Bootstrap Prediction Intervals for Autoregression

by

Lori A. Thombs and William R. Schucany

Technical Report No. SMU/DS/TR-201
Department of Statistics, ONR Contract

March 1987

Research sponsored by the Office of Naval Research
Contract N00014-85-K-0340
Project NR 042-479

Department of Statistics
Southern Methodist University
Dallas, Texas 75275

BOOTSTRAP PREDICTION INTERVALS FOR AUTOREGRESSION

Lori A. Thombs and William R. Schucany*

ABSTRACT

The nonparametric bootstrap is applied to the problem of prediction in autoregression. Let $\{Y_t: t=0,\pm1,\pm2,\ldots\}$ be a stationary autoregressive process of known order p. Given a realization of the series up to time t, $(y_1,y_2,\ldots,y_t)$, a $100(1-\alpha)\%$ prediction interval for Y_{t+k} is desired. Standard forecasting techniques, which assume that the error sequence of the process $\{Y_t\}$ is Gaussian, rely upon the fact that the conditional distribution of Y_{t+k} given the data is also Gaussian. As a nonparametric alternative, the bootstrap provides an estimate of the conditional distribution of Y_{t+k}. The method is similar to other applications of the bootstrap for linear models due to the residuals being resampled. The proposed methodology represents a different approach, since an alternative representation for AR(p) series is used allowing for bootstrap replicates generated backward in time. It follows that the resulting replicates all have the same conditionally fixed values at the end of every series. A simulation which compares the proposed technique with the standard technique for low-order Gaussian and Non-Gaussian autoregressive models demonstrates the potential of the bootstrap technique.

KEY WORDS: Forecasting; Non-Gaussian Time Series; Resampling

---

* Lori A. Thombs is Assistant Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. William R. Schucany is Professor and Chair of the Department of Statistical Science, Southern Methodist University, Dallas, TX 75275. This work was partially supported by Office of Naval Research Contract N00014-85-K-0340.

I. INTRODUCTION

An important problem in the statistical analysis of time series is the following: Given a realization of a series up to time t, what can be said about the observation at time (t+k)? The methodology presented here is an application of the bootstrap to the problem of interval forecasting for p-th order autoregressive (AR(p)) time series. Clearly, this addresses only one facet of this large problem, but a sufficiently interesting collection of practical problems are adequately approximated by AR(p).

Initially, the bootstrap technique was used in the case of independent, identically distributed (iid) random variables, but recently the scope of the bootstrap has been extended to include regression and AR(p) models. Freedman (1981) established the validity of the bootstrap as a technique for estimating standard errors of parameter estimates in a regression model. Freedman and Peters (1984a) present empirical evidence that the bootstrap provides good estimates of standard errors of estimates in a multi-equation linear dynamic model. Unlike either the iid or regression setting, there is no well-defined general method for bootstrapping dependent data. Stine (1982) and Findley (1985) proposed procedures for using the bootstrap to estimate the mean squared error of forecasts for the AR(p) model. The methods presented here represent yet a different application of the bootstrap to prediction in autoregression.

There are several methods available for predicting time series. Prediction intervals, random sets designed to contain the random variable Y_{t+k}, can be more informative than a point estimate of the future value Y_{t+k} and an estimated variance. Classical prediction intervals for autoregression, (such as the widely-used Box-Jenkins method,) require specifying the distribution of the error sequence associated with the process. The Box-Jenkins procedure assumes a normal error distribution and can be adversely affected by departures from normality. Furthermore, these approaches typically do not attempt to take account of the sampling variability of the estimated coefficients. This can also lead to lower than nominal coverage of the Box-Jenkins prediction intervals.

As a nonparametric alternative, a bootstrap procedure for estimating the conditional distribution of Y_{t+k}, given the observations up to time t, is proposed here. This estimate reflects departures from normality of the underlying error sequence, and yields a $100(1-\alpha)\%$ prediction interval for Y_{t+k}. It also captures the additional variability inherent in coefficient estimation. Such intervals are useful when the series length is short or the assumption of Gaussian errors may not be justified.

The next section contains a description of the method for obtaining bootstrap replicates of the observed series. Section 3 presents an algorithm for obtaining the bootstrap distribution of Y_{t+k}, and the bootstrap prediction interval is defined. Asymptotic justification of the bootstrap method is also given in Section 3. The last section contains a small sample Monte Carlo study which compares the relative performances of the bootstrap and Box-Jenkins prediction intervals for several error distributions. It also reports a simulation study which provides an empirical illustration of the asymptotic theory.

II. THE BOOTSTRAP REPLICATE

A discrete time, real valued autoregressive series of known order p is defined by:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t, \qquad (2.1)$$
$$t = 0, \pm 1, \pm 2, \ldots,$$

where

(i) $\{a_j\}$ is a sequence of zero mean, uncorrelated random variables with common distribution function F_a and $E[a_j^2]=\sigma^2<\infty$,

and

(ii) $\delta, \phi_1, \phi_2, \ldots, \phi_p$ are unknown constants.

First we discuss some general issues that are relevant to bootstrapping in this context. Assume that the (potentially infinite) sequence of n observations $(y_{t-n+1}, \ldots, y_t)$ are available from an AR(p) process. (Due to our emphasis on conditionally fixing y_t and other previous values, the device of indexing backward to produce n data values will be used throughout.) There is no obvious conventional procedure for obtaining a bootstrap replicate $(y^*_{t-n+1}, \ldots, y^*_t)$. The various procedures recently proposed in the literature share the characteristic of resampling the residuals. Define the i^{th} residual by

$$\hat{a}_i = y_i - \hat{\delta} - \hat{\phi}_1 y_{i-1} - \cdots - \hat{\phi}_p y_{i-p}, \qquad (2.2)$$
$$i=t, t-1, \ldots t-n+p+1,$$

where $(\hat{\delta}, \hat{\phi}_1, \ldots, \hat{\phi}_p)$ are the least squares estimates of the parameters. Since the $\hat{a}_i$ replace the true errors a_i of (2.1) in the generation of bootstrap replicates, it is important that they provide a good estimate of F_a. If another estimation criterion other than least squares is used, then the resulting residuals may not be centered. In this case, centering of the residuals by,

$$\tilde{a}_i = \hat{a}_i - \bar{a},$$

where

$$\bar{a} = \sum_{i=t}^{t-n+p+1} \hat{a}_i/(n-p)$$

is recommended. Exact results concerning the variance of the $\hat{a}_i$ are difficult to obtain, but empirical evidence indicates that the residuals have been deflated due to fitting (Freedman and Peters, 1984b). Their studies suggest that bootstrap techniques will give better results if the residuals are rescaled by the factor $\left[\frac{n}{n-p}\right]^{1/2}$. Let $\hat{F}_a$ denote the empirical cumulative distribution function (cdf) of the centered, rescaled residuals.

Due to the recursive nature of (2.1), p starting values are needed in order to generate a bootstrap replicate. Proposals include using the first p values of the series (Efron and Tibshirani, 1986) or randomly selecting a block of p adjacent values from the observed series (Stine). Given the p starting values, the remainder of the bootstrap replicate would be generated by the recursive equation defining the AR(p) process:

$$y_j^* = \hat{\delta} + \hat{\phi}_1 y_{j-1}^* + \cdots + \hat{\phi}_p y_{j-p}^* + \hat{a}_j^*, \qquad (2.3)$$

$$j=t-n+p+1,\ldots,t-1,t,$$

where $\hat{a}_j^*$ is a random draw from $\hat{F}_a$ and $(\hat{\delta},\hat{\phi}_1,\ldots,\hat{\phi}_p)$ are the parameter estimates in (2.2).

As it is usually employed, B bootstrap replicates $\underline{y}_b^* = (y_{t-n+1}^*,\ldots,y_t^*)_b'$ are needed to apply the bootstrap technique. The method given above has been shown to work well when the problem is estimation of the standard error of the estimates $\hat{\phi}_i$, i=1,...,p. However, when the bootstrap is used to estimate the actual conditional distribution of Y_{t+k}, the above procedure is not applicable. It is well known that for AR(p) models, the distribution of Y_{t+k} conditional on all past observations is the same as the conditional distribution of Y_{t+k} given the last p values. Ideally, for the bootstrap to effectively simulate the conditional distribution of Y_{t+k}, the method of generating bootstrap replicates should:

(i) produce bootstrap realizations $\underline{y}_1^*, \underline{y}_2^*, \ldots, \underline{y}_B^*$ that mimic the correlation structure of the series being predicted, and

(ii) conditionally fix the last p values; that is, for every replicate, $y_t^* = y_t$, $y_{t-1}^* = y_{t-1}, \ldots, y_{t-p+1}^* = y_{t-p+1}$.

There is an alternative representation for stationary AR(p) series, called the "backward representation," in which the process is generated by a forward difference operator (Box and Jenkins, pp. 197-200.) In this representation, the random variable Y_t is expressed as a linear combination of future values plus an error term,

$$Y_t = \delta + \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \cdots + \phi_p Y_{t+p} + e_t. \qquad (2.4)$$

An important result is that the correlation structures of the process generated by the forward representation (2.1) and the backward representation (2.4) are identical. In the bootstrap setting, this "time-reversible" property of stationary AR(p) processes is useful. Since (2.4) requires the variables y_j^* to be generated backward in time, a natural choice for the starting values of the bootstrap replicate are the <u>last</u> p values of the observed series. This provides a method of generating conditional bootstrap replicates that have the same last p values of the series and also the same correlation structure as the series being predicted.

This section concludes with a description of the new proposal for obtaining bootstrap replicates. Using the same notation as in (2.2), the errors e_j of the backward representation (2.4) are estimated by

$$\hat{e}_j = y_j - \hat{\delta} - \hat{\phi}_1 y_{j+1} - \cdots - \hat{\phi}_p y_{j+p},$$

$$j = t-p, t-p-1, \ldots, t-n+1.$$

Let $\hat{F}_e$ denote the cdf of the centered, rescaled "backward residuals." To obtain a bootstrap replicate, set

$$\begin{aligned} y_t^* &= y_t \\ y_{t-1}^* &= y_{t-1} \\ &\vdots \\ y_{t-p+1}^* &= y_{t-p+1}, \end{aligned} \tag{2.5}$$

and generate the remainder of $\underline{y}^*$ by the recursive equations

$$y_j^* = \hat{\delta} + \hat{\phi}_1 y_{j+1}^* + \cdots + \hat{\phi}_p y_{j+p}^* + \hat{e}_j^* , \tag{2.6}$$

$$j = t-p, \ldots, t-n+p,$$

where $\hat{e}_j^*$ are iid from $\hat{F}_e$. Hence, a typical bootstrap replicate is $(y_{t-n+1}^*, y_{t-n+2}^*, \ldots, y_{t-p}^*, y_{t-p+1}, \ldots, y_t)$. It can be shown that the true error distributions F_a and F_e of the forward and backward models are the same when they are normal. For non-Gaussian time series, it is important to make a distinction between F_a and F_e.

An important feature of the forecasting problem has been retained in the proposed resampling method. Since the last p values of the replicates are fixed, all bootstrap inference is with respect to the conditional distribution of Y_{t+k}, which is the relevent distribution in the forecast problem. This feature is illustrated in Figure 1, which shows several bootstrap replicates from a realization of a second order AR process.

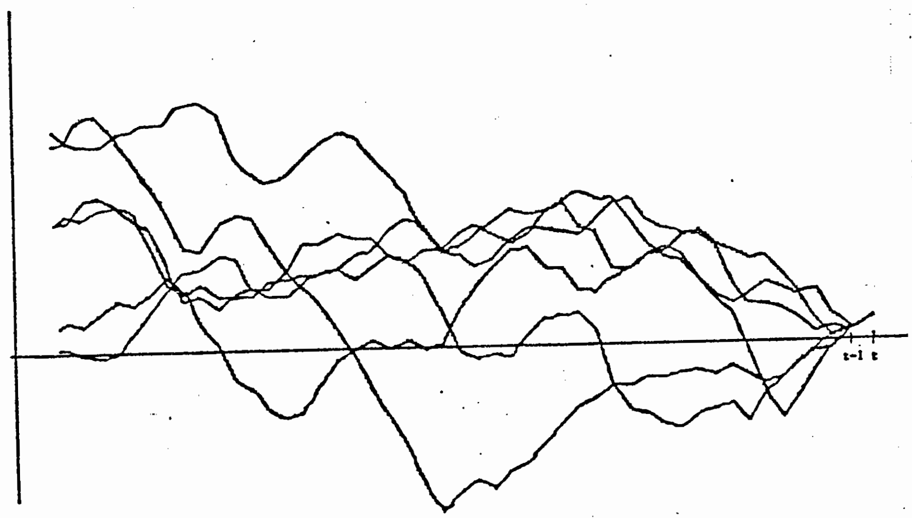

Figure 1: Six Bootstrap Replicates from an AR(2) process (δ=0, ϕ_1= 1.75, ϕ_2 = -.76) of length n=50, conditional upon having the same last p=2 values, y_{49} and y_{50}.

III. ALGORITHM FOR FORECAST INTERVAL AND ASYMPTOTIC RESULTS

Suppose that observations $(y_{t-n+1}, y_{t-n+2}, \ldots, y_t)$ are available from an AR(p) process and let $(y^*_{t-n+1}, \ldots, y^*_{t-p}, y_{t-p+1}, \ldots, y_t)$ denote a typical bootstrap replicate. A $100(1-\alpha)\%$ prediction interval for Y_{t+k} is desired. The proposed procedure begins by calculating a "bootstrap future value" Y^*_{t+k} for each of the B replicates. For a single replicate, Y^*_{t+k} is obtained from

$$Y^*_{t+k} = \hat{\delta}^* + \hat{\phi}^*_1 y^*_{t+k-1} + \cdots + \hat{\phi}^*_p y^*_{t+k-p} + \hat{a}^*_{t+k}, \tag{3.1}$$

where $(\hat{\delta}^*, \hat{\underline{\phi}}^*)$ are the least squares estimates based on the bootstrap data and $\hat{a}^*_{t+k}$ is a random draw from $\hat{F}_a$.

Depending on the lag k, the values y^*_{t+k-i}, $i=1,\ldots p$, in (3.1) will be (i) one of the last p values of the bootstrap replicate, or (ii) a bootstrap future value Y^*_{t+k}. In case (i), due to the conditional nature of the bootstrap data, y^*_{t+k-i} will be the same for each future value that is calculated. Hence y^*_{t+k-i} does not contribute to the variability of the Y^{*b}_{t+k} $(b=1,\ldots,B)$ values. As an illustration, consider the second order model with k=1. In this case,

$$Y^*_{t+1} = \hat{\delta}^* + \hat{\phi}^*_1 y^*_t + \hat{\phi}^*_2 y^*_{t-1} + \hat{a}^*_{t+1} \ . \tag{3.2}$$

Since $y^*_{t-1} = y_{t-1}$ and $y^*_t = y_t$, variability in the Y^*_{t+1} values is determined by the $(\hat{\delta}^*, \hat{\phi}^*)$ and $\hat{a}^*$ values, which vary with each bootstrap replicate. This is consistent with the result that a point predictor of Y_{t+k} (with estimated parameters) has prediction error variance that can be decomposed into two parts: (i) variability due to the residual term and (ii) variability due to parameter estimation.

For leads greater than one, Y^*_{t+k} must be calculated in a recursive manner. In the AR(2) example, for k=2,

$$Y^*_{t+2} = \hat{\delta}^* + \hat{\phi}^*_1 Y^*_{t+1} + \hat{\phi}^*_2 y_t + \hat{a}^*_{t+2} ,$$

where Y^*_{t+1} must first be calculated by (3.2). In this situation, since Y^*_{t+1} is a future value, it will vary with each bootstrap replicate, thus increasing the variance of Y^*_{t+2}. This is reasonable, since one expects the precision of any estimate of Y_{t+k}, (based on data available only up to time t) to decrease as the lag k increases.

Having obtained the set of B bootstrap future values, $(Y^{*1}_{t+k},\ldots,Y^{*B}_{t+k})$, the prediction limits are defined as quantiles of (the Monte Carlo estimate of) the bootstrap cumulative distribution function of Y^*_{t+k}. More specifically, define the bootstrap cdf of Y^*_{t+k} by

$$G^*(h) = \Pr\{Y^*_{t+k} \le h\}$$

and its Monte Carlo estimate by

$$G^*_B(h) = \#\{Y^{*b}_{t+k} \le h\}/B.$$

For a given α, a $100(1-\alpha)\%$ prediction interval for Y_{t+k} is given by

$$\left[\, L^*_B(\underset{\sim}{y}),\ U^*_B(\underset{\sim}{y}) \,\right] = \left[\, Q^*_B\left(\frac{\alpha}{2}\right), Q^*_B\left(1-\frac{\alpha}{2}\right) \right] ,$$

where $Q^*_B = G^{*-1}_B$.

The following algorithm summarizes the necessary steps for obtaining the bootstrap prediction interval.

<u>Step 1</u> Compute forward residuals $\hat{a}_j$ and backward residuals $\hat{e}_j$. Let $\hat{F}_a$ ($\hat{F}_e$) be the empirical cdf of the centered and rescaled forward (backward) residuals.

Step 2 Generate a bootstrap replicate using (2.5) and (2.6) from the backward representation.

Step 3 Compute estimates $(\hat{\delta}^*, \hat{\phi}_1^*, \ldots, \hat{\phi}_p^*)$ from the bootstrap replicate. Compute a "bootstrap future value",

$$Y_{t+k}^* = \hat{\delta}^* + \sum_{j=1}^{p} \hat{\phi}_j^* y_{t+k-j}^* + \hat{a}_{t+k}^* ,$$

with y_{t+k-j}^* computed from this relation as necessary and $\hat{a}_{t+k}^*$ a random draw from $\hat{F}_a$.

Step 4 If B bootstrap replicates (and future values) have been obtained, go to Step 5. Otherwise, repeat Steps 2-3.

Step 5 Let $G_B^*(\cdot)$ be the bootstrap CDF of Y_{t+k}^*. The endpoints of the prediction interval are given by quantiles of G_B^*.

We now present some results about the limiting behavior of the bootstrap distribution $G^*(\cdot)$ that establish the large sample validity of the proposed prediction intervals. Recall that the bootstrap procedure contributes in two ways to the problem of estimating the conditional distribution $G_{Y_{t+k}|\underline{Y}}(\cdot)$. First, the estimates $(\hat{\delta}^*, \underline{\hat{\phi}}^*)$ provide information about the distribution of $(\hat{\delta}, \underline{\hat{\phi}})$. The resampling of of $\hat{F}_a$ contributes a second source of variability in the future values Y_{t+k}^*. One would hope that as the past becomes infinite $(n \longrightarrow \infty)$ the estimates $(\hat{\delta}^*, \underline{\hat{\phi}}^*)$ converge appropriately to the true parameters $(\delta, \underline{\phi})$ and that $\hat{F}_a$ converges to F_a. This result is stated formally in the following theorem.

Theorem 3.1 Let $\{Y_j\}$ be a stationary autoregressive process with $E\{a\}=0$ and $E\{|a^{\alpha}|\} < \infty$ for some $\alpha > 2$. Let $(y_{t-n+1}, \ldots, y_t)$

denote a realization from $\{Y_j\}$. Then, along almost all sample sequences, as $n \longrightarrow \infty$,

(i) $$\begin{bmatrix} \hat{\delta}^* \\ \hat{\phi}_1^* \\ \cdot \\ \cdot \\ \cdot \\ \hat{\phi}_p^* \end{bmatrix} \longrightarrow \begin{bmatrix} \delta \\ \phi_1 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{bmatrix}$$, in conditional probability,

and

(ii) $Y^*_{t+k} \overset{d}{\rightarrow} Y_{t+k}$.

Proof:

(i) This result follows from Freedman (1985).

(ii) This is a proof by induction. We begin by showing the theorem is true for k=1. In this case

$$Y^*_{t+1} = \hat{\delta}^* + \hat{\phi}^*_1 y_t + \dots + \hat{\phi}^*_p y_{t+1-p} + \hat{a}^*_{t+1} . \tag{3.3}$$

The first term, $\hat{\delta}^*$, converges in conditional probability to δ by part (i). Consider the next p terms of (3.3). The values $(y_t,\dots,y_{t-p+1})$ are fixed and can be regarded as constants. Applying part (i) to these product terms gives $\hat{\phi}^*_j y_{t+1-j} \xrightarrow{p} \phi_j y_{t+1-j}$. Thus Y^*_{t+1} is a sum whose first (p+1) terms converge in probability to the constant $\delta+\Sigma\phi_j y_{t+1-j}$. The last term, $\hat{a}^*$, has distribution function $\hat{F}_a$, and Freedman (1985) proved $|\hat{F}_a - F_a| \xrightarrow{d} 0$, or $\hat{a}^* \xrightarrow{d} a$. It follows from Slutsky's Theorem that $Y^*_{t+1} \xrightarrow{d} Y_{t+1}$. For k=2,

$$Y^*_{t+2} = \hat{\delta}^* + \hat{\phi}^*_1 Y^*_{t+1} + \hat{\phi}^*_2 y_t + \dots + \hat{\phi}^*_p y_{t+2-p} + \hat{a}^*_{t+2} . \tag{3.4}$$

Applying Slutsky's Theorem and the result for k=1 to the term $\hat{\phi}^*_1 Y^*_{t+1}$ gives $\hat{\phi}^*_1 Y^*_{t+1} \xrightarrow{d} \phi_1 Y_{t+1}$. By the same argument as above, $Y^*_{t+2} \xrightarrow{d} Y_{t+2}$.

The induction argument completes the proof. Suppose that the theorem holds true for leads up to an arbitrary integer, say m-1. Then

$$Y^*_{t+m} = \hat{\delta}^* + \hat{\phi}^*_1 Y^*_{t+m-1} + \cdots + \hat{\phi}^*_p Y^*_{t+m-p} + \hat{a}^*_{t+m},$$

and by assumption (and Slutsky's Theorem) we have $\hat{\phi}^*_j Y^*_{t+m-j} \xrightarrow{d} \phi_j Y_{t+m-j}$ for j=1,...,p. Applying Slutsky's Theorem to each term gives the desired result, namely $Y^*_{t+m} \xrightarrow{d} Y_{t+m}$ in conditional distribution. ▒

Theorem 3.1 can be stated in different terms. It says that the bootstrap cdf of Y_{t+k} (denoted by G^* and obtained by conditionally fixing the last p values of each replicate) converges weakly to the true conditional distribution. It follows that $Q^*(u) \longrightarrow Q(u)$, (pointwise for $0<u<1$) where $Q^* = G^{*-1}$ is the theoretical bootstrap quantile function. In practice the endpoints of the interval are estimates of $Q^*(\alpha)$ and $Q^*(1-\frac{\alpha}{2})$ since a finite value of B is used. Convergence in probability of $Q^*_B(u)$ to $Q^*(u)$ as $B\to\infty$ follows from the Glivenko-Cantelli Theorem. Together these results establish large sample validity of the bootstrap interval. Namely, that the 100(1-α)% prediction limits obtained from the proposed bootstrap procedure are asymptotically correct, that is,

$$\lim_{n\to\infty} \left[\lim_{B\to\infty} \Pr \{ L^*_B(\underline{y}) \le Y_{t+k} \le U^*_B(\underline{y}) \} \right] = 1-\alpha .$$

The small sample coverage of these intervals may be improved by the accelerated, bias-corrected intervals proposed by Efron (1987) but this will not be examined here.

IV. SIMULATION RESULTS

In this section the results of a simulation study of the bootstrap prediction intervals are presented. Although asymptotically valid, the true content of the intervals for finite samples is affected by parameter estimation, the nature of the error distribution and the form of the model. Estimates of the coverage and length of the bootstrap interval (BOOT) and standard Box-Jenkins (ST) are compared. Some additional simulations for bootstrap intervals with increasing sample size provide an illustration of the theoretical results of Section 3.

Our attention here will be limited to the following models:

MODEL I: $Y_t = .95Y_{t-1} + a_t$

MODEL II: $Y_t = 1.75Y_{t-1} - .76Y_{t-2} + a_t$

MODEL III: $Y_t = -.80Y_{t-1} + a_t$

MODEL IV: $Y_t = .75Y_{t-1} - .50Y_{t-2} + a_t$

The three error distributions F_a considered are the Normal, exponential and Laplace, each centered and scaled to have zero mean and unit variance. These distributions represent the ideal, skewed and heavy-tailed symmetric alternatives, respectively. We consider sample sizes n = 50 and 100 and leads k = 1,2 and 3.

For each combination of model, sample size, error distribution and lead time, sets of M=100 realizations of an observed series are generated and summary measures of performance are calculated. To estimate the probability content and average length of the interval (and standard deviations of these estimates) we

(i) simulate a series of a specified structure, length and error distribution, and also generate R=100 true future values Y_{t+k} from that series, using F_a and the true parameter values,

(ii) use the bootstrap procedure to obtain a $100(1-\alpha)\%$ prediction interval to be denoted by (L^*,U^*) (based on B=1000) and also obtain the standard symmetric Box-Jenkins interval (L_{ST},U_{ST}) with the same nominal $1-\alpha$ content (Box and Jenkins, pp. 126-129).

(iii) Based on the simulated sample $\underline{y}_i$, estimate the coverage for each of the two methods by

$$\hat{\beta}^*_i = \#\{L^* \leq Y^r_{t+k} \leq U^*\}/R,$$

and

$$\hat{\beta}_i = \#\{L_{ST} \leq Y^r_{t+k} \leq U_{ST}\}/R,$$

where Y^r_{t+k}, $r=1,\ldots,R$, are the true future values generated in (i), and calculate the length of the intervals,

$$\hat{L}^*_i = U^* - L^*,$$

and

$$\hat{L}_i = U_{ST} - L_{ST}.$$

(iv) Repeat steps (i)-(iii) M=100 times to get a collection of summary measures $(\hat{\beta}^*_i,\hat{L}^*_i,\hat{\beta}_i,\hat{L}_i)$ $i=1,\ldots,M$, and obtain

$$\bar{\beta}^* = \Sigma\, \hat{\beta}^*_i \,/\, M\ ,$$

$$SE(\bar{\beta}^*) = \{[\Sigma(\hat{\beta}^*_i - \bar{\beta}^*)^2/(M-1)]/M\}^{1/2}\ ,$$

$$\overline{Len}^* = \Sigma\, \hat{L}^*_i \,/\, M,$$

$$SE(\overline{Len}^*) = \{[\Sigma(\hat{L}^{*2}_i - \overline{Len}^*)^2/(M-1)]/M\}^{1/2}\ ,$$

The statistics $(\bar{\beta}^*, SE(\bar{\beta}^*), \overline{Len}^*, SE(\overline{Len}^*))$ measure average coverage, variability of average coverage, average length and variability of average length for the bootstrap intervals. The measures for the standard method, $(\bar{\beta}, SE(\bar{\beta}), \overline{Len}, SE(\overline{Len}))$ are defined similarly.

Since prediction intervals are random sets designed to contain a random variable with stated probability, they can be viewed as tolerance intervals. The coverage,

$$C(\underline{y}) = \int_{L^*}^{U^*} g(\xi)d\xi$$

$$= \Pr [L^*(\underline{y}) \leq Y_{t+k} \leq U^*(\underline{y})] ,$$

is a random variable intended to satisfy

$$E[C(\underline{Y})] = \beta = 1-\alpha.$$

In this case $[L^*(\underline{y}),U^*(\underline{y})]$ is called a β-expectation tolerance interval. Note that as a result $\Pr[C(\underline{Y})\geq\beta] = \gamma \simeq .50$, approximately. In the simulations reported here, the nominal coverage $\beta = .95$ and the statistic

$$\hat{\gamma}^* = \#\{ \hat{\beta}_i^* \geq .95 \}/M$$

is computed as a second check on the desired center of the distribution of the content of the bootstrap intervals. If the true performance is close to the nominal, one would expect $\hat{\gamma}^* \simeq .50$. The sizes M=100 and R=100 for for this experiment were chosen to yield stable estimates $\bar{\beta}^*$ and $\hat{\gamma}^*$. To see this, note that $R\hat{\beta}_i^* = \#\{L^*(\underline{y}) \leq Y_{t+k}^r \leq U^*(\underline{y})\}$ is conditionally a binomial with parameter R and β_i, where $\beta_i = \Pr[L^*(\underline{y}_i) \leq Y_{t+k} \leq U^*(\underline{y}_i)]$ is the coverage for the i^{th} generated sample. The values of β_i $(i=1,\ldots,M)$ are conditional upon the sample, $\underline{y}_i$, and will vary over the M different realizations. Hence,

$$\begin{aligned} M\cdot \mathrm{Var}(\bar{\beta}^*) &= \mathrm{Var}(\hat{\beta}_i^*) \\ &= E\{\mathrm{Var}(\hat{\beta}_i^* | \beta_i)\} + \mathrm{Var}\{E(\hat{\beta}_i^* | \beta_i)\} \\ &= E\{\beta_i(1-\beta_i)/R\} + \mathrm{Var}(\beta_i) \\ &\leq \beta(1-\beta)/R + \mathrm{Var}(\beta_i), \end{aligned} \qquad (4.1)$$

by Jensen's inequality. To obtain a reasonable bound on $\mathrm{Var}(\beta_i)$ consider that ideally the $L^*(y)$ and $U^*(y)$ are order statistics of the same distribution. As such, the coverages would follow a beta distribution with parameters 950 and 51. Of course the bootstrap future values are not samples of size B from the same population and thus the variance will exceed that of the beta. The actual distribution is unimodal and concentrated between .75 and .99. Hence, a workable approximate bound on $\mathrm{Var}(\beta_i)$ is provided by a beta with parameters 90 and 10, which has variance of approximately .0009. For the nominal β the choice R=100 is large enough that the second term of (4.1) dominates. The cost of increasing M greatly exceeds the cost increasing R, and a large value of M is required to obtain a good estimate $\hat{\gamma}^*$. The values M=100 and R=100 yielded anticipated standard errors for $\bar{\beta}^*$ and $\hat{\gamma}^*$ of .004 and .05, respectively.

The bootstrap interval for $\alpha = .05$ is

$$\left[\; Y_{t+k}^{*(25)} \; , \; Y_{t+k}^{*(976)} \; \right] ,$$

where $[Y_{t+k}^{*(1)}, \ldots, Y_{t+k}^{*(1000)}]$ are the ordered bootstrap future values. The 95% normal-theory (ST) prediction interval endpoints (L_{ST}, U_{ST}) are given by

$$\hat{Y}_t(k) \pm 1.96 \left[\hat{\sigma}_a^2 \cdot \sum_{j=0}^{k-1} \hat{\psi}_j \right]^{1/2} ,$$

where $\hat{Y}_t(k) = \hat{\delta} + \Sigma\, \hat{\phi}_j \hat{Y}_t(k-j)$ and the $\hat{\psi}_j$ weights are calculated from the relation $\Psi(B) = \Phi^{-1}(B)$ (see Box and Jenkins, pp. 132-138.)

Some results are presented in Table 1 and Figures 2-5 for sample size n=50. When the error distribution is normal, the nonparametric intervals compete well with the normal theory intervals. For the Laplace distribution, the bootstrap intervals consistently have better estimated coverages

TABLE 1. Simulation results for MODELS I & II
(Nominal coverage = .95, Sample size = 50)

Model	Lead	Distri-bution	Interval	Coverage Avg	Coverage SE	Length Avg	Length SE	$\hat{\gamma}^*$
I	1	Normal	BOOT	.935	(.0036)	3.912	(.0461)	.46
			ST	.934	(.0037)	3.835	(.0419)	.42
		Expon	BOOT	.938	(.0064)	3.793	(.0953)	.56
			ST	.931	(.0046)	3.764	(.0802)	.49
		Laplace	BOOT	.932	(.0045)	4.313	(.0938)	.45
			ST	.924	(.0045)	3.861	(.0640)	.35
	3	Normal	BOOT	.900	(.0058)	5.860	(.0782)	.24
			ST	.903	(.0058)	5.990	(.0797)	.27
		Expon	BOOT	.895	(.0086)	5.787	(.1344)	.34
			ST	.903	(.0072)	5.787	(.1260)	.32
		Laplace	BOOT	.898	(.0067)	5.892	(.1021)	.23
			ST	.898	(.0067)	5.828	(.0979)	.24
II	1	Normal	BOOT	.942	(.0036)	4.181	(.0551)	.43
			ST	.941	(.0036)	4.054	(.0484)	.42
		Expon	BOOT	.949	(.0069)	4.223	(.0936)	.63
			ST	.941	(.0033)	4.030	(.0760)	.36
		Laplace	BOOT	.940	(.0041)	4.708	(.0697)	.37
			ST	.928	(.0038)	4.000	(.0565)	.25
	3	Normal	BOOT	.810	(.0163)	12.100	(.2066)	.13
			ST	.796	(.0182)	11.472	(.1835)	.09
		Expon	BOOT	.864	(.0157)	12.556	(.3334)	.22
			ST	.837	(.0129)	12.033	(.2815)	.29
		Laplace	BOOT	.829	(.0148)	12.100	(.2632)	.12
			ST	.817	(.0160)	11.107	(.2260)	.04

NOTE: ST-Standard Box Jenkins Interval
BOOT-Bootstrap Interval

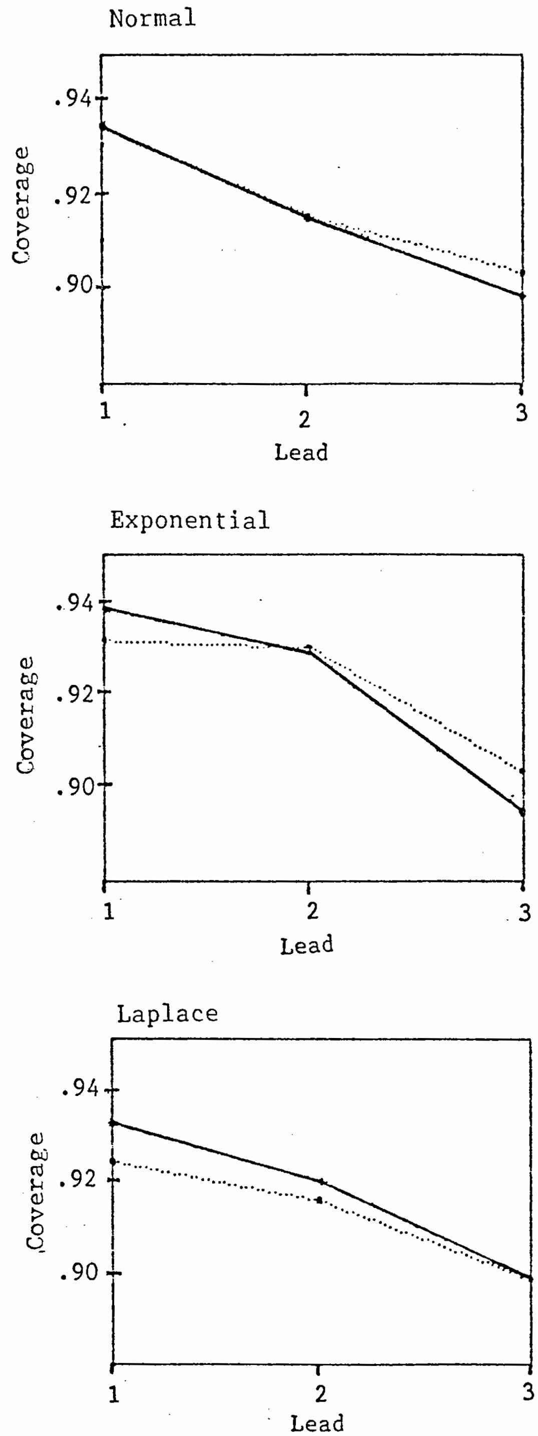

FIGURE 2
Comparison of Estimated
Coverages for the Bootstrap
and Standard Intervals.
(Model I: $Y_t = .95Y_{t-1} + a_t$)

—— Bootstrap
······· Standard

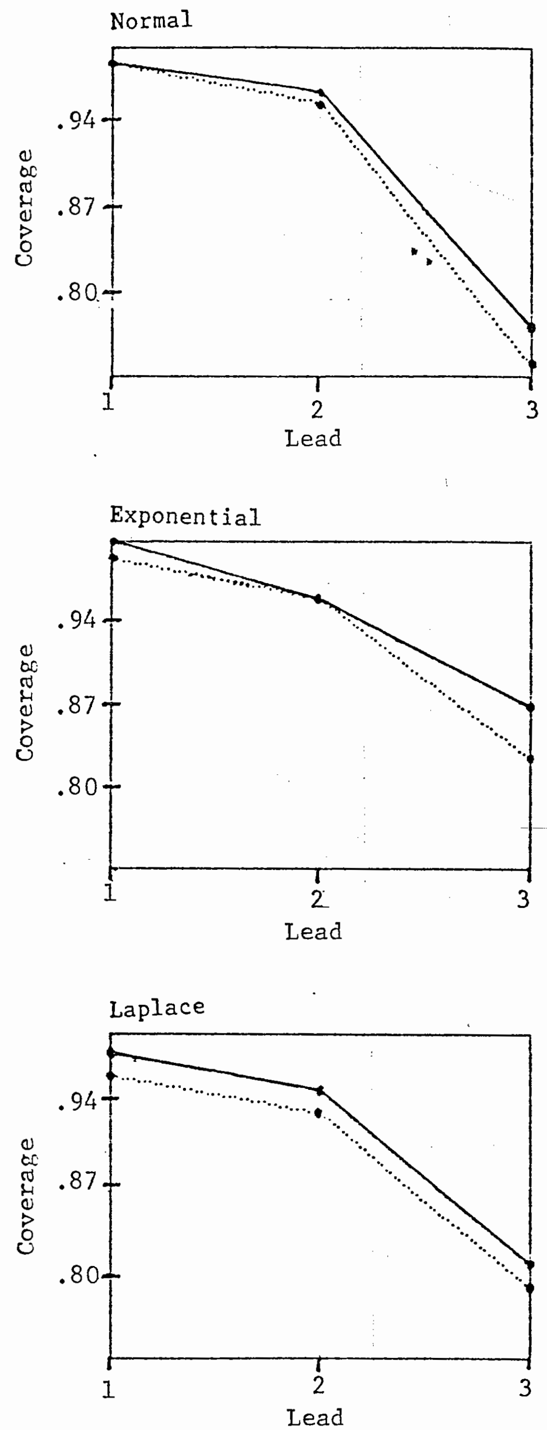

FIGURE 3
Comparison of Estimated
Coverages for the Bootstrap
and Standard Intervals.
(Model II: $Y_t = 1.75Y_{t-1} - .76Y_{t-2} + a_t$)

—— Bootstrap
······· Standard

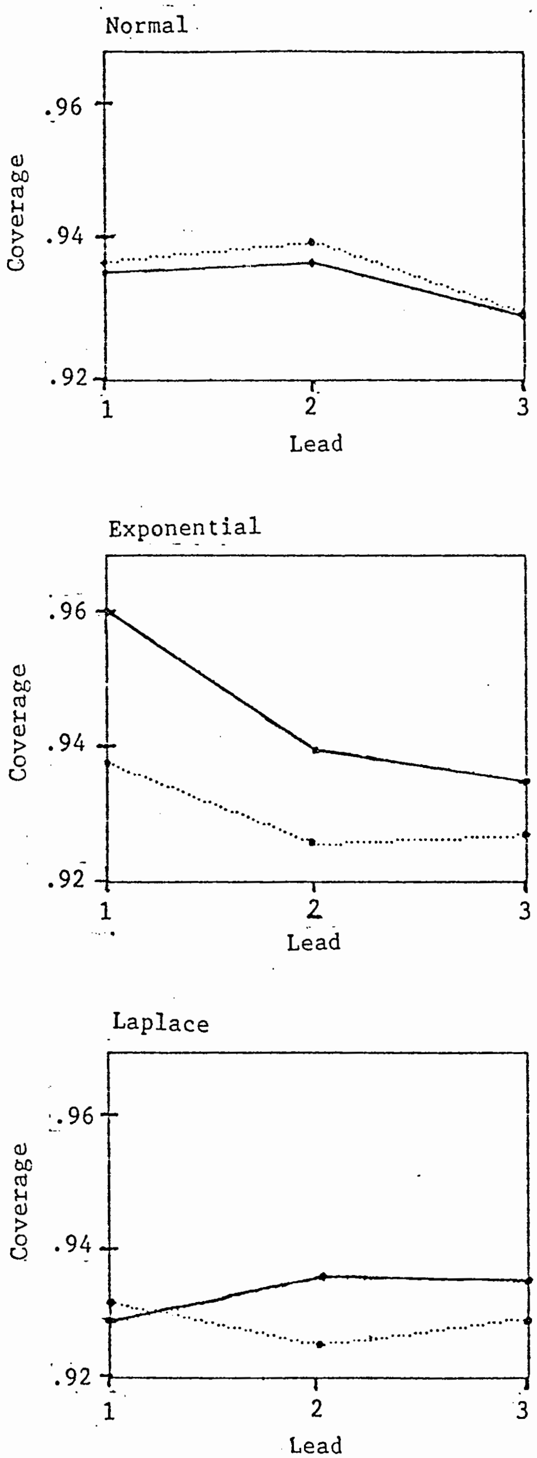

FIGURE 4
Comparison of Estimated
Coverages for the Bootstrap
and Standard Intervals
(Model III: $Y_t = -.8Y_{t-1} + a_t$)

—— Bootstrap
······· Standard

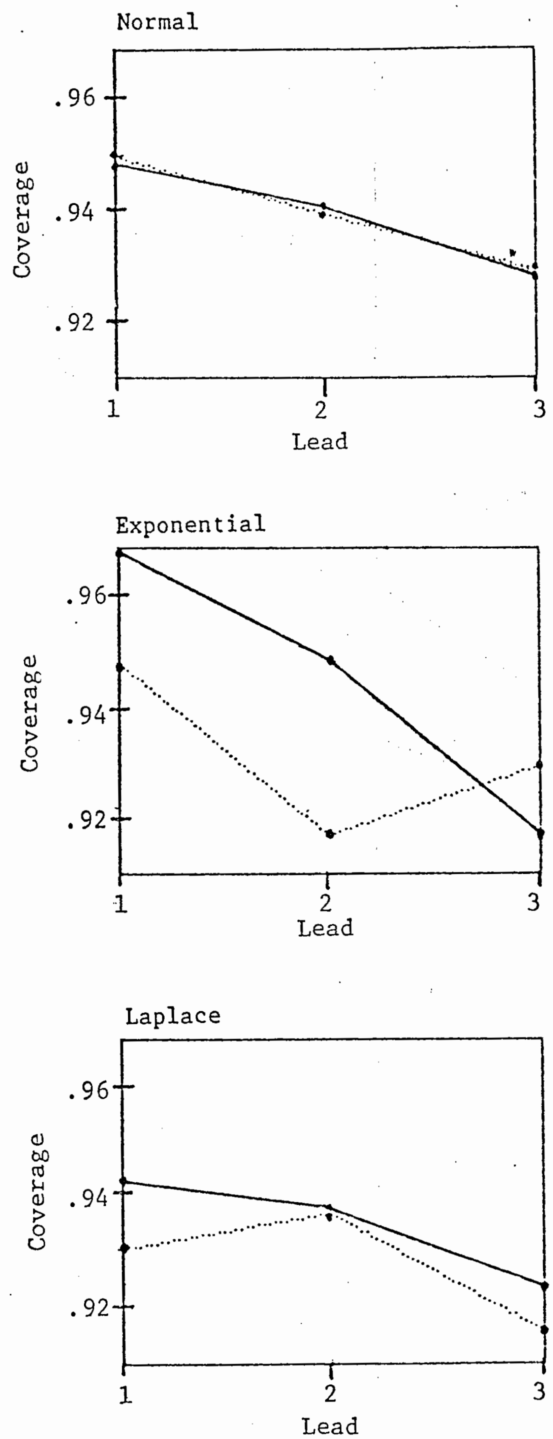

FIGURE 5
Comparison of Estimated
Coverages for the Bootstrap
and Standard Intervals.
(Model IV: $Y_t = .75Y_{t-1} - .50Y_{t-2} + a_t$)

—— Bootstrap
······· Standard

with slightly longer lengths than the standard method. Figures 2-5 clearly illustrate the negative effect of increased lead on coverage for both methods. Simulations with different sample sizes yielded similar results concerning the relative performances of the two methods.

Note that Models I and II (Figures 2 and 3) have parameters close to the region of stationarity and the parameters of Models III and IV are well within the region of stationarity. It appears that there is no significant difference (regarding estimated coverages and lengths) between the two classes of AR(p) models. Among all simulations, the difference in coverage and lengths for the two methods is most apparent for the models not near stationarity (Figures 4 and 5) with exponential error distribution and n=100.

The simulation study is designed to yield standard errors of the average coverage ($\bar{\beta}^*$) of .004. For the two symmetric error distributions the standard errors presented in Table 1 for lead-1 intervals are close to .004, and the standard errors generally increase with lead. The $\hat{\gamma}^*$ values indicate that the Beta approximation for the distribution of the coverages is appropriate for the lead-1 intervals with symmetric error distribution. Note that for the bootstrap lead-1 intervals, $\hat{\gamma}^*>.50$ and $\bar{\beta}^*<.95$, suggesting coverage distribution that is skewed.

According to the theoretical results of the previous section, the sequence $\bar{\beta}_n^*$, n=24,50,75,100, should tend towards .95, for all error distributions. This is illustrated in Table 2, which presents coverages and lengths for the lead-1 bootstrap prediction intervals. As n increases, the estimated coverage increases (and the standard deviation decreases.) The average length of the intervals does not change significantly with sample size, but variability in the lengths decreases with increased sample size.

TABLE 2. Coverages and lengths of the Lead-1 BOOTSTRAP Prediction Interval for increasing sample sizes

Model	Distribution	n	Avg	SE	Avg	SE	$\hat{\gamma}^*$
I	Normal	24	.920	(.0061)	3.933	(.0674)	.39
		50	.935	(.0036)	3.912	(.0461)	.43
		75	.936	(.0038)	3.937	(.0440)	.49
		100	.937	(.0035)	3.905	(.0359)	.45
	Expon	24	.933	(.0094)	4.321	(.1643)	.50
		50	.938	(.0064)	3.794	(.0938)	.56
		75	.960	(.0043)	3.961	(.0856)	.73
		100	.960	(.0048)	3.922	(.0559)	.62
	Laplace	24	.916	(.0061)	4.237	(.1143)	.34
		50	.932	(.0045)	4.313	(.0938)	.45
		75	.936	(.0037)	4.263	(.0707)	.46
		100	.940	(.0031)	4.193	(.0653)	.50
II	Normal	24	.902	(.0098)	4.047	(.0803)	.34
		50	.942	(.0036)	4.184	(.0551)	.43
		75	.946	(.0049)	4.169	(.0423)	.52
		100	.950	(.0029)	4.083	(.0393)	.48
	Expon	24	.903	(.0122)	3.980	(.1047)	.45
		50	.949	(.0069)	4.223	(.0936)	.63
		75	.965	(.0033)	4.032	(.0775)	.72
		100	.965	(.0042)	4.219	(.0878)	.77
	Laplace	24	.914	(.0056)	4.300	(.1110)	.24
		50	.940	(.0041)	4.708	(.0697)	.37
		75	.952	(.0029)	4.516	(.0679)	.54
		100	.953	(.0028)	4.613	(.0803)	.50

The greatest improvement in coverage occurs when n is increased from 24 to 50. The low coverages for n=24 may be due in part to the least squares estimates, which can exhibit severe bias when the series is short and the parameters are close to nonstationarity. Simulation results for other first and second order AR models were very similar to those presented in Table 2.

The proposed bootstrap methodology provides a useful nonparametric alternative to the widely used Box-Jenkins procedure. However, low coverages for some small sample simulations may indicate the need for refinement of the bootstrap intervals. A modified bootstrap prediction interval based on Efron's bias-corrected percentile interval is currently under study. Preliminary results indicate that correcting the bootstrap distribution of Y^*_{t+k} for bias can improve coverage without increasing length. Smoothing of the empirical distributions of the residuals before resampling also has the potential to improve the observed coverage. We are also investigating the performance of the bootstrap interval for ARMA(p,q) models.

BIBLIOGRAPHY

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

Efron, B. (1987). "Better Bootstrap Confidence Intervals," to appear in Journal of the American Statistical Association.

Efron, B. and Tibshirani, R. (1986) "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," Statistical Science, 1, 54-75.

Findley, D. F. (1984). "On the Use of the Bootstrap to Obtain Estimates of Mean Square Error for Multi-Step-Ahead Forecasts of Short Time Series from Autoregressive Processes," Technical Report, U.S. Census Bureau.

Freedman, D. A. (1981). "Bootstrapping Regression Models," Annals of Statistics, 9, 1218-1228.

Freedman, D. A. (1985). "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models," Annals of Statistics, 12, 827-842.

Freedman, D. A. and Peters, S. C. (1984a). "Bootstrapping an Econometric Model: Some Empirical Results," Journal of Business and Economic Statistics, 2, 150-158.

Freedman, D. A. and Peters, S.C. (1984b). "Bootstrapping a Regression Equation: Some Empirical Results," Journal of the American Statistical Association, 79, 97-106.

Stine, R. A. (1982). "Prediction Intervals for Time Series," Ph.D. Dissertation, Princeton University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: SMU/DS/TR-201	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Bootstrap Prediction Intervals for Autoregression	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Lori A. Thombs and William R. Schucany	8. CONTRACT OR GRANT NUMBER(s): N00014-85-K-0340
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Southern Methodist University, Department of Statistics, Dallas, Texas 75275	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-479
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, VA 22217	12. REPORT DATE: March 1987
	13. NUMBER OF PAGES: 24
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecasting; Non-Gaussian Time Series; Resampling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The nonparametric bootstrap is applied to the problem of prediction in autoregression. Let $\{Y_t: t=0,\pm1,\pm2,...\}$ be a stationary autoregressive process of known order p. Given a realization of the series up to time t, $(y_1,y_2,...,y_t)$, a $100(1-\alpha)\%$ prediction interval for Y_{t+k} is desired. Standard forecasting techniques, which assume that the error sequence of the process $\{Y_t\}$ is Gaussian, rely upon the fact that the conditional distribution of Y_{t+k} given the data is also Gaussian. As a nonparametric alternative, the bootstrap

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) provides an estimate of the conditional distribution of Y_{t+k}. The method is similar to other applications of the bootstrap for linear models due to the residuals being resampled. The proposed methodology represents a different approach, since an alternative representation for AR(p) series is used allowing for bootstrap replicates generated backward in time. It follows that the resulting replicates all have the same conditionally fixed values at the end of every series. A simulation which compares the proposed technique with the standard technique for low-order Gaussian and Non-Gaussian autoregressive models demonstrates the potential of the bootstrap technique.