

Härdle, W. and Bowman, A. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands, *J. Amer. Statist. Assoc.*, **83**, 100-110.

Hinkley, D. (1988). Bootstrap methods. *J. Roy. Statist. Soc., B*, **50**, 321-337.

Ingram, C.R. and Scheaffer, R.L. (1976). On consistent estimation of age replacement intervals, *Technometrics*, **18**, 213-219.

Krutchkoff, R.G. (1967). Classical and inverse regression methods of calibration, *Technometrics*, **9**, 425-440.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, **17**, 1217-1241.

Lahiri, S.N. (1990). Second order optimality of stationary bootstrap, *Proc. of 21/4th IMS Meeting*, East Lansing, Michigan. (Also: Technical Report No. 90-1, Department of Statistics, Iowa State University.)

Léger, C. and Cléroux, R. (1990). Nonparametric age replacement: Bootstrap confidence interval for the optimal cost, Publication 731, Université de Montréal, Département d'informatique et de recherche opérationnelle.

Léger, C. and Romano, J.P. (1990a). Bootstrap choice of tuning parameters, *Annals of the Institute of Statistical Mathematics*, 4, 709-735.

Léger, C. and Romano, J.P. (1990b). Bootstrap adaptive estimation: The trimmed mean example, *Canadian Journal of Statistics*, **18**, 297-314.

Lewis, P. and Stevens, J.C. (1990). Using MARS to investigate non-linear time series, *Proc. Interface '90 Symposium*, East Lansing, Michigan.

Liu, R.Y. and Singh, K. (1988). Moving blocks jackknife and bootstrap capture weak dependence, Unpublished manuscript, Department of Statistics, Rutgers University.

Loh, W. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Ass.*, **82**, 155-162.

Martin, M. (1990). On bootstrap iteration for coverage correction in confidence intervals. *J. Amer. Statist. Ass.* **85**, 1105-1118.

Miller, R.G. (1986). *Beyond ANOVA, Basics of Applied Statistics*, John Wiley.

Morey, M.J. and Schenck, L.M. (1984). Small sample behavior of bootstrapped and jackknifed regression estimators, in *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 437-442.

Odeh, R. E. (1990). Two-sided prediction intervals to contain at least k out of m future observations from a normal distribution. *Technometrics*, **32**, 203-216.

Politis, D.N. and Romano, J.P. (1989). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, Technical Report No. 338, Department of Statistics, Stanford University.

Politis, D.N. and Romano, J.P. (1990). A nonparametric resampling procedure for multivariate confidence regions in time series analysis, Proceedings of INTERFACE '90, 22nd Symposium on the Interface of Computing Science and Statistics.

Politis, D.N. and Romano, J.P. (1991). The stationary bootstrap, Technical Report No. 365, Department of Statistics, Stanford University.

Politis, D.N., Romano, J.P., and Lai, T.-L. (1990). Bootstrap confidence bands for spectra and cross-spectra, Technical Report No. 342, Department of Statistics, Stanford University.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press.

Ramos, E. (1988). Resampling methods for time series, Technical Report No. ONR-C-2, Department of Statistics, Harvard University.

Romano, J.P. (1988a). A bootstrap revival of some nonparametric distance tests. *J. Amer. Statist. Ass.* **83**, 698-708.

Romano, J.P. (1988b). On weak convergence and optimality of kernel density estimates of the mode, *Annals of Statistics*, **16**, 629-647.

Romano, J.P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141-159.

Romano, J.P. (1990). On the behavior of randomization tests without a group invariance assumption. *J. Amer. Statist. Ass.* **85**, 686-692.

Shorack, G.R. (1982). Bootstrapping robust regression, *Communications of Statistics - Theory and Methods*, **11**, 961-972.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, New York: Chapman & Hall.

Silverman, B. and Young, G. (1987). The bootstrap: to smooth or not to smooth, *Biometrika*, **74**, 469-480.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.*, **9**, 11887-1195.

Srivastava, V.K. and Singh, N. (1989). Small-disturbance asymptotic theory for linear-calibration estimators, *Technometrics*, **31**, 373-378.

Stangenhaus, G. (1987). Bootstrap and inference procedures for L_1 regression, in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y. Dodge, Amsterdam: Elsevier Science Publishers (North Holland).

Stine, R.A. (1985). Bootstrap prediction intervals for regression, *Journal of the American Statistical Association*, **80**, 1026-1031.

Subba Rao, T. and Gabr, M.M. (1980). A test for linearity of stationary time series, *J. Time Ser. Anal.*, **1**, 145-158.

Swanepoel, J.W.H. and van Wyk, J.W.J. (1986). The bootstrap applied to spectral density function estimation, *Biometrika* **73**, 135-142.

Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation, *Biometrika*, **76**, 705-712.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Springer Lecture Notes in Statistics, 21.

Weber, N.C. (1984). On resampling techniques for regression models, *Statistics & Probability Letters*, **2**, 275-278.

Weisberg, S. (1985). *Applied Linear Regression* (2nd ed.), New York: John Wiley.

World Almanac and Book of Facts (1989), New York: Pharaos Books.

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics*, **14**, 1261-1295.

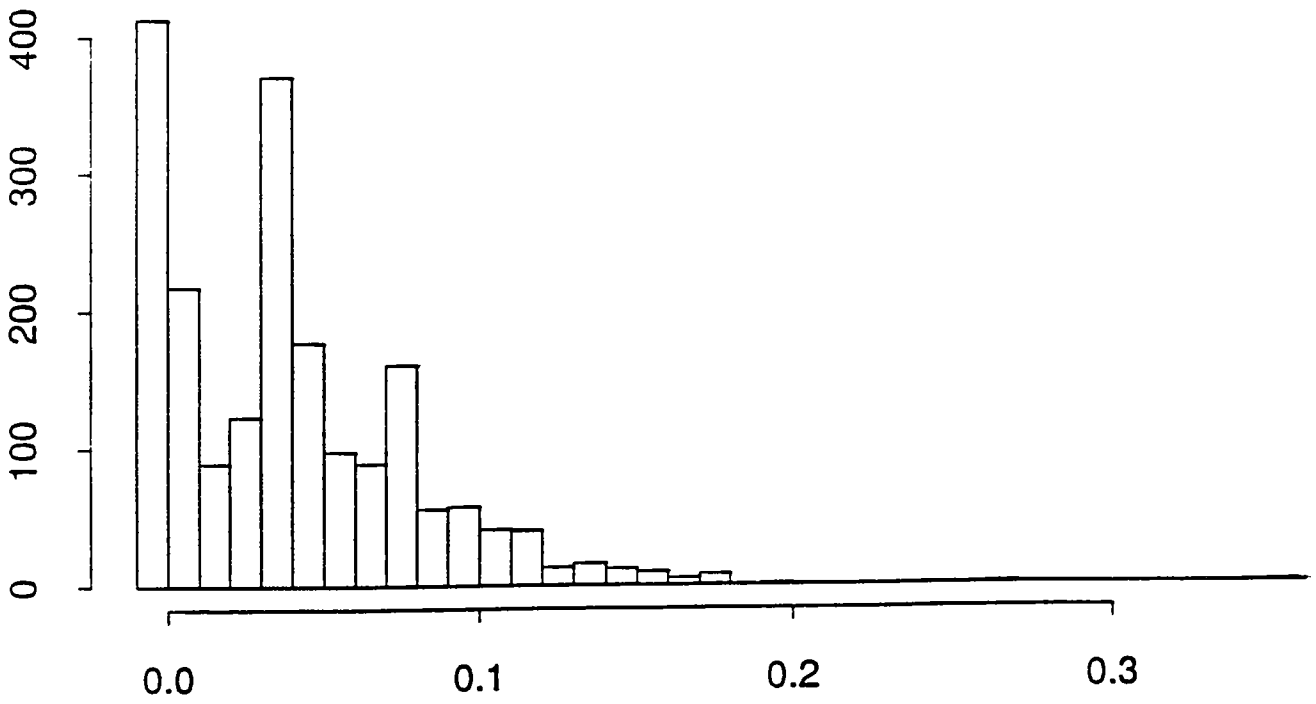


FIGURE 2.1. Bootstrap histogram of Mann-Whitney statistic

Estimator	Variables				
	Intercept	Tax	Dlic	Inc	Road
Least Squares Coef.	235.5	-8.2	11.9	-68.8	2.9
Classical 95% CI	[-223.9,695.0]	[-37.6,21.2]	[7.2,16.6]	[-111.5,-26.2]	[-5.1,10.8]
Bootstrap 95% CI	[-188.4,684.4]	[-35.8,19.4]	[7.1,16.2]	[-110.5,-29.5]	[-4.3,10.5]
LS Coef. w/o Hawaii	479.8	-40.0	11.5	-53.6	-4.1
Classical 95% CI	[88.5,871.1]	[-67.6,-12.3]	[7.6,15.4]	[-89.3,-18.0]	[-11.3,3.1]
Bootstrap 95% CI	[112.3,834.4]	[-65.5,-14.2]	[8.1,15.2]	[-84.1,-23.5]	[-10.6,2.0]
Robust Coef.	503.3	-24.8	9.6	-69.2	-0.5
Bootstrap 95% CI	[210.3,786.2]	[-43.3,-4.8]	[6.8,12.6]	[-96.4,-42.0]	[-5.9,4.2]

Table 3.1: Estimates and confidence intervals for the fuel data

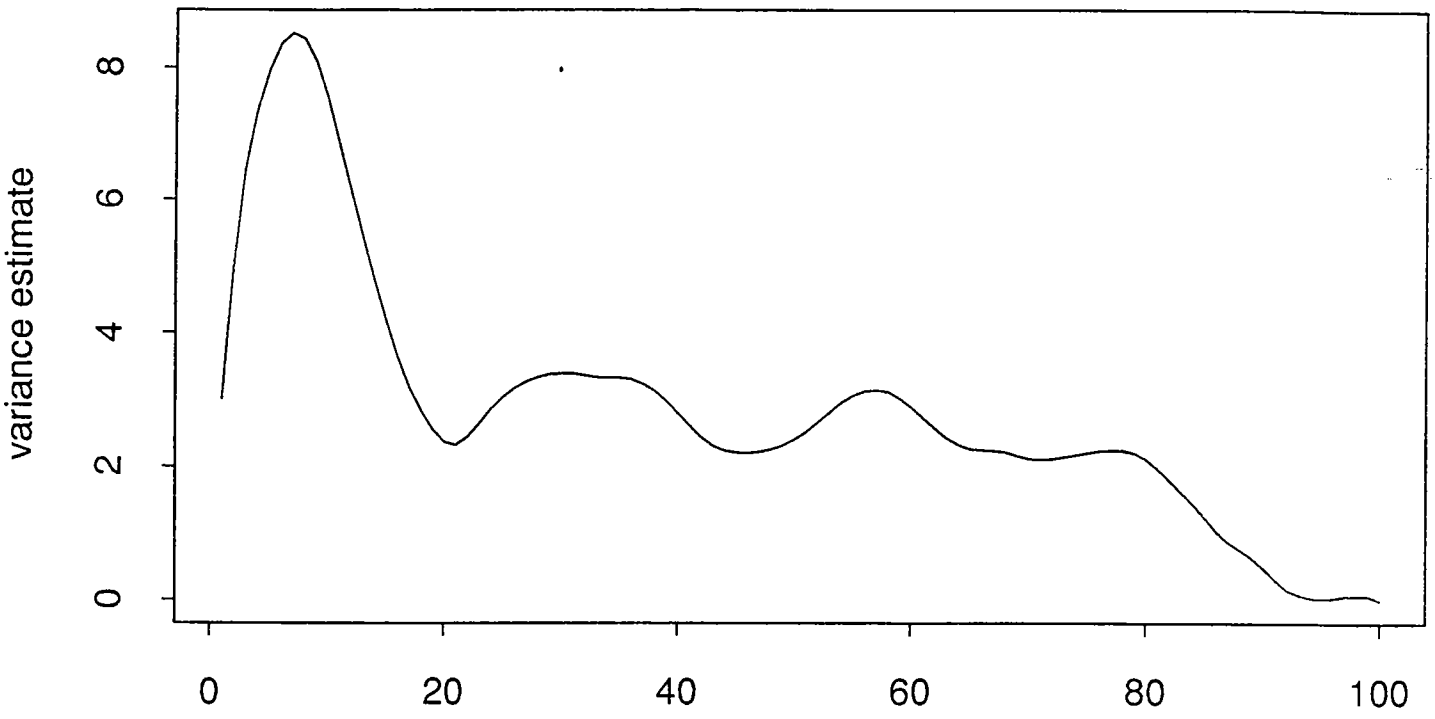


FIGURE 4.1. The moving blocks bootstrap variance estimate

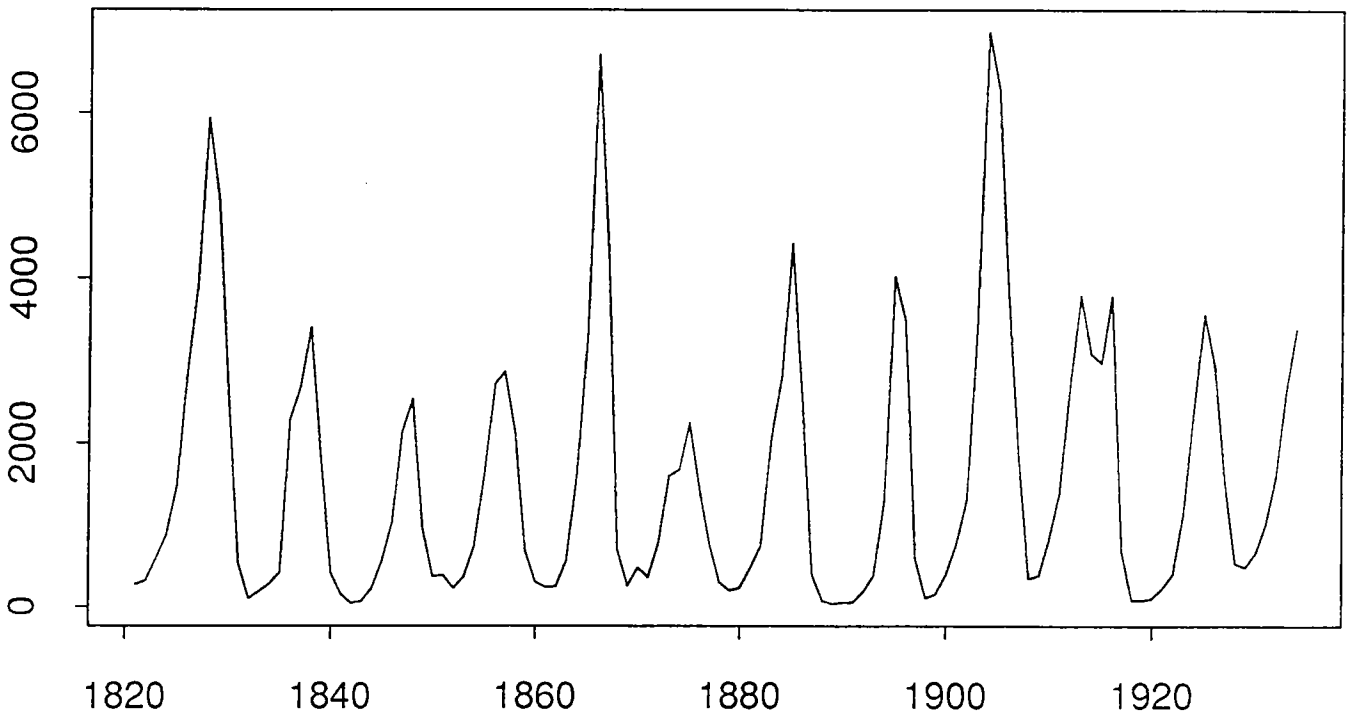


FIGURE 4.2. The time series of annual number of lynx trappings

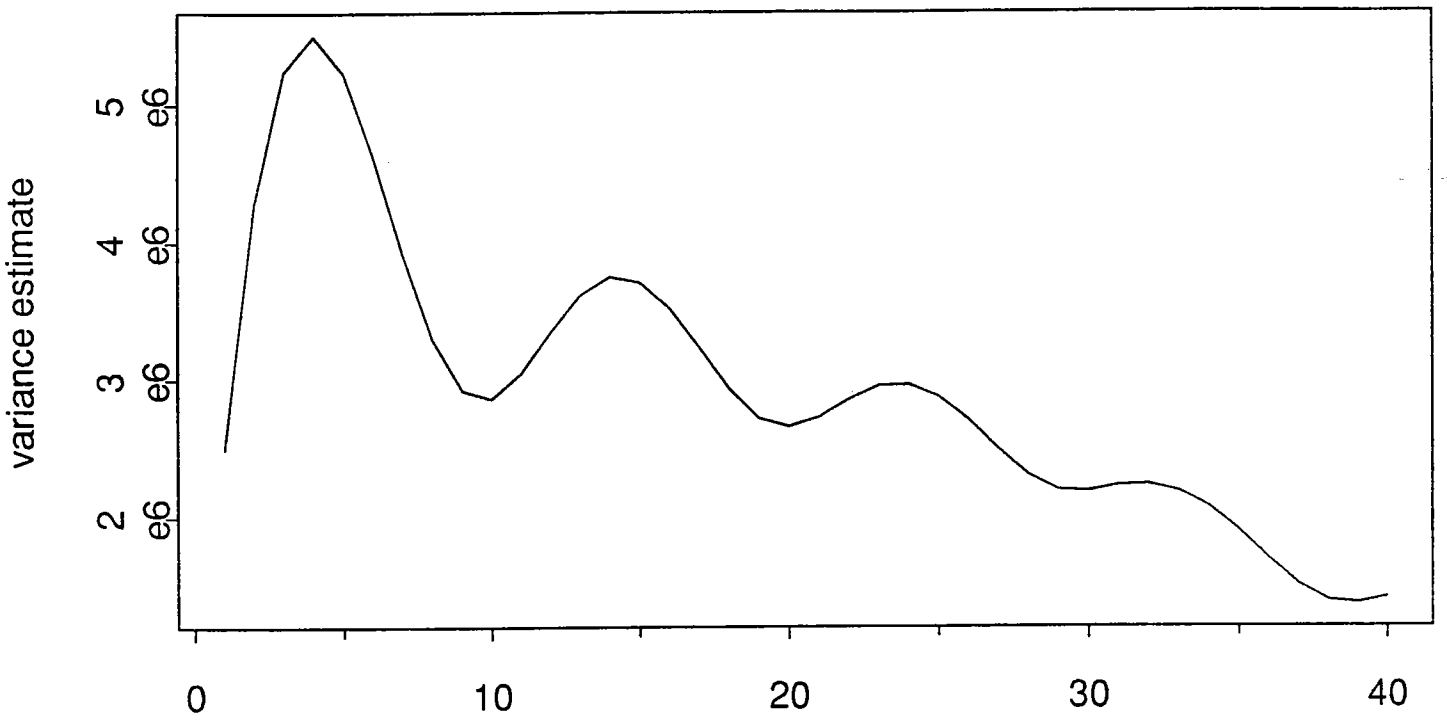


FIGURE 4.3. The moving blocks bootstrap variance estimate

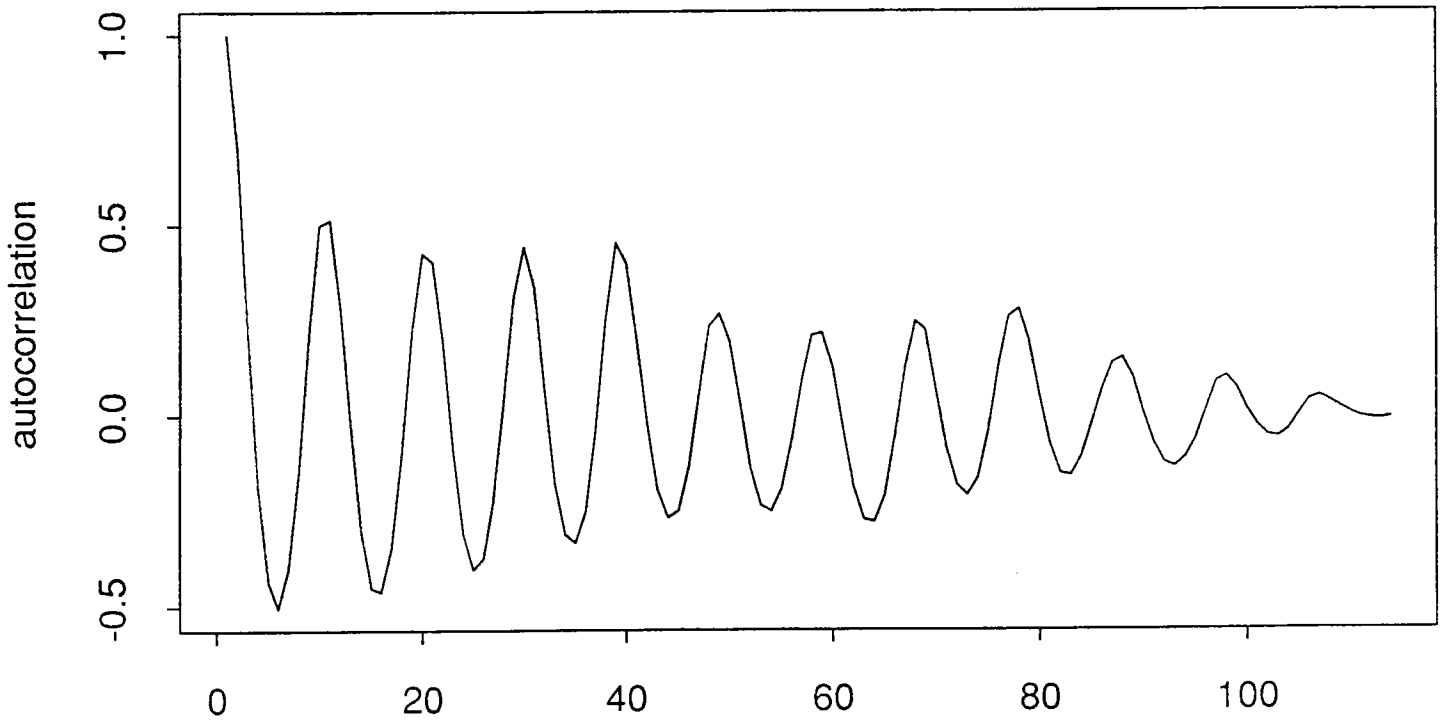


FIGURE 4.4. Sample autocorrelation sequence of Lynx data

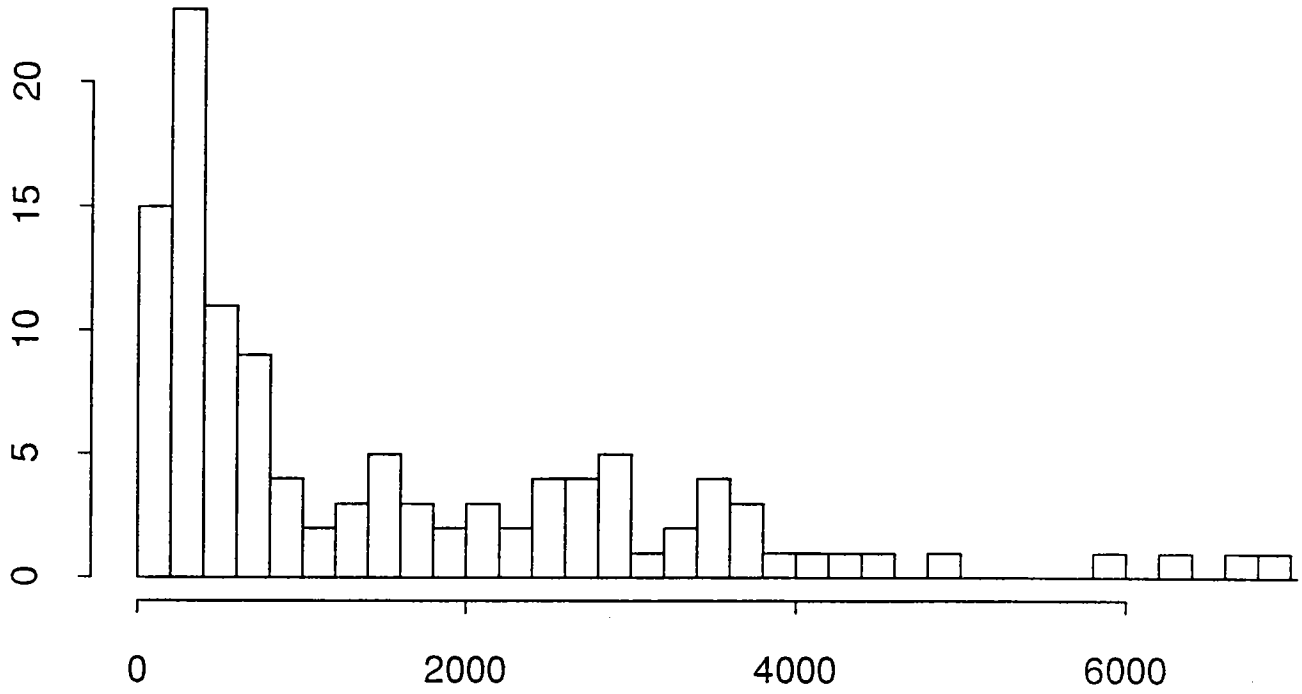


FIGURE 4.5. Histogram of Lynx data

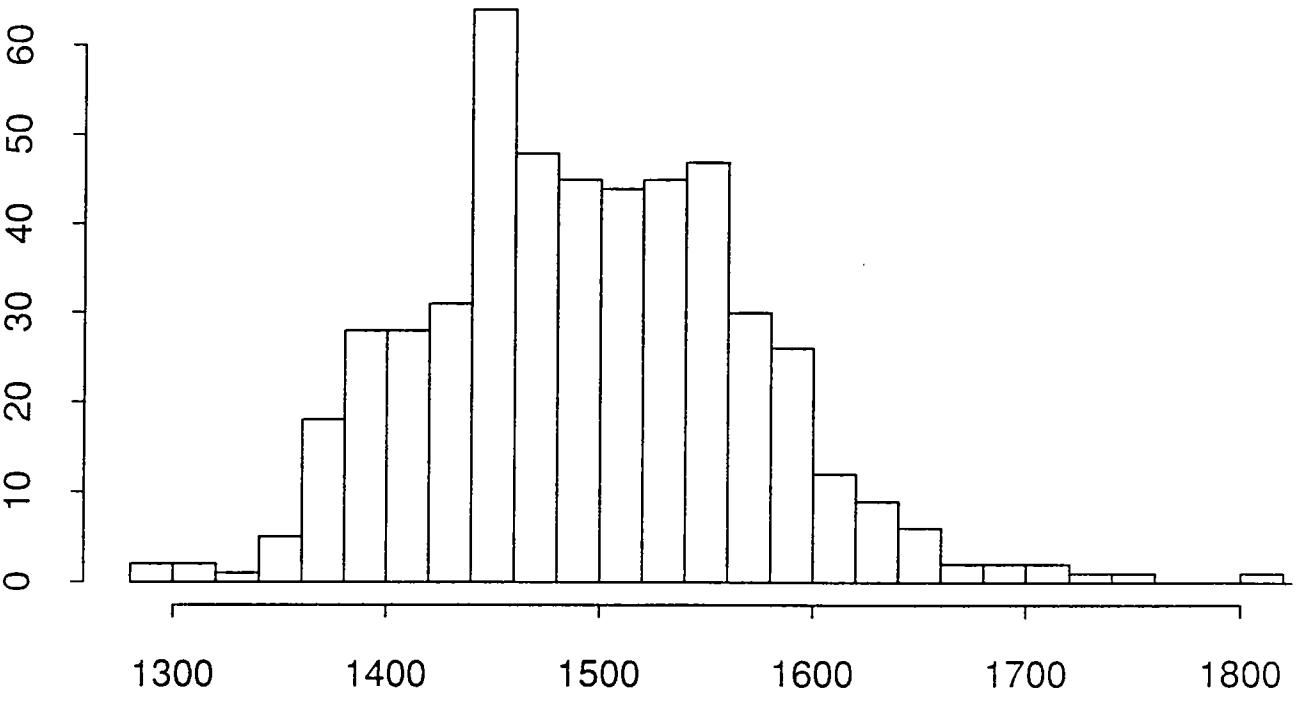


FIGURE 4.6. Bootstrap distribution of Lynx data sample mean, 500 replications

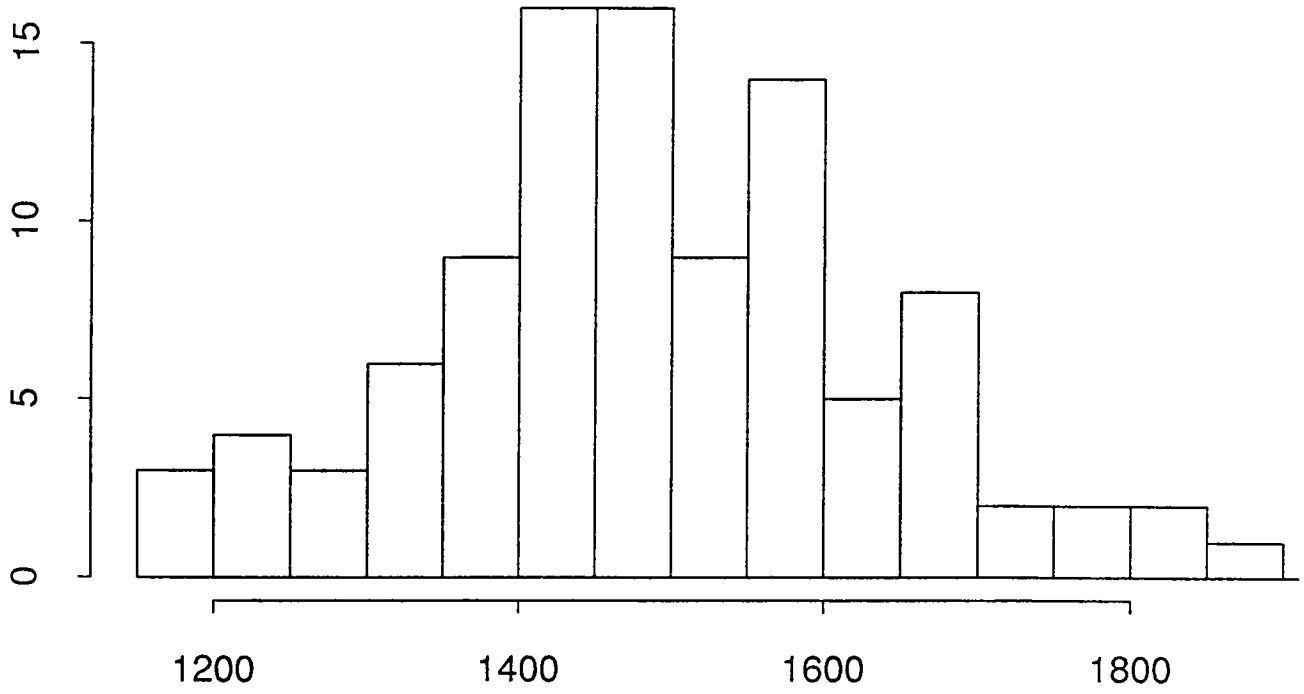


FIGURE 4.7. Bootstrap distribution of Lynx data sample mean, 100 replications

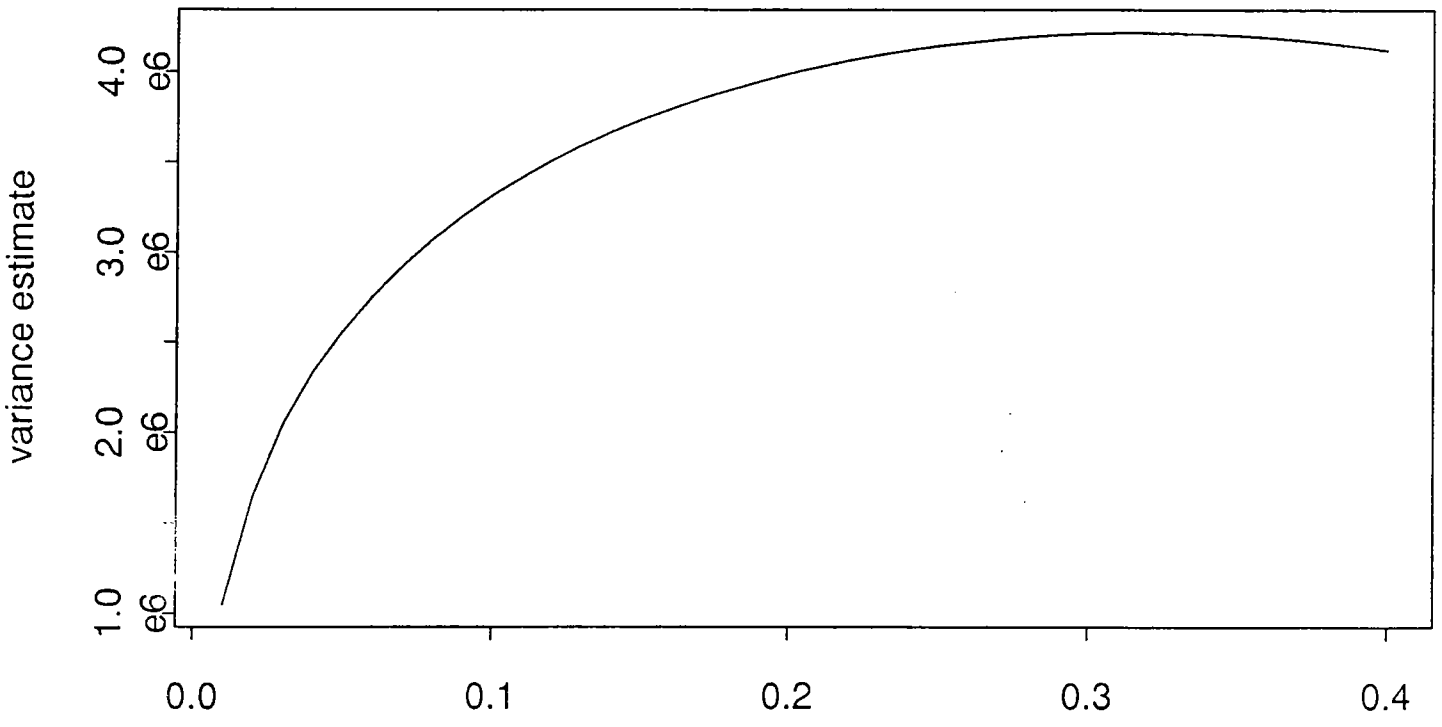


FIGURE 4.8. The stationary bootstrap variance estimate

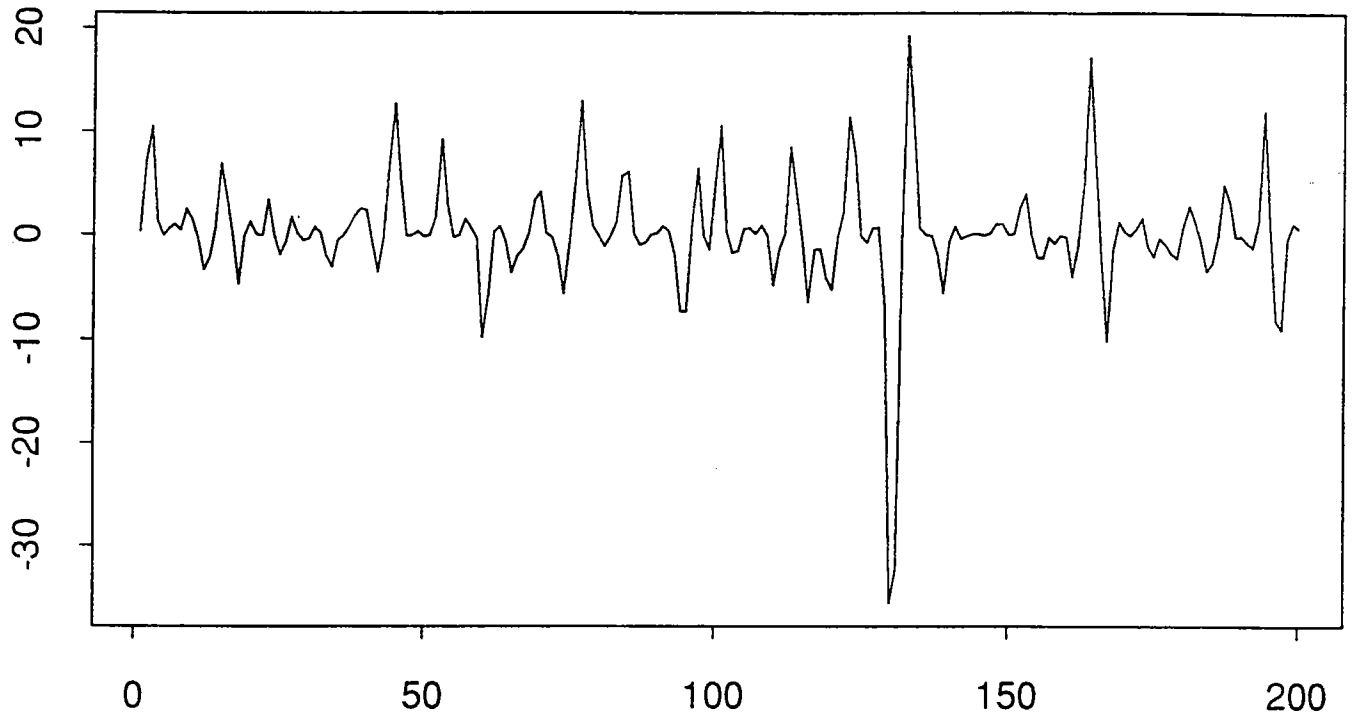


FIGURE 4.9. The time series Y

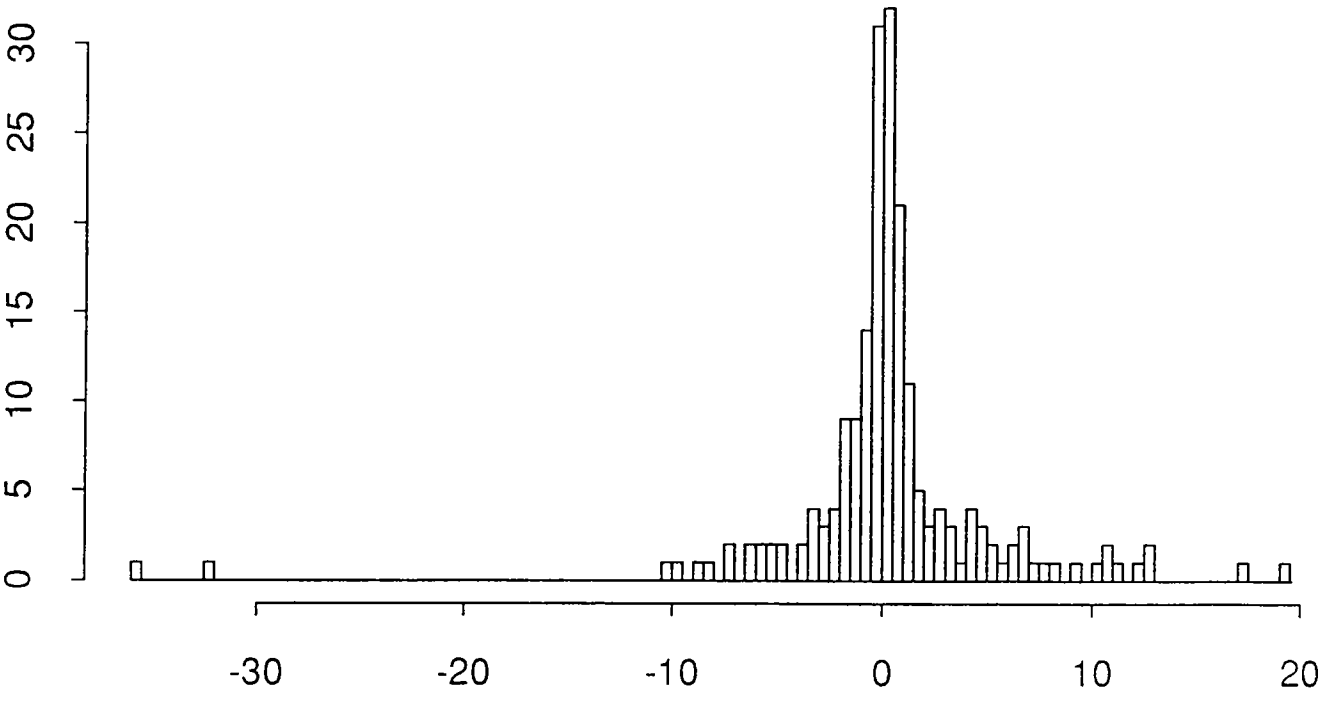


FIGURE 4.10. Histogram of the Y data

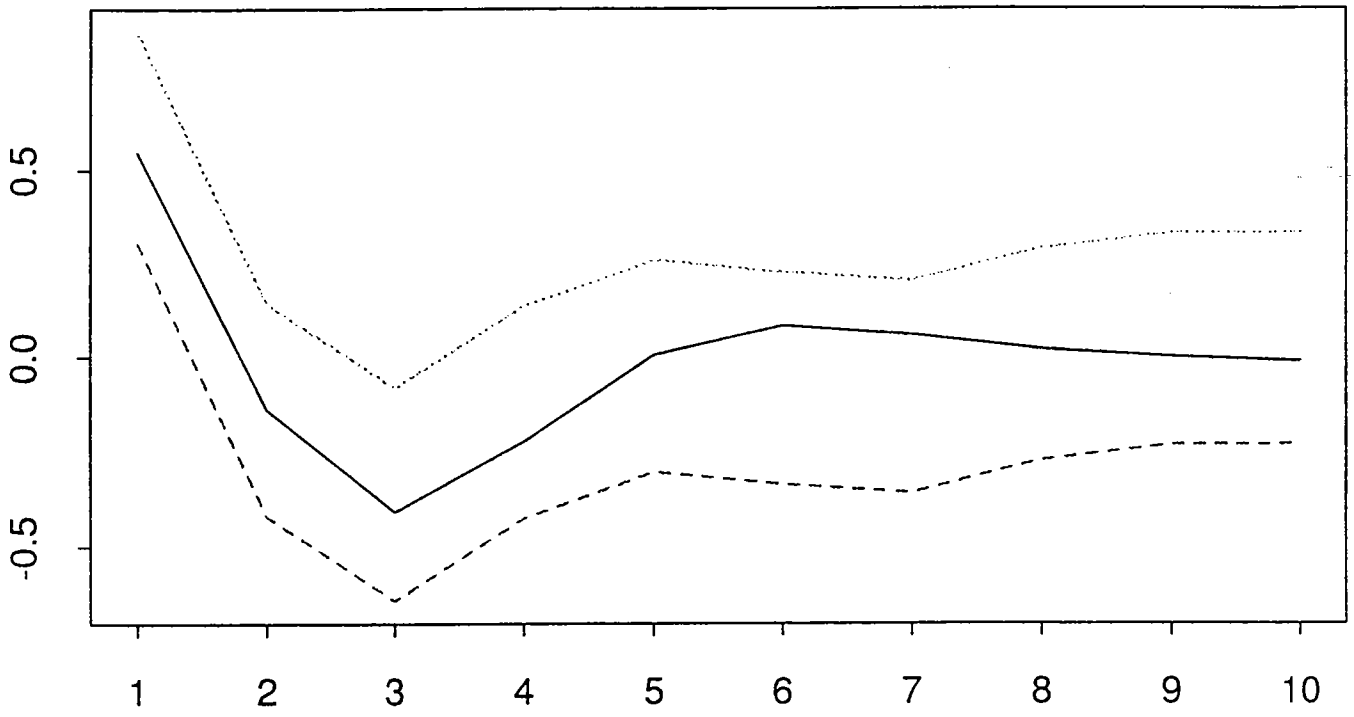


FIGURE 4.11. 95% confidence band for the trasformed autocorrelations

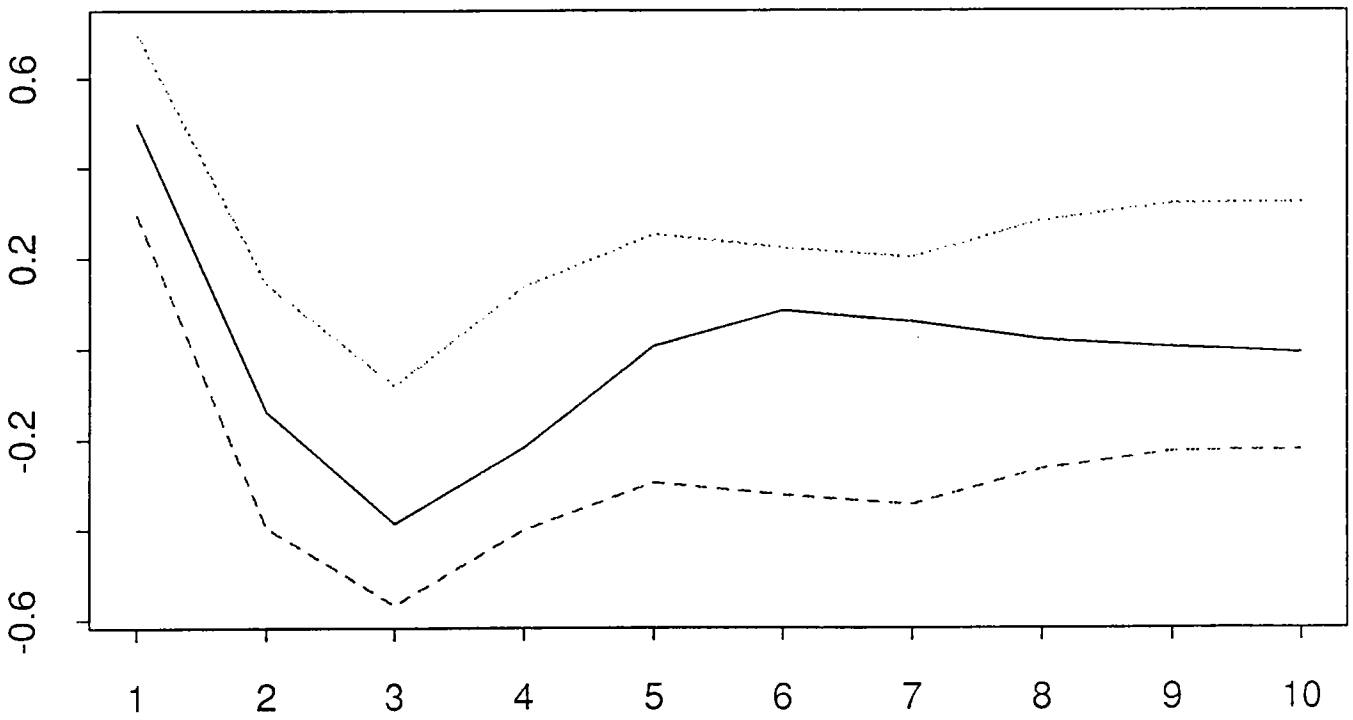


FIGURE 4.12. 95% confidence band for the autocorrelation coefficients

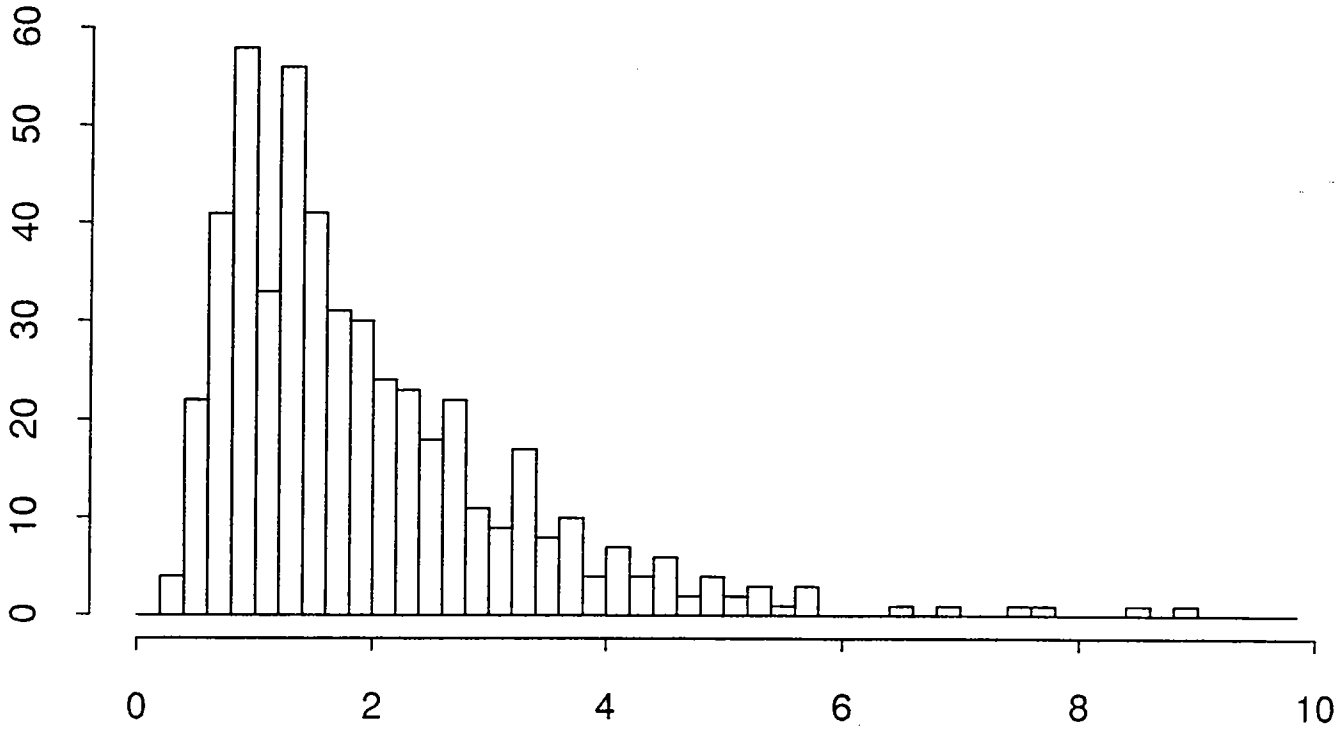


FIGURE 4.13. Bootstrap histogram of maximum modulus z value

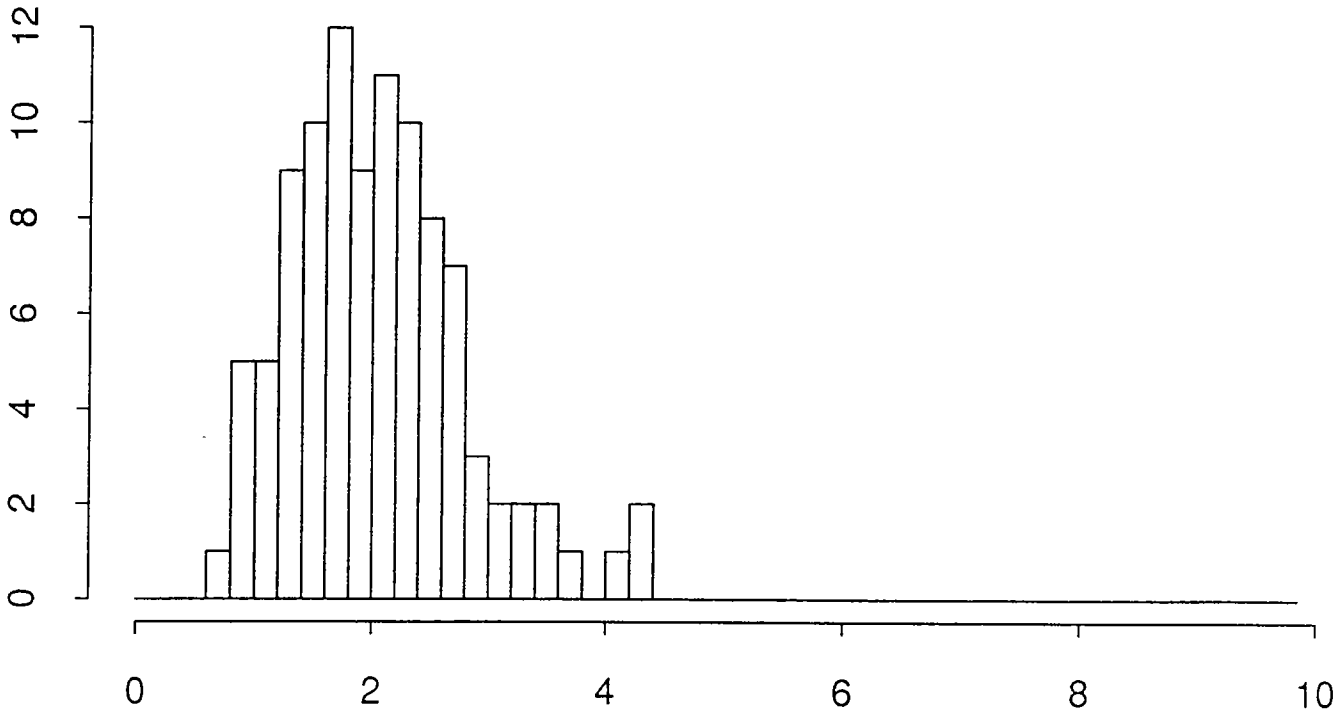
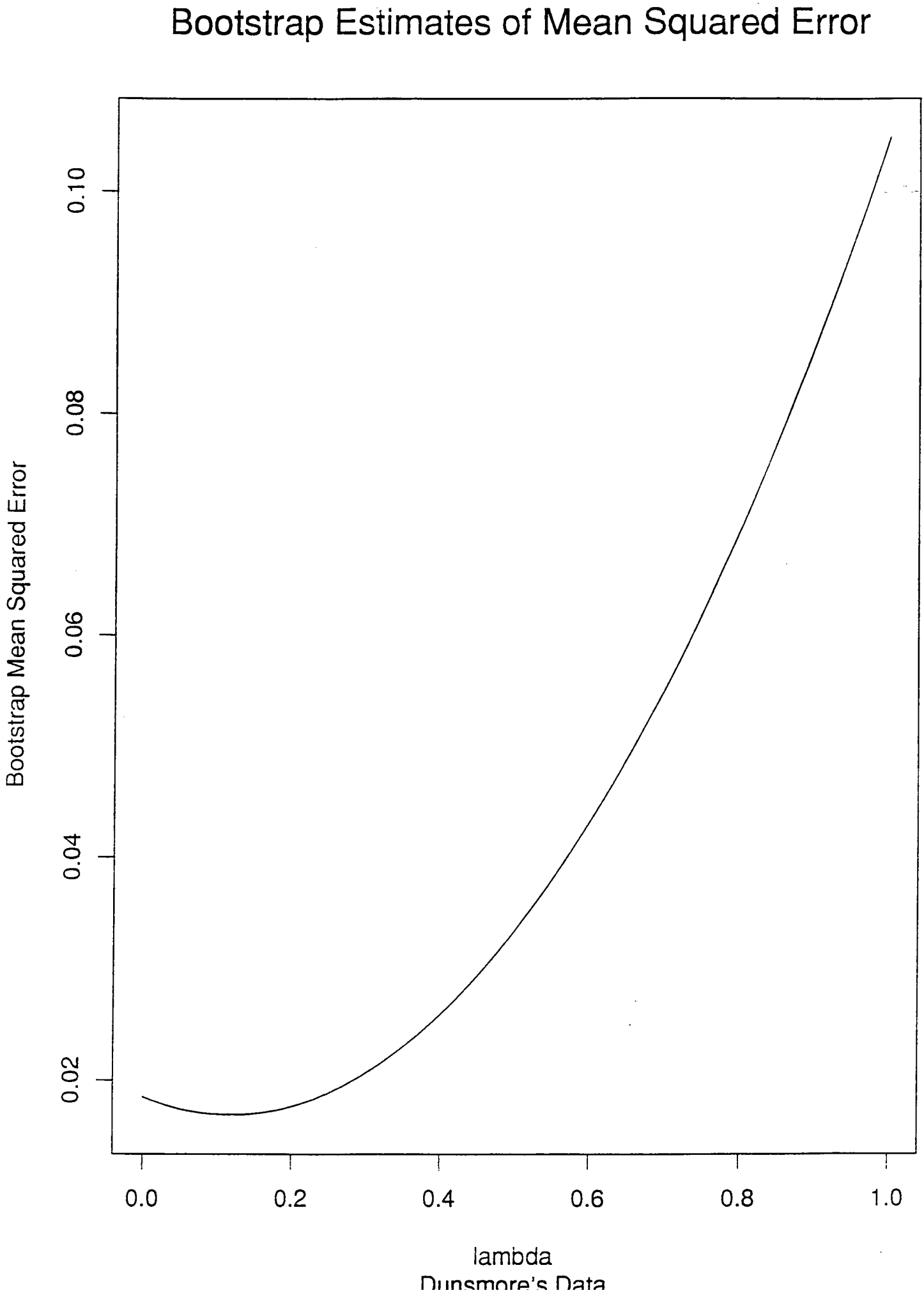


FIGURE 4.14. Empirical histogram of maximum modulus z value

FIGURE 5.1


Bootstrap Estimates of Mean Squared Error
Bootstrap Mean Squared Error
0.02
0.04
0.06
0.08
0.10
0.0
0.2
0.4
0.6
0.8
1.0
lambda
Dunsmore's Data