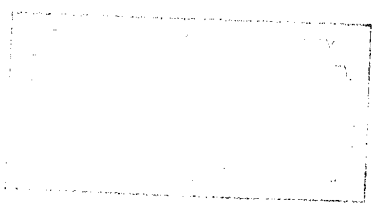
Frequency domain bootstrap methods for time series

- **Author/Creator:** Clifford Marc. Hurvich
- **Author/Creator:** Scott Zeger
- **Related Titles: Series:** Working paper series (New York University. Graduate School of Business Administration) ; 87-14.
- **Publisher:** New York, New York University, Graduate School of Business Administration
- **Language:** English
- **Publication Date:** 1987
- **Description:** 19 p. ; 28 cm.

Location:

New York University:

- Check Availability:
  - NYU Bobst 1st Floor Reference (WP 87-01 - 87-29 Non-circulating)
- Check Availability:
  - NYU Bobst Main Collection (WP 87-01 - 87-29)



FREQUENCY DOMAIN BOOTSTRAP METHODS FOR TIME SERIES

by

Clifford M. Hurvich and Scott Zeger

#87-14

February 1987

New York University
Graduate School of Business Administration
College of Business and Public Administration

New York University
Faculty of Business Administration
Working Paper Series
100 Trinity Place
New York, N.Y. 10006
Telephone: (212) 285-6204

Frequency Domain Bootstrap Methods For Time Series

Clifford M. Hurvich and Scott Zeger

Dept. of Statistics and Operations Research, New York University,

Dept. of Biostatistics, Johns Hopkins University

ABSTRACT

Two frequency domain bootstrap methods for weakly stationary time series will be proposed. The motivations for the proposed methods will be discussed, and the performance of the first method will be compared with that of a recently proposed method of Swanepoel and van Wyk, in a Monte Carlo study. It is found that, when applied to the problem of estimating the variance of a log spectrum estimate, all methods under consideration can sometimes perform poorly. Overall, the frequency domain method used in conjunction with an automatic spectrum estimate choice criterion developed by Hurvich, is found to perform best.

1: INTRODUCTION

In a recent paper, Swanepoel and Van Wyk (1986) proposed a bootstrap method designed for finite-order autoregressive (AR) stochastic processes. The method consists of resampling from the residuals of a fitted autoregression to construct AR bootstrap replications of the process. In a Monte Carlo study in which the simulated processes were all low order autoregressions, they examined the performance of the simultaneous spectral

confidence bands generated by the bootstrap distribution of an AR spectrum estimate. Swanepoel and van Wyk's method is time domain in character and seems to depend on the assumption that the process is either a finite order AR, or that it can be well modeled by a fitted finite order AR. (Analogous comments would apply to a new ARMA version of the method, which is under development.)

In this paper, we will propose two frequency domain bootstrap methods for weakly stationary stochastic processes. Our methods are free of any finite order time domain parametric assumptions (e.g., ARMA), and rely instead on the distributional properties of the discrete Fourier transform of the data (for Method 1) and the infinite order moving average representation of the process (for Method 2). Both methods require a basic spectrum estimate, and their success depends strongly on the quality of this estimate. We will use the cross-validatory WHOOSH method of Hurvich (1985) to automatically and objectively select both the estimate type (either AR or Blackman-Tukey), and the corresponding smoothness parameter (i.e, the model order or bandwidth). In a Monte Carlo study, we will treat the problem of estimating the variance of an AR spectrum estimate with the model order either fixed at 5, or variable and determined by Parzen's CAT criterion (Parzen, 1977). The case of CAT-selected order is particularly interesting, since the model order selection stage contributes to the variability of the resulting spectrum estimate. We will also treat the problem of estimating the variance of a Blackman-Tukey spectrum estimate having a fixed bandwidth. We will compare the performance of Method 1 with Swanepoel and van Wyk's method for three different simulated Gaussian processes: a superposition of complex exponentials with random phases and amplitudes, a third order moving average, and a third order autoregression. For Method 1, we will use a variety of basic estimates (of varying quality), including the automatically-selected WHOOSH estimate mentioned above.

Method 1 was independently proposed by Ramos (1984). The method relies on the approximate independence and normality of the Fourier coefficients: a sequence of bootstrap Fourier coefficients is generated by multiplying independent standard Gaussians by (a

constant multiple of) the square root of the spectrum estimate. These bootstrap Fourier coefficients are then inverse Fourier transformed to yield a bootstrap replication of the process. Since the resulting bootstrap replications are Gaussian, Method 1 cannot be expected to give accurate bootstrap distributions for time domain statistics (such as fitted AR coefficients) if the process is non Gaussian. (For frequency domain statistics such as spectrum estimates the problem may not be as severe, but some deterioration for non Gaussian processes is still to be expected.) Ramos (1984) and Stine (1985) propose bootstrap methods involving empirical resampling in the Frequency domain. (Ramos resamples from the normalized Fourier coefficients, while Stine resamples from the normalized periodogram. In both cases, normalization by a suitable function of the spectrum estimate is essential.) We, however, feel that these methods will be (slightly) outperformed by Method 1 for Gaussian processes, while for non-Gaussian processes they will still fail to capture any pertinent non-Gaussian aspects of the data.

Our Method 2, which we believe to be new, involves estimating the moving average transfer function of the process through the frequency domain FLES (Factorized Log of Estimated Spectrum) method of Bhansali to obtain estimates of the underlying innovations process (assumed to be strict white noise), resampling from the estimated innovations, and then reversing the whitening procedure to obtain a bootstrap replication. Method 2 is similar in spirit to Swanepoel and van Wyk's method in that both involve resampling from estimated innovations. The key difference is that Method 2 employs a frequency domain method to estimate the innovations while Swanepoel and van Wyk's method employs a time domain method. Both methods can be expected to work even if the process is non Gaussian.

2: MATHEMATICAL BACKGROUND

Suppose that we have data $x_0, \ldots, x_{n-1}$ generated by a weakly stationary zero mean real valued stochastic process $\{X_t\}$ with covariance function $\{c_\tau\}$, $c_\tau = E[X_t X_{t-\tau}]$, and spectrum

$$f(\omega) = \frac{1}{2\pi}\sum c_r \exp(ir\omega) \quad , \quad \omega \in [-\pi, \pi] \quad .$$

The Discrete Fourier Transform (DFT) of the data is $\{J_j\}_{j=0}^{n-1}$, where

$$J_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=0}^{n-1} x_t \exp(-i\omega_j t) \quad , \quad \omega_j = \frac{2\pi j}{n} \quad .$$

(ω_j is called the j'th Fourier Frequency.) We can write $J_j = A_j - iB_j$ where the Fourier coefficients $\{A_j\}$ and $\{B_j\}$ are given by

$$A_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=0}^{n-1} x_t \cos\omega_j t \quad ,$$

$$B_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=0}^{n-1} x_t \sin\omega_j t \quad .$$

Note that $A_j = A_{n-j}$, $B_j = -B_{n-j}$ $(j = 1, \ldots, n-1)$, and also that $B_0 = 0$, $B_{n/2} = 0$ (if n is even). The original data can be recovered from the DFT by the inversion formula

$$x_t = \frac{\sqrt{2\pi}}{\sqrt{n}} \sum_{j=0}^{n-1} J_j \exp(i\omega_j t) = \frac{\sqrt{2\pi}}{\sqrt{n}} \sum_{j=0}^{n-1} (A_j \cos\omega_j t + B_j \sin\omega_j t) \quad .$$

The normalized Fourier coefficients $\{A_j/\sqrt{f_j}\}_{j=1}^{\tilde{n}}$, $\{B_j/\sqrt{f_j}\}_{j=1}^{\tilde{n}}$ (with $f_j = f(\omega_j)$ and $\tilde{n} = [(n-1)/2]$) are often treated as *iid* $N(0, \frac{1}{2})$ random variables. This distributional result indeed holds asymptotically as $n \to \infty$ if the process is Gaussian and if the coefficients g_k of its moving average representation satisfy $\sum_k |g_k||k|^\alpha < \infty$ for some $\alpha > 0$. (See Priestley, 1981.) Without assuming a Gaussian process, Brillinger (1975) shows that under fairly weak conditions, any *finite* number of these normalized Fourier coefficients are asymptotically $N(0, \frac{1}{2})$.

Since f is unknown, we cannot actually compute the normalized Fourier coefficients. We can, however, use a consistent estimate $\hat{f}$ in place of f . We will consider two basic methods of spectrum estimation, the classical Blackman-Tukey (BT) method and the autoregressive (AR) method. The BT method is based on the periodogram

$$\{I_j\}_{j=0}^{n-1} = \{|J_j|^2\}_{j=0}^{n-1}$$

and takes the form

$$\hat{f}_b(\omega_j) = \sum_{k=-nb}^{nb} g_{k,b} I_{j-k} \quad , \tag{1}$$

where the $\{g_{k,b}\}_k$ are fixed weights, and b is a bandwidth parameter to be chosen by the user. The m'th order AR estimate takes the form

$$\hat{f}(\omega) = \frac{1}{2\pi} \frac{P_m}{|\sum_{k=0}^{m} a_k exp(i\omega_j k)|^2} \quad ,$$

where $a_0 = 1$, and $a_1, \ldots, a_m, P_m$ are determined from the sample covariances $\{\hat{c}_r\}_{r=0}^m$,

$$\hat{c}_r = \frac{1}{n} \sum_{t=r}^{n-1} x_{t-r} x_t \quad ,$$

through the Yule-Walker Equations,

$$\begin{bmatrix} \hat{c}_0 & \hat{c}_1 & \cdots & \hat{c}_m \\ \hat{c}_1 & \hat{c}_0 & \cdots & \hat{c}_{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{c}_m & \hat{c}_{m-1} & \cdots & \hat{c}_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

The model order m is to be chosen by the user.

A key question is how to choose the estimate type and corresponding smoothness parameter solely on the basis of the data. The WHOOSH method of Hurvich (1985), which will be used here, allows for the objective choice of an estimate from an arbitrary finite collection C of candidate estimates. In this paper, we will take C to be the union of the Yule-Walker AR estimates of orders $1, \ldots, 20$, and the modified discrete Daniell BT estimates (1) with

$$g_{k,b} = \begin{cases} 1/2nb & |k| < nb \\ 1/4nb & |k| = nb \end{cases} ,$$

$n = 100$, for $b = 1, \ldots, 24$. For an estimate $\hat{f}$ in C , define

$$WHOOSH(\hat{f}) = \frac{1}{n} \sum_{j=1}^{n} \{[\log \hat{f}^{-j}(\omega_j) - (\log I_j + C)]^2 - \frac{\pi^2}{6}\} \quad ,$$

where $C = .577216$ is Euler's constant, and $\hat{f}^{-j}(\omega_j)$, the leave-out-j version of the estimate, is defined below. The chosen estimate $\hat{f}$ is the minimizer of $WHOOSH(\hat{f})$ over the class C . The motivation for the WHOOSH criterion is that $WHOOSH(\hat{f})$ gives an approximately

unbiased estimate of the mean integrated squared error of $\hat{f}$,

$$MISE(\hat{f}) = \frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} E(\log\hat{f}(\omega_j) - \log f_j)^2 \quad ,$$

provided that $\hat{f}^{-j}(\omega_j)$ and I_j are approximately independent, and that $\hat{f}^{-j}(\omega_j)$ and $\hat{f}(\omega_j)$ are similar. Since all spectrum estimates considered here can be written as functions of the sequence $\{\hat{c}_r\}$ of estimated covariances,

$$\hat{f}(\omega;\{\hat{c}_r\}) \quad ,$$

the leave-out-j version of $\hat{f}$ can be defined (for a given fixed j) as

$$\hat{f}^{-j}(\omega) = \hat{f}(\omega;\{\hat{c}_r^{-j}\}) \quad ,$$

where

$$\hat{c}_r^{-j} = \frac{2\pi}{n'}\sum_{k=0}^{n'-1} I^{-j}(\omega'_k)\exp(ir\omega'_k) \quad ,$$

$n' = 2n$, $\omega'_k = \frac{2\pi k}{n'}$ and $I^{-j}(\omega)$ is the periodic extension of the function defined on $[-\pi,\pi]$ by

$$I^{-j}(\omega) = \begin{cases} I(\omega) & \omega \notin \{(\omega_{j-1},\omega_{j-1}) \cup (\omega_{-j-1},\omega_{-j-1})\} \\ \theta_{1,\omega} I(\omega_{j-1}) + \theta_{2,\omega} I(\omega_{j-1}) & \omega\in(\omega_{j-1},\omega_{j-1}) \\ I^{-j}(-\omega) & \omega\in(\omega_{-j-1},\omega_{-j-1}) \end{cases}$$

$$\theta_{1,\omega} = 1 - \frac{\omega-\omega_{j-1}}{\omega_{j-1}-\omega_{j-1}} \quad , \quad \theta_{2,\omega} = \frac{\omega-\omega_{j-1}}{\omega_{j-1}-\omega_{j-1}} \quad .$$

Note that $I^{-j}(\omega)$ is identical to $I(\omega)$ except on the intervals $(\omega_{j-1},\omega_{j-1})$ and $(\omega_{-j-1},\omega_{-j-1})$, but that the computation of $\hat{f}^{-j}(\omega_j)$ does not involve $I(\omega)$ for ω in these intervals. Thus, $\hat{f}^{-j}(\omega_j)$ and $\hat{f}(\omega_j)$ will be similar and $\hat{f}^{-j}(\omega_j)$ and I_j will be approximately independent.

Finally, we describe the bootstrap method of Swanepoel and van Wyk.

1) Given data $x_0,\ldots,x_{n-1}$, fit an m'th order autoregression, with parameters $\hat{a}_1,\ldots,\hat{a}_m$. Swanepoel and van Wyk select m by a procedure of Hannan and Quinn (1979) and fit the parameters by least squares. We will select m by BIC (see Priestley 1981, p. 375) and fit the parameters by the Yule-Walker method.

2) Compute the residuals $\hat{\epsilon}_t = \sum_{k=0}^{m}\hat{a}_k x_{t-k}$, $t = m,\ldots,n-1$.

3) Resample from the $\{\hat{\epsilon}_t\}_{t=0}^{n-1}$ to obtain $\{\epsilon_t^*\}_{t=0}^{N-1}$ where $N >> n$.

4) Use the $\{\epsilon_t^*\}$ to generate a bootstrap replication $\{x_t^*\}_{t=0}^{n-1}$: Define

$$y_{-m} = \cdots = y_{-1} = 0 \quad ,$$

$$Y_t = -\sum_{k=1}^{m} \hat{a}_k y_{t-k} + \epsilon_t^* \quad , \quad t = 0, \ldots, N-1 \quad ,$$

and

$$x_t^* = y_{t-N+n} \quad t = 0, \ldots, n-1 \quad .$$

3: TWO TIME SERIES BOOTSTRAP METHODS

Suppose that we have time series data $x_0, \ldots, x_{n-1}$ and that we are interested in the distribution of a statistic $T(x_0, \ldots, x_{n-1})$. Suppose also that a spectrum estimate $\hat{f}$ has already been calculated. (The choice of estimate is important, and was discussed in Section 2.) Define

$$\tilde{n} = [(n-1)/2] = \begin{cases} n/2 - 1 & n \text{ even} \\ (n-1)/2 & n \text{ odd} \end{cases} \quad ,$$

$$\delta\{n \text{ even}\} = \begin{cases} 1 & n \text{ even} \\ 0 & n \text{ odd} \end{cases} \quad .$$

Finally, define the inverse Fourier transform (IFT) of the sequence $\{z_j\}_{j=0}^{n-1}$ of complex numbers by

$$IFT_t\left(\{z_j\}_j\right) = \sum_{j=0}^{n-1} z_j \exp(i\omega_j t) \quad , \quad t = 0, \ldots, n-1 \quad .$$

Then our bootstrap methods can be described as follows.

Method 1: Parametric bootstrap

1) Generate n *iid* $N(0, \frac{1}{2})$ random variables

$$\{Z_{A,j}\}_{j=0}^{\tilde{n}-\delta\{n \text{ even}\}} \quad , \quad \{Z_{B,j}\}_{j=1}^{\tilde{n}} \quad .$$

2) Form simulated Fourier coefficients

$$A_j^* = Z_{A,j}\sqrt{\hat{f}_j} \qquad j = 0, \ldots, \tilde{n} + \delta\{n \text{ even}\} \quad ,$$

$$B_j^* = Z_{B,j}\sqrt{\hat{f}_j} \qquad j = 1, \ldots, \tilde{n} \quad .$$

3) Obtain a bootstrap data set $\{x_t^*\}_{t=0}^{n-1}$ by

$$x_t^* = \frac{\sqrt{2\pi}}{\sqrt{n}} \sum_{j=0}^{n-1} (A_j^* \cos\omega_j t + B_j^* \sin\omega_j t) \quad .$$

4) Evaluate $T(\{x_t^*\}_{t=0}^{n-1})$.

5) Repeat steps 1-4 B times (with B large) and approximate the true distribution of $T(\{x_t\})$ by the bootstrap distribution of $T(\{x_t^*\})$.

Method 2: The Nonparametric Bootstrap

This method assumes that $\{X_t\}$ has a one-sided moving average representation $X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k}$ where the $\{\epsilon_t\}$ are *iid* . Let $\Gamma(\omega) = \sum_{k=0}^{\infty} g_k \exp(-i\omega k)$. Note that $f(\omega) = |\Gamma(\omega)|^2$. The method is as follows:

1) Obtain $\{z_j\} = DFT \ \{x_t\}$

2) Obtain an estimate $\hat{\Gamma}(\omega)$ of $\Gamma(\omega)$

3) Form $\{\hat{\epsilon}_t\} = IFT \ \{z_j / \Gamma(\omega_j)\}$

4) Resample from $\{\hat{\epsilon}_t\}$ to obtain $\{\epsilon_t^*\}$

5) Form $\{W_j\} = DFT \ \{\epsilon_t^*\}$

6) Define $\{X_t^*\} = IFT\{\Gamma(\omega_j) \ W_j\}$.

Motivation for the bootstrap methods

Method 1:

The key assumption here is that the true normalized Fourier coefficients are *exactly iid* $N(0, \frac{1}{2})$. The parametric bootstrap relies on this assumption and uses $\hat{f}$ in place of f . Thus, *iid* Gaussians are generated, multiplied by $\sqrt{\hat{f}}$ and inverse Fourier transformed to obtain a bootstrap sample. If the process is exactly Gaussian and $\hat{f}$ is close to f , the parametric bootstrap can be expected to work well. Note, however, that the method indeed assumes that

the process is exactly Gaussian, since the bootstrap distribution of $\{x_t^*\}$ is always exactly Gaussian. As indicated earlier, the choice of spectrum estimate may be crucial to the success of this method.

Method 2:

To motivate method 2, note first that if $x_1, \ldots, x_n$ are known to be *iid* , then the method reduces to the ordinary bootstrap of Efron. This follows since we can take

$$\Gamma(\omega) = \hat{f}(\omega) = 1 \quad ,$$

so that the $\{\hat{\epsilon}_t\}$ sequence defined in step 3) and resampled from in step 4) is

$$\{\hat{\epsilon}_t\} = IFT \ \{z_t\} = \{x_t\} \quad .$$

Suppose next that $\Gamma(\omega)$ is known, but not necessarily constant. Then $\{X_t\}$ has a spectral representation

$$X_t = \int_{-\pi}^{\pi} \exp(i\omega t)\Gamma(\omega)\, dZ_\epsilon(\omega) \quad ,$$

and the DFT's of $\{x_t\}_{t=0}^{n-1}$ and $\{\epsilon_t\}_{t=0}^{n-1}$ are related by

$$J_x(\omega) = \Gamma(\omega)J_\epsilon(\omega) + r_n(\omega) \quad , \tag{2}$$

where $E\,|r_n(\omega)|^2 = O(n^{-1})$, uniformly in ω . (Equation (1) holds under fairly weak conditions. See Priestley, 1981.) Thus, step 3) yields

$$\{\hat{\epsilon}_t\} = IFT \ \{z_j/\Gamma(\omega_j)\} \approx IFT \ \{J_\epsilon(\omega_j)\} = \{\epsilon_t\} \quad .$$

We see that the resampling in Step 4 is well-motivated, since the $\{\hat{\epsilon}_t\}$ are approximately equal to the underlying *iid* random variables $\{\epsilon_t\}$. Thus if $\{X_t\}$ has a known one-sided moving average representation with respect to *iid* white noise, then Method 2 is clearly superior to Method 1. In practice $\Gamma(\omega)$ must be estimated using, for example, the method of Bhansali (1974, 1977).

4: MONTE CARLO COMPARISONS OF METHOD 1 WITH SWANEPOEL AND VAN WYK'S METHOD

Here, we compare the performance of the parametric frequency domain bootstrap (Method 1) with the time domain method of Swanepoel and van Wyk. The particular application of the bootstrap we consider here is the assessment of the variances of log spectrum estimates. For simplicity, we focus on this application instead of the simultaneous spectral confidence interval application. (Schenker (1985) discusses some drawbacks inherent in (individual) bootstrap confidence intervals. In addition, it is not clear how best to judge the quality of the simultaneous confidence intervals considered by Swanepoel and van Wyk. Finally, we feel that more can be learned about the advantages and disadvantages of the methods, and about the reasons for these advantages and disadvantages, by studying the behavior of the bootstrap variances.)

For each of three Gaussian processes, 100 simulated realizations were generated. Each realization was of length $n = 100$. The three process types were a superposition of complex exponentials with random phases and amplitudes (called the Sinusoids process), a third order moving average process, and a third order autoregression. (These are the same three processes used in Hurvich, 1985.) Details on the generation of these processes is given in the Appendix. The bootstrap methods to be compared are the time domain bootstrap method of Swanepoel and van Wyk (SVW), and the frequency domain bootstrap method (i.e., Method 1). For the latter, two different normalizing spectrum estimates are considered: WHOOSH, and BIC. Thus, in all, there are three different methods considered, and they are denoted by SVW, WHOOSH and BIC. For each realization of the process, the three bootstrap methods were used to estimate the variances of the logarithms of three different spectrum estimates: a discrete Daniell estimate (1) with bandwidth $b = 5$ (denoted BT5), an autoregressive estimate of order 5 (AR5), and an autoregressive estimate with the order automatically selected by the CAT criterion (ARCAT). In all cases, the bootstrap variance estimates were based on $B = 100$ bootstrap replications of the process.

In Figures 4.1-4.3, we plot the averages of these bootstrap variance estimates, taken over the 100 realizations of the process. We also plot the "true" variances, which are based

on 10000 realizations of the process. The true variances are given by the solid line, the averages of the WHOOSH variances are given by the dotted line, the averages of the BIC variances are given by the dashed line (long dashes), and the averages of the SVW variances are given by the dashed line (short dashes). In Figure 4.4, we give boxplots of the integrated squared errors of the log spectrum estimates,

$$ISE(\hat{f}) = \frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}(\log\hat{f}(\omega_j) - \log f(\omega_j))^2 \quad .$$

The principle finding of our study is that bootstrap variance estimates can be highly inaccurate unless the spectrum used to generate the bootstrap replications is close to the actual spectrum. Of course, in practice the actual spectrum is unknown and hence bootstrap variances for spectrum estimates must be used with caution. The largest bias was found in bootstrap variances for ARCAT spectrum estimates, followed by BT5, then AR5 estimates. We have found the parametric frequency domain bootstrap with the BIC normalization to be essentially equivalent to the time domain method of Swanepoel and van Wyk (1986). These two methods perform well when the data satisfy an autoregressive model but perform poorly for the MA and Sinusoids process. Overall, the WHOOSH parametric bootstrap was preferable.

Figure 4.1 displays the bootstrap variances for the ARCAT spectrum estimate for the MA, Sinusoids and AR processes. The actual variance and parametric bootstrap estimates normalized by WHOOSH and BIC as well as the variance obtained by the time domain bootstrap are shown. For the MA and Sinusoids process, the time domain and frequency domain BIC variances substantially underestimate the true variance, in the MA case by a factor of 1.5 to 2. (Assuming the log spectrum estimates are approximately Gaussian and unbiased for the log spectrum, underestimating the variance by a factor of 2 corresponds to 95% nominal coverage probabilities with actual coverages of only 85%.) These methods generate bootstrap replications having the correlation structure of the AR process. When the data are autoregressive, they perform well (Figure 4.1c); when they are not, they perform poorly, in part because they fail to properly reflect the contribution to the spectrum

estimate's variance from the order-selection stage. Since for a given data set all bootstrap replications have the correlation structure of a fixed-order AR process, variability in the CAT-selected order of the bootstrap-replicated spectrum estimates will be minimized. The WHOOSH variance is less biased for the MA and Sinusoids process. One reason is that the WHOOSH spectrum estimates are superior to the AR-BIC spectrum estimates for these processes. Another contributing factor is that the WHOOSH-chosen spectrum estimates are often not autoregressive. When they are, however, the bootstrap variances will be biased downward, for reasons described above. Thus the average WHOOSH bootstrap variances shown in Figures 4.1a and 4.1b are still somewhat smaller than the true variances. For the AR data, WHOOSH overestimates the variance of the ARCAT spectrum estimates by up to a factor of 2. This problem may be linked to the inferior performance of WHOOSH for selecting a spectrum estimate when the underlying process is a small-order autoregression.

Figure 4.2 displays the variance of the Blackman-Tukey spectrum estimate (BT5) for the three processes. WHOOSH gives the best variance estimate for the MA and Sinusoids process. For the AR process, BIC and WHOOSH are best. Interestingly, this is the one case where the time domain and BIC frequency domain variances differ, but this occurs only at low frequencies. The WHOOSH variance appears best overall.

For AR5 spectrum estimates, all methods perform reasonably well, as is evidenced in Figure 4.3.

5: APPENDIX: SIMULATION METHODS

Using the AT&T Bell Laboratories S random number generator on a Vax 11/750, we simulated three different types of zero mean Gaussian processes, all of length $n = 100$: a superposition of sinusoids, a third order autoregression, and a third order moving average. The basic uniform random number generator used by S is the McGill generator, described in Chambers (1977, pp. 174, 191).

The superposition of sinusoids (referred to simply as "sinusoids" in the text) has form

$$x_t = \sum_{j=1}^{n/2} (A_j \cos\omega_j t + B_j \sin\omega_j t) \tag{3}$$

where for each j, A_j and B_j are independent $N(0,\sigma_j^2)$ and where the pairs (A_i,B_i) and (A_j,B_j) are independent if $i \neq j$. Here, the σ_j $(j = 1, \ldots, n/2)$ are constants. It is easy to show that $\{x_t\}$ has a *discrete* spectrum with power $\frac{1}{2}\sigma_j^2$ at ω_j. Since the interval $(\omega_{j-1},\omega_j]$ has length $\frac{2\pi}{n}$, we see that $\{x_t\}$ provides an approximation to a process with spectral density $f(\omega)$, where

$$f(\omega_j) = \frac{n\sigma_j^2}{4\pi} \quad j = 1, \ldots, n/2 \ .$$

Thus, by properly choosing the σ_j^2, we can approximate processes with spectral densities $f(\omega)$ having any desired shape. Simulation of $\{x_t\}$ is straightforward: if

$$w_1, \ldots, w_{n/2}, y_1, \ldots, y_{n/2}$$

are simulated *iid* $N(0,1)$ random variables, then we define

$$A_j = \sigma_j w_j \ , \quad B_j = \sigma_j y_j \qquad (j = 1, \ldots, n/2)$$

and then form the sum (3). The values of σ_j we have chosen to use here are

$$\sigma_j = \sin\left(\frac{2\pi j}{101}\right) \quad (j = 1, \ldots, 50) \ .$$

The autoregression is

$$x_t + .9x_{t-1} + .8x_{t-2} + .6x_{t-3} = \epsilon_t \ , \quad \epsilon_t \overset{iid}{\sim} N(0,1) \ .$$

To simulate $\{x_t\}$, we first generate $\{\epsilon_t\}_{t=0}^{199}$, a set of 200 simulated *iid* $N(0,1)$ random

variables. Next, we define $\{y_t\}$ recursively by

$$y_{-3} = y_{-2} = y_{-1} = 0 \quad ,$$

$$y_t = -.9y_{t-1} - .8y_{t-2} - .6y_{t-3} + \epsilon_t \qquad t = 0, \ldots, 199$$

and finally, we set

$$x_t = y_{t+100} \qquad t = 0, \ldots, 99 \quad .$$

The moving average is

$$x_t = \epsilon_t + .9\epsilon_{t-1} + .8\epsilon_{t-2} + .6\epsilon_{t-3} \quad , \quad \epsilon_t \overset{iid}{\sim} N(0,1) \quad .$$

Simulation of $\{x_t\}$ is straightforward: generate the $\{\epsilon_t\}$ and form the indicated sum.

REFERENCES

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

CHAMBERS, J. (1977), *Computational Methods* for *Data Analysis* , New York: John Wiley.

HURVICH, C. (1985), "Objective Choice of a Spectrum Estimate: Extending the Applicability of Cross Validation Methods," *Journal of the American Statistical Association* , 80, 934-940.

PARZEN, E. (1977), "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes," *Multivariate Analysis – IV*, P.R. Krishnaiah, Ed., Amsterdam: North-Holland.

PRIESTLEY, M.B. (1981), *Spectral Analysis* and *Time Series*, New York: Academic Press.

RAMOS, E. (1984), "A Bootstrap For Time Series," Qualifying Paper, Department of Statistics, Harvard University.

SCHENKER, N. (1985), "Qualms About Bootstrap Confidence Intervals," *Journal of the American Statistical Association* , 80, 360-361.

STINE, R. (1985), "A Frequency Domain Bootstrap For Time Series," Unpublished Manuscript, Department of Statistics, Wharton School, University of Pennsylvania.

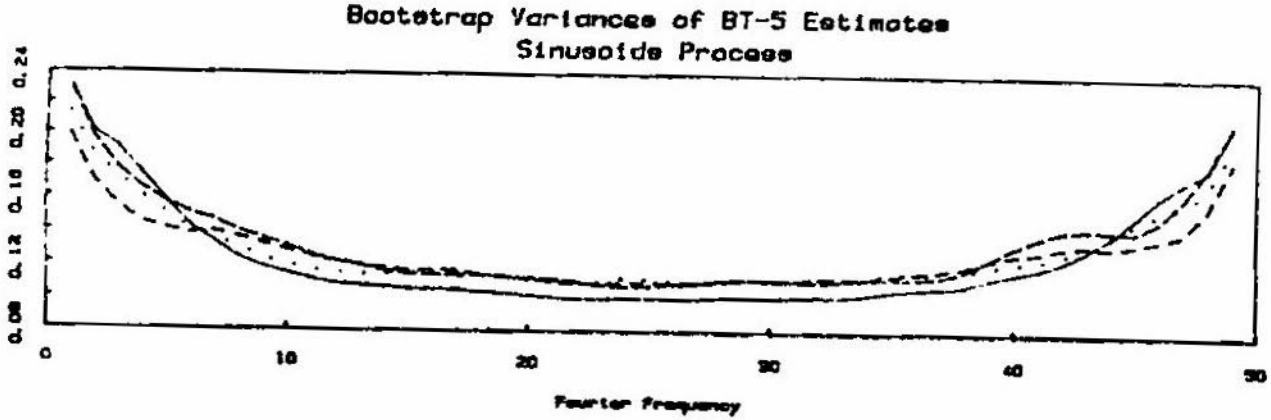


Fig. 4.2a

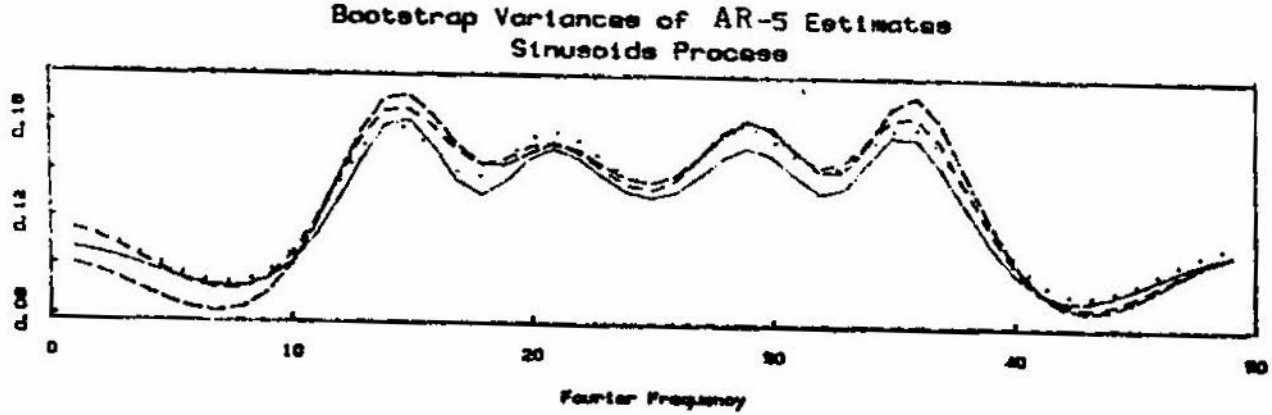


Fig. 4.3a

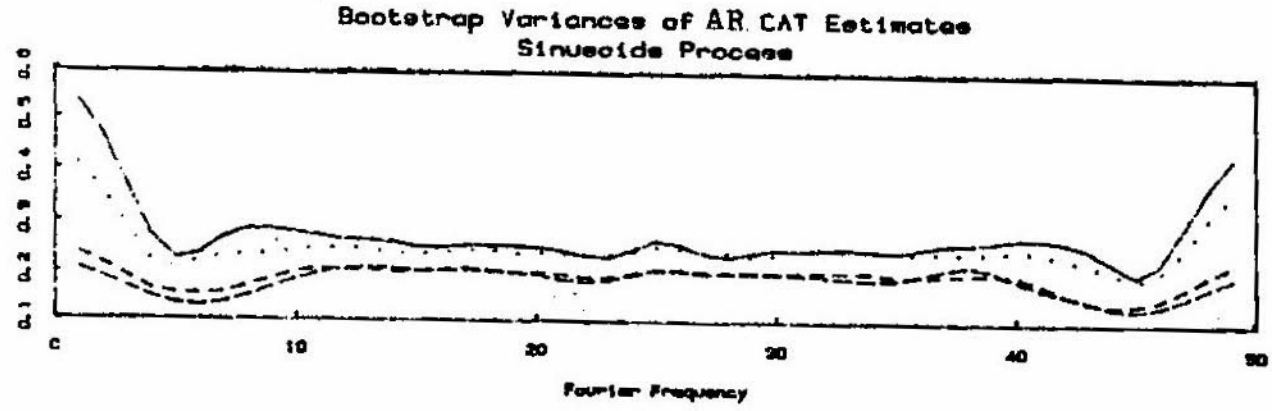


Fig. 4.1a

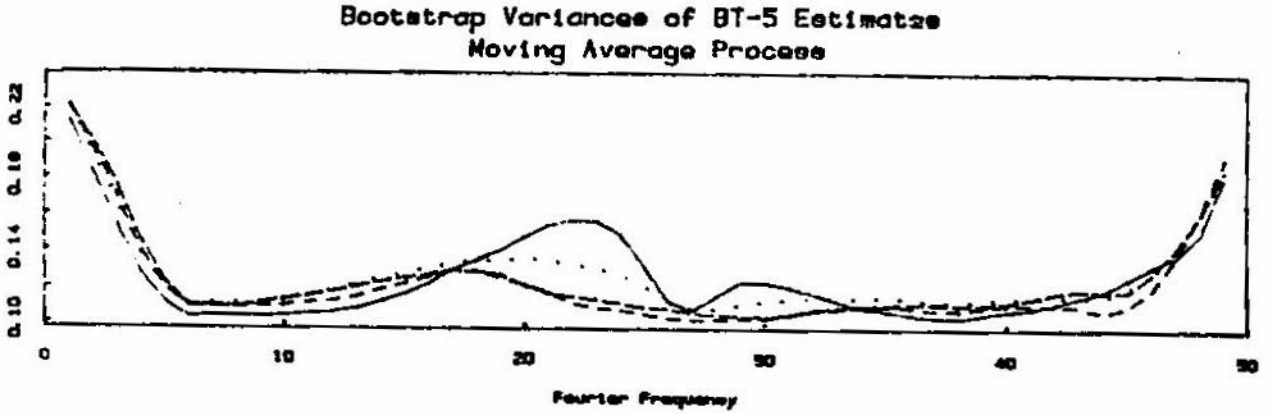


Fig. 4.2b

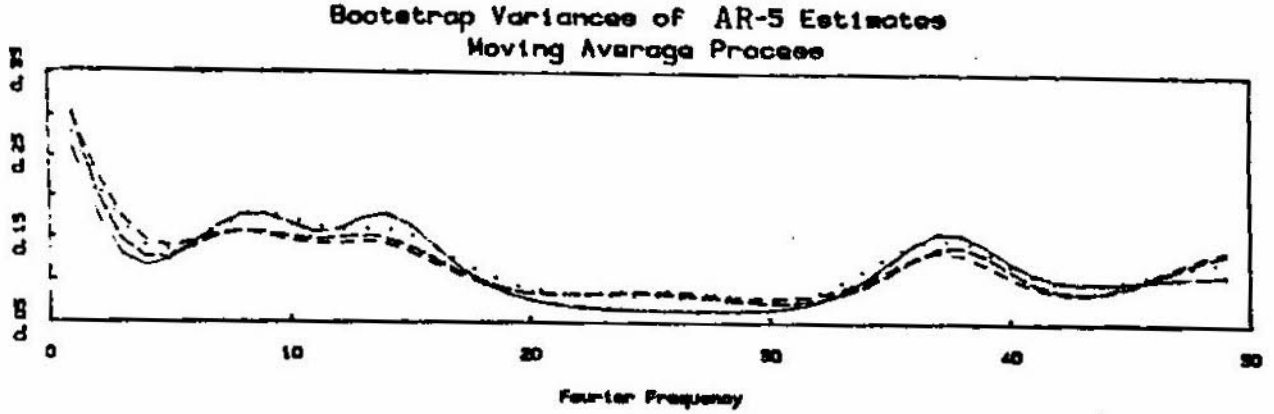


Fig. 4.3b

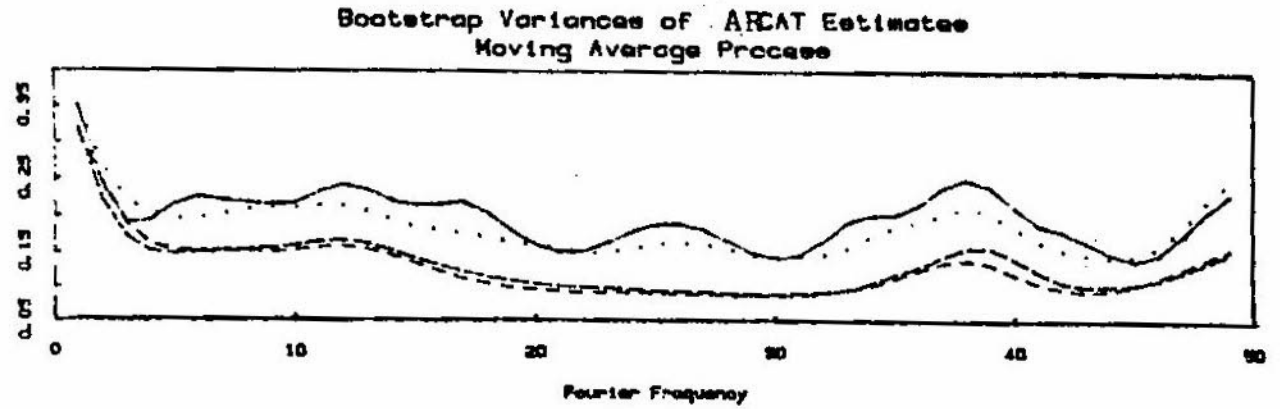


Fig. 4.1b

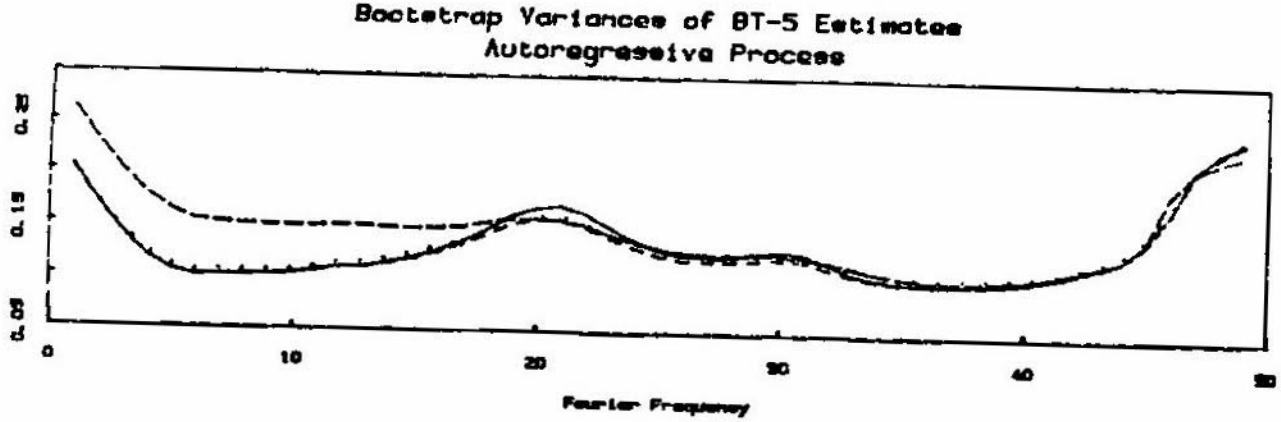


Fig. 4.2c

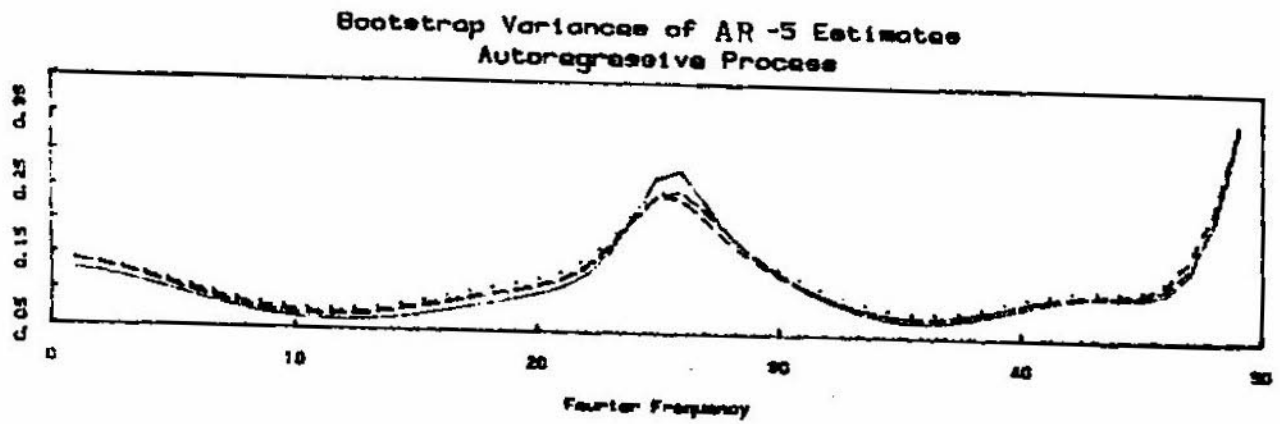


Fig. 4.3c

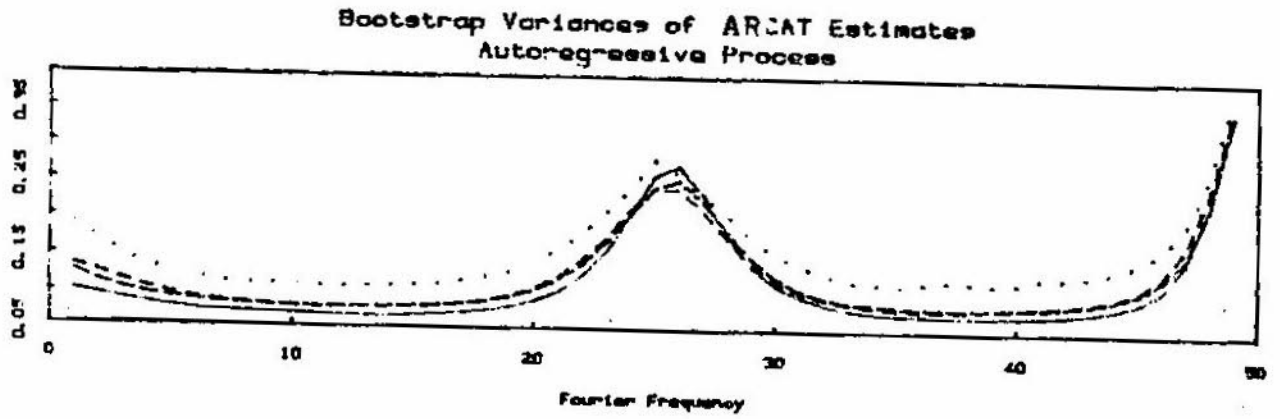
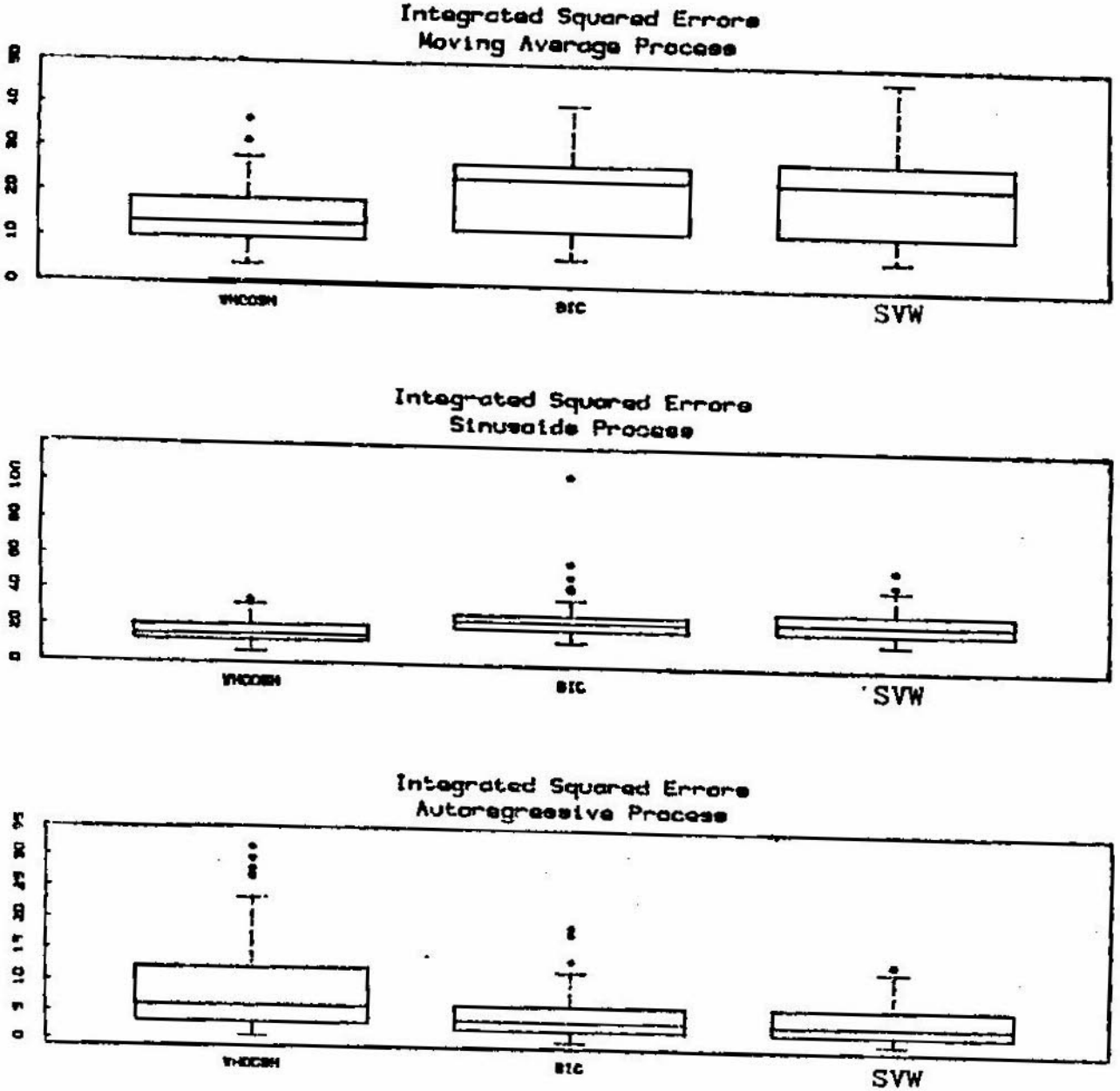


Fig. 4.1c

FIGURE 4.4


Integrated Squared Errors
Moving Average Process
BIC
SVW
Integrated Squared Errors
Sinusoide Process
BIC
SVW
Integrated Squared Errors
Autoregressive Process
BIC
SVW

Frequency domain bootstrap methods for time series

- **Author/Creator:** Clifford Marc. Hurvich
- **Author/Creator:** Scott Zeger
- **Related Titles: Series:** Working paper series (New York University. Graduate School of Business Administration) : 87-14.
- **Publisher:** New York, New York University, Graduate School of Business Administration
- **Language:** English
- **Publication Date:** 1987
- **Description:** 19 p. ; 28 cm.

Location:

New York University:

- Check Availability:
  - NYU Bobst 1st Floor Reference (WP 87-01 - 87-29 Non-circulating)
- Check Availability:
  - NYU Bobst Main Collection (WP 87-01 - 87-29)