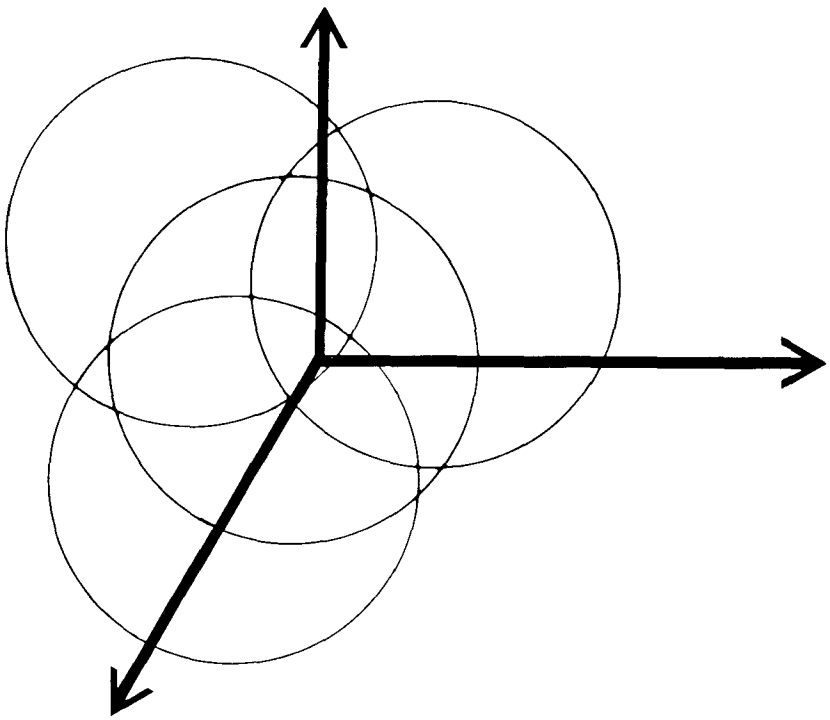

communications in statistics

SIMULATION AND COMPUTATION

COMMUN. STATIST.—SIMULA., 26(3), 961-978 (1997)

SAVING COMPUTER TIME IN CONSTRUCTING CONSISTENT BOOTSTRAP PREDICTION INTERVALS FOR AUTOREGRESSIVE PROCESSES

Cao, R.

Febrero-Bande, M., González-Manteiga, W.,

Prada-Sánchez, J.M. and García-Jurado, I.

Departamento de Matemáticas, Facultade de Informática,

Campus de Elviña, Universidade da Coruña, 15071 A Coruña

and

Departamento de Estadística e Investigación Operativa

Facultade de Matemáticas

Universidade de Santiago de Compostela

15171 Santiago de Compostela

Key Words and Phrases: smoothed bootstrap; time series.

ABSTRACT

This paper presents consistent and fast bootstrap methods for constructing nonparametric prediction intervals for autoregressive processes. These methods are compared, in a simulation study, with the Box-Jenkins approach and Thombs-Schucany's bootstrap method.

1 INTRODUCTION

One of the main aims of statistical analysis is prediction. Often, statisticians formulate a model which incorporates the essential characteristics of a real phenomenon. This model usually depends on some finite number of parameters which have to be properly estimated. This is useful not only to examine the influence of each of the variables which take part in the real situation, but also to predict the future behaviour of the real system when the values of some of these variables are known. This is of special interest in time series analysis.

In many practical situation we are given a sample of real data Y_t observed in time. Often these data are dependent upon each other, and in many practical problems they are succesfully assumed to follow an autoregressive model. In this context, our prediction problem may be the following: find a random set (an interval for instance) which covers, with a high probability, the possible values of a future observation Y_{t+k}, given the series up to time t.

The classical approaches to the problem of finding prediction intervals for autoregressive time series assume that the underlying distribution of the error process is known. Typically, this is not the case in practice and prediction intervals constructed under the assumption of a specific distribution (the normal for the Box and Jenkins (1976) procedure), may produce poor results when this condition fails.

The arguments given above have motivated some authors to introduce bootstrap methods, not only for constructing prediction intervals but also for forecasting. This is the case of Stine (1987) and Findley (1986) who used the bootstrap to estimate the mean squared error of the forecast. More connected to our problem, Thombs and Schucany (1990) propose a bootstrap resampling plan to compute prediction intervals for an autoregressive model.

Suppose we are dealing with an $AR(p)$ process. Thombs and Schucany propose to fix the last p values of the series and draw first backward and then forward bootstrap realizations: $Y^*_{t+1}, Y^*_{t+2}, \ldots, Y^*_{t+k}$. Details about the way in which this procedure is carried out will be given in Section 2.

The bootstrap prediction interval is based on the bootstrap distribution of Y^*_{t+k}. These authors prove that the bootstrap distribution is asymptotically correct and they also present a simulation study in which this bootstrap approach clearly beats Box-Jenkins intervals when the error distribution is not normal (the results for both methods are very similar for the normal case).

The major drawback of Thombs and Schucany's approach is that the algorithm is relatively time-consuming. This is due partly to the fact that we have to draw plenty of bootstrap replications (this is unavoidable in most of uses of the bootstrap), but mainly to the part of the algorithm in which every backward boostrap resample has to be modeled by an $AR(p)$ structure, and the parameters estimated. In some contexts the requirement of a fast method of predicting future values is absolutely needed. An example, involving fast prediction of concentrations of SO_2 in six tracking stations in the surroundings of a coal-fired power station in Spain, can be found in the paper by García-Jurado, González-Manteiga, Prada-Sánchez, Febrero-Bande and Cao (1995). These authors propose an extension of the Thombs and Schucany's method to the ARI case. However, the computational speed already mentioned in that paper had to be improved due to an important increase in the number of tracking stations for which the predictions were needed (from 6 to 17 tracking stations).

In this paper we present alternative consistent bootstrap methods for constructing prediction intervals for an $AR(p)$ model, designed to save computer time. These methods, as Thombs and Schucany's method, are nonparametric approaches to the problem and are still valid when the error distribution is not Gaussian. Our methods differ from Thombs and Schucany's approach in the backward resampling. We do not draw backward resamples, consequently we do not incorporate the variability coming from parameters estimation. For this reason, we refer to both methods as the *conditional bootstrap*. It is intuitive that such algorithms are much less time-consuming than Thombs and Schucany's and, even so, are proved to work asymptotically in the same sense as Thombs and Schucany did.

Section 2 contains a description of the proposed bootstraps. The correct asymptotic behaviour of these resamplings is detailed in the results presented in Section 3. Finally, Section 4 contains a small sample Monte Carlo study which compares the two bootstrap methods proposed here with Thombs and Schucany's bootstrap and the standard Box-Jenkins method.

2 BOOTSTRAP PREDICTION INTERVALS

Assume that we are given the data vector $(y_1, y_2, \ldots, y_t)$ which consists of t consecutive observations of an $AR(p)$ process:

$$Y_s = \delta + \phi_1 Y_{s-1} + \ldots + \phi_p Y_{s-p} + a_s \ , \quad s = \ldots, -2, -1, 0, 1, 2, \ldots, \qquad (1)$$

where the errors a_s are zero-mean and independent random variables with common distribution F_a and $Var(a_s) = \sigma_a^2$. Let us assume, in addition, that the roots of the polynomial $\Phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ lie outside the unit circle –in other words the process is stationary– and fix our attention in time $t + k$.

As mentioned above, there exist several methods to construct prediction intervals for the variable Y_{t+k}. We mention the widely used Box-Jenkins procedure (see Box and Jenkins (1976)) given by

$$\left(\hat{Y}_{t+k} - z_\beta \left[\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j \right]^{1/2} , \hat{Y}_{t+k} + z_\beta \left[\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j \right]^{1/2} \right), \tag{2}$$

where $\hat{Y}_{t+k} = \hat{\delta} + \sum_{j=1}^{p} \hat{\phi}_j \hat{Y}_{t+k-j}$, z_β is the two-tailed β quantile of the standard normal distribution, $\hat{\delta}$, $\hat{\sigma}_a^2$ and $\hat{\phi}_j$, $j = 1, 2, \ldots, p$ are the least squares estimates of the parameters and $\hat{\psi}_j$, $j = 1, 2, \ldots$ are the coefficients of the polynomial $\hat{\Psi}(B)$, computed from the formula $\hat{\Psi}(B) = \hat{\Phi}^{-1}(B)$.

There is a refinment of this prediction interval, given by Stine (1987), which incorporates a bias-corrected estimate of the second order term in the asymptotic formula of the prediction mean squared error. In this case, the standard normal quantile is replaced by the quantile of a t distribution with estimated degrees of freedom.

A nonparametric alternative to the intervals mentioned above are the bootstrap-based prediction intervals. They are constructed in order to produce asymptotically correct coverage rates under no parametric assumption on the error distribution. The idea behind the method is quite simple.

Let us denote by $H_k(x)$ the unconditional distribution function of the random variable $Y_{t+k} - \hat{Y}_{t+k}$ (a more plausible possibility is to deal with the distribution of the previous variable conditionally on the last p observations of the series, or even to the whole series). Assume for a moment that $H_k(x)$ were known (of course this is only a theoretical device, never true in practice), then a theoretical prediction interval at time $t + k$ would be:

$$\left(\hat{Y}_{t+k} + H_k^{-1}\left(\frac{\beta}{2}\right) , \hat{Y}_{t+k} + H_k^{-1}\left(1 - \frac{\beta}{2}\right) \right), \tag{3}$$

where

$$H_k^{-1}(p) = \inf\{x | H_k(x) \geq p\}.$$

Since the distribution function H_k is not known, one can estimate it by using bootstrap methods and then plug its quantiles in (3) to get a bootstrap prediction interval. The differences among the possible bootstrap approaches to the problem will come from the differences in the resampling plans used.

Stine's (1987) proposal proceeds as follows:

- First, draw a bootstrap initial block $(Y^*_{-p+1}, Y^*_{-p+2}, \ldots, Y^*_0)$ of p consecutive observations of the series. This is done by drawing with equiprobability one of the $t-p$ possible blocks.
- Draw the quantities a^*_s with replacement from a corrected empirical distribution function of the residuals.
- Construct further bootstrap observations using the equation:

$$Y^*_s = \hat{\delta} + \hat{\phi}_1 Y^*_{s-1} + \ldots + \hat{\phi}_p Y^*_{s-p} + a^*_s \ , \quad s = 1, 2, \ldots, t \ ,$$

  namely, the bootstrap analog of (1).
- Compute $\hat{\delta}^*$ and $\hat{\phi}^*_j$, $j = 1, 2, \ldots, p$, the least squares estimates of the parameters for the replicated series $(Y^*_1, Y^*_2, \ldots, Y^*_t)$.
- Compute the value $\hat{Y}^*_{t+k}$ in the same manner as $\hat{Y}_{t+k}$ in (2) but replacing the quantities based on the sample by their bootstrap analogs.

This resampling plan is not effective for bootstrapping prediction, although it is very appropriate for bootstrapping the parameters of the series. The reason is that it does not focus on replicating the conditional distribution of the value Y_{t+k} given the observed values (or equivalently, for the $AR(p)$ case, the last p values), but on replicating the random mechanism in parameter estimation.

On the other hand, Thombs and Schucany's (1990) bootstrap is designed to mimic the conditional distribution of the future value given $Y_{t-p+1}, \ldots, Y_t$. The method consists of the following steps:

- Fix the last p values of the series and construct the centered and rescaled backward residuals:

$$\hat{e}'_i = \left[\frac{n-p}{n-2p}\right]^{1/2}\left(\hat{e}_i - \frac{1}{n-p}\sum_{j=1}^{n-p}\hat{e}_j\right),$$

where

$$\hat{e}_i = Y_i - \hat{\delta} - \hat{\phi}_1 Y_{i+1} - \ldots - \hat{\phi}_p Y_{i+p}\ , \quad i = 1, 2, \ldots, t-p\ ,$$

- Draw backward bootstrap replications of the form

$$(Y_1^*, \ldots, Y_{t-p}^*, Y_{t-p+1}^*, \ldots, Y_t^*),$$

where

$$Y_j^* = Y_j, \text{ for } j = n-p+1, \ldots, t,$$

$$Y_j^* = \hat{\delta} + \hat{\phi}_1 Y_{j+1}^* + \ldots + \hat{\phi}_p Y_{j+p}^* + \hat{e}'^*_j, \quad j = 1, 2, \ldots, t-p$$

and $\hat{e}'^*_j$ are artificial values drawn from the empirical distribution of the backward residuals $\hat{e}'_j$, $j = 1, 2, \ldots, t-p$.

- Fit the replicated series to an $AR(p)$ model and estimate the parameters. Denote by δ^* and $\hat{\phi}_1{}^*, \ldots, \hat{\phi}_p{}^*$ the paramaters estimates of the bootstrap verision of the series.

- Draw forward bootstrap errors $\hat{a}'^*_j$ from the empirical distribution of the centered and rescaled forward residuals:

$$\hat{a}'_i = \left[\frac{n-p}{n-2p}\right]^{1/2}\left(\hat{a}_i - \frac{1}{n-p}\sum_{j=1}^{n-p}\hat{a}_j\right),$$

where

$$\hat{a}_i = Y_i - \hat{\delta} - \hat{\phi}_1 Y_{i-1} - \ldots - \hat{\phi}_p Y_{i-p}\ , \quad i = p+1, p+2, \ldots, t$$

and construct the forward bootstrap realizations $Y_{t+1}^*, Y_{t+2}^*, \ldots Y_{t+k}^*$ based on these bootstrap errors and the estimated bootstrap parameters:

$$Y_{t+j}^* = \hat{\delta} + \hat{\phi}_1 Y_{t+j-1}^* + \ldots + \hat{\phi}_p Y_{t+j-p}^* + \hat{a}'^*_{t+j}, \quad j = 1, 2, \ldots, k.$$

As already mentioned, the major drawback of Thombs and Schucany's bootstrap is the long computations needed for constructing a single bootstrap realization Y^*_{t+k}.

The methods that we present below avoid this fact by means of a completely *conditional bootstrap*. The only difference between them comes from $\hat{F}_a$, the estimator of F_a which is used in each. Once we have estimated the coefficients of the model, we only *resample* the future values $Y^*_{t+1}, \ldots, Y^*_{t+k}$. The resampling proceeds as follows:

- Estimate the parameters of the model using the least squares estimators, center the residuals, rescale them – as suggested by Stine (1987) – and construct the empirical distribution function of the residuals $\hat{F}_a$, as done in Step 4 of Thombs and Schucany's method.

- Draw the bootstrap errors a^*_s, $s = t+1, \ldots, t+k$ independently and with replacement from $\hat{F}_a$.

- Define $Y^*_s = Y_s$ for every $s = t-p+1, \ldots, t$ and compute future bootstrap observations using the equation:

$$Y^*_s = \hat{\delta} + \hat{\phi}_1 Y^*_{s-1} + \ldots + \hat{\phi}_p Y^*_{s-p} + a^*_s \ , \quad s = t+1, \ldots, t+k \ ,$$

  which are, in a sense, estimated possible future values of the series.

- The bootstrap distribution of Y^*_{t+k} is used to approximate the unknown distribution of Y_{t+k} given the observed sample. Hence, bootstrap prediction intervals are found based on this bootstrap distribution.

There is much to say about the question: *is the previous method a bootstrap method?* Indeed, it is a bootstrap method as far as it proposes to draw artificial samples from a distribution which was constructed as an estimator of some underlying unknown distribution. On the other hand, these *resamples* are only drawn for constructing *future bootstrap* realizations. In other terms, the

resampling mechanism is not used to draw bootstrap replicates of the original series but of some *future values* of the series, which are not observed at the sampling time but will be observed after some time. This is the reason why we call it the *conditional bootstrap*.

For better understanding, one may think of the method proposed here in the following way: using expression (1) one can regard Y_{t+k} as a sum of some terms depending only on the observed series and the true parameters and other terms which are linear combinations of the *future errors* a_s, with $s = t+1, \ldots, t+k$, whose coefficients depend on the unknown parameters. As a consequence, the conditional distribution of Y_{t+k} given the observed series $(Y_1, Y_2, \ldots, Y_t)$ is essentially a convolution of k distributions obtained from rescaling the error distribution F_a. The conditional distribution of Y_{t+k} would be known if the true parameters and the error distribution were known. A reasonable estimator of that conditional distribution could be defined by replacing those theoretical values and distribution by the least squares estimators of the parameters and the estimator $\hat{F}_a$ of the error distribution. Since the exact computation of this distribution is very time-consuming, in general, the method proposed here approximates it by Monte Carlo, i.e., drawing *bootstrap resamples* a_s^*, $s = t+1, \ldots, t+k$. Nevertheless, for very small k ($k = 1, 2$), the calculations of the exact *bootstrap distribution* in a computer are less time-consuming than the Monte Carlo approach.

As an alternative, we could also use a kernel estimator $\hat{F}_{a,h}$ of the error distribution function, with bandwidth h. This version only differs in the resampling plan used. If we have strong evidence that the error distribution is continuous, we may draw bootstrap errors a_s^* from a kernel density estimator, see Parzen (1962), based on the recentered and rescaled residuals $\hat{a}'_s$, $s = p+1, p+2, \ldots, t$, with a given bandwidth h. This smoothed version of the bootstrap typically has second order efficiency properties (see De Angelis and Young (1992) for a nice review and Wang (1995) for some improved version

of the smoothed bootstrap). Of course, any other information (parametric or not) about the error distribution should be incorporated in the method proposed above. For instance, if we happen to know that the error distribution is normal, we do not need to approximate the conditional distribution of Y_{t+k} given $(Y_1, Y_2, \ldots, Y_t)$, because the convolutive part of it is also normal with some variance depending on the estimated parameters.

Extensions of these conditional boostraps to the context of ARI models (with a similar spirit of that presented in the paper by García-Jurado et al (1995) for the Thombs and Schucany's method) are open problems of future research.

3 ASYMPTOTIC RESULTS

Suppose we have observed the series $(y_1, y_2, \ldots, y_t)$ from an $AR(p)$ process and we want to predict k instants ahead. The new algorithm presented in Section 2 yields a bootstrap prediction interval for Y_{t+k}. We intend to state the large-sample validity of the intervals obtained using this algorithm. Since the interval is based on the bootstrap distribution of Y^*_{t+k}, to have asymptotically correct coverage probability it suffices to prove convergence in conditional distribution of the bootstrap version Y^*_{t+k} to Y_{t+k}. The following theorem states this result.

Theorem 1 *Let Y_s be an $AR(p)$ process with $E(a_s) = 0$ and $E(|a_s|^\alpha) < \infty$ for some $\alpha > 2$. Assume in addition that the bandwidth parameter h tends to zero. Then, under either the smoothed or nonsmoothed resampling plans of the bootstrap algorithm presented in Section 2, we have that:*

$$Y^*_{t+k} \rightarrow^d Y_{t+k}$$

for almost all sample sequences.

Sketch of the proof:

First, we deal with the nonsmoothed case. For $k = 1$ the proof is exactly the same as in Theorem 3.1 of Thombs and Schucany (1990) but replacing expression (3.3) by

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

$$= \hat{\delta} + \hat{\phi}_1 Y_t + \ldots + \hat{\phi}_p Y_{t+1-p} + a^*_{t+1}.$$

Observe that in this case it is trivial that the *bootstrap* convergence in conditional probability,

$$\hat{\phi}_k Y^*_{t+1-k} \rightarrow \hat{\phi}_k Y_{t+1-k},$$

holds almost surely.

The proof continues with an induction argument in an analogous way to the proof of Thombs and Schucany's result.

The same arguments may be used in the proof of the smoothed case. The only part which is really different is the proof of the convergence in distribution,

$$\hat{a}^* \rightarrow^d a,$$

with probability one.

Denote by $\hat{\varphi}_a$, $\hat{\varphi}_{a,h}$ and φ_K the characteristic functions of $\hat{F}_a$, $\hat{F}_{a,h}$ and the kernel K, respectively and recall the convolutive structure of the kernel estimate of the error distribution (see Parzen (1962)),

$$\hat{F}_{a,h}(x) = \int_{-\infty}^{x} \hat{f}_{a,h}(t)\, dt,$$

where $\hat{f}_{a,h}(t) = \int 1/hK((t-y)/h)\hat{F}_{a,h}(dy)$. Then the following equation holds:

$$\hat{\varphi}_{a,h}(t) = \hat{\varphi}_K(ht)\hat{\varphi}_a(t).$$

Now, it becomes trivial that

$$\lim_{n\rightarrow\infty} \hat{\varphi}_{a,h}(t) = \lim_{n\rightarrow\infty} \hat{\varphi}_a(t)$$

and since convergence in distribution is equivalent to pointwise convergence of the characteristic functions, the almost sure convergence in distribution for the smoothed bootstrap follows from the unsmoothed case.

4 A SIMULATION STUDY

Some properties of the Box-Jenkins method (BJ), the Thombs and Schucany's bootstrap (TS), and this new conditional bootstrap (CB) with its smoothed version (SCB) are compared by means of a simulation study. The $AR(2)$ model given by $Y_t = 0.75Y_{t-1} - 0.5Y_{t-2} + a_t$ was considered and three different distributions for the error term a_t specified: a standard normal (N), a shifted exponential distribution on the interval $[-1, \infty)$ with scale parameter equal to one and zero mean (E) and the normal mixture consisting of a $N(-1, 1)$ with probability 0.9 and $N(9, 1)$ with probability 0.1 (M).

In every case a series of length 50 was generated from the model given above. These values correspond to the generation after an initial period of stabilization (starting at zero and neglecting the first 300 values drawn). With every sample 1, 2 and 3-lag 99% prediction intervals were constructed using the four mechanisms and their coverage probabilities and mean lengths were estimated by Monte Carlo using 1000 simulated values of the series at instants 51, 52 and 53. The number of bootstrap replications was set to 1000.

Finally, the previous process was carried out with a total number of 100 simulated series –to provide the randomness due to the conditional construction of these intervals– and the mean and standard deviation of the coverage rates and the lengths were computed. The probability that the conditional prediction intervals exceed the nominal value –denoted by γ– was estimated along these 100 simulated samples. The results are in Table I.

TABLE I

error dist.	lag	method	coverage mean	coverage std. dev.	length mean	length std. dev.	?
N	1	BJ	97.76	0.16	4.86	0.05	0.19
		TS	97.26	0.21	4.88	0.07	0.16
		CB	95.53	0.29	4.42	0.06	0.06
		SCB	98.49	0.14	5.37	0.07	0.46
N	2	BJ	97.19	0.25	5.98	0.07	0.22
		TS	97.55	0.25	6.28	0.08	0.26
		CB	97.07	0.28	6.12	0.08	0.14
		SCB	98.36	0.19	6.73	0.09	0.41
N	3	BJ	97.53	0.23	6.03	0.07	0.16
		TS	97.86	0.22	6.40	0.08	0.28
		CB	97.33	0.26	6.17	0.08	0.19
		SCB	98.54	0.20	6.80	0.09	0.47
E	1	BJ	95.85	0.25	4.72	0.10	0.03
		TS	97.97	0.22	4.97	0.14	0.38
		CB	96.13	0.46	4.50	0.13	0.24
		SCB	98.35	0.18	5.43	0.15	0.42
E	2	BJ	95.73	0.34	5.78	0.12	0.05
		TS	97.86	0.26	6.10	0.15	0.37
		CB	96.74	0.33	5.78	0.14	0.22
		SCB	98.19	0.22	6.57	0.16	0.39
E	3	BJ	95.88	0.34	5.82	0.12	0.06
		TS	97.95	0.26	6.20	0.15	0.39
		CB	97.08	0.31	5.85	0.14	0.25
		SCB	98.17	0.24	6.60	0.16	0.43
M	1	BJ	92.01	0.30	14.85	0.31	0.05
		TS	97.78	0.28	14.79	0.19	0.43
		CB	95.39	0.34	13.38	0.14	0.08
		SCB	99.10	0.20	16.45	0.20	0.80
M	2	BJ	94.08	0.59	18.15	0.40	0.12
		TS	98.13	0.27	18.96	0.35	0.43
		CB	97.00	0.53	18.43	0.35	0.34
		SCB	98.68	0.24	20.58	0.38	0.60
M	3	BJ	94.22	0.59	18.29	0.40	0.13
		TS	98.20	0.34	19.29	0.36	0.46
		CB	96.99	0.63	18.49	0.37	0.32
		SCB	98.49	0.36	20.70	0.39	0.57

Means and standard deviations of the coverage probabilities and the lengths for the 1, 2 and 3-lag 99% prediction intervals constructed using the Box-Jenkins method, Thombs and Schucany's bootstrap, the conditional bootstrap method and its smoothed version for 100 samples of series of length 50.

An important matter concerning the SCB method is the choice of the bandwidth in the resampling. The criterion we adopted was to select the bandwidth in order to minimize the mean integrated squared error when estimating the underlying distribution function by means of a smoothed version of the empirical cdf. Azzalini (1981) shows that such a minimizer is asymptotically equal to:

$$h_{AMISE} = \left[\frac{1 - \int_{-1}^{1} \left(\int_{-1}^{v} K(z)dz \right)^2 dv}{\left(\int_{-1}^{1} z^2 K(z)dz \right) n \int f'(x)^2 dx} \right]^{1/3},$$

when the kernel support is [-1,1]. In the simulation study the triangular kernel was used. The only unknown quantity in this expression is $I = \int f'(x)^2 dx$, which may be estimated, once again, using the kernel method by

$$\hat{I} = -\frac{1}{n^2 g^3} \sum_{i \neq j} K^{(2)} \left(\frac{X_i - X_j}{g} \right).$$

See Hall and Marron (1987) for details.

The asymptotically optimal choice of the pilot bandwidth g, in terms of the mean squared error, is

$$g_0 = \left[\frac{5 \int K^{(2)2} \int f^2}{n^2 \left(\int K^2 \right)^2 \left(\int f^{(2)2} \right)^2} \right]^{1/9}.$$

For the particular choice of Gaussian kernel and the specific assumption of a normal distribution, g_0 is only a function of the underlying standard deviation, namely.

$$g_0 = \left(\frac{80\pi}{3n^2} \right)^{1/9} \sigma,$$

which is estimated by replacing the standard deviation by its empirical analogue. This data-driven pilot bandwidth, $\hat{g}_0$, was used in the simulation study. This bandwidth selector is close related to the so called plug-in rule of density estimation (see, for instance, the comparative study by Cao, Cuevas and González-Manteiga (1994)). Here we only consider one step and a final normal reference. The good performance of this kind of bandwidth selection rule in the context of density estimation (see for instance Sheather and Jones (1991)) can be also expected in our context of smooth cdf estimation. Finally, using g_0 as an auxiliary bandwidth to compute $\hat{I}$ and then plugging this estimator in the expression of h_{AMISE}, one gets the bandwidth selector $\hat{h}$ that is used in the smoothed bootstrap resampling. The bootstrap residuals are obtained by

drawing observations of the recentered residuals and then adding up the value of a deviate with $N(0, \hat{h}^2)$ distribution (since the primary kernel is Gaussian).

In view of the results in Table I some remarks can be made:

- All the coverage rates summarized in the previous table are significantly different from 0.99. This is stated in the sense that the mean coverage probability can not be assumed to be equal to 0.99 (when performing a statistical testing of hypothesis about the mean at a level 0.05). Furthermore, within every simulation, the differences between every pair of the four methods are also statistically significant (at a level of 0.05).

- The smoothed conditional bootstrap (SCB) is the best method in terms of the mean and variance of the coverage rates for the three error distribution used in the simulation. This is even more evident with respect to the quantity γ, whose optimal value is 0.5. When the error is normal, the reason for this may be that the SCB is the only method that reasonably fits the length of the theoretical (non observable) Box-Jenkins prediction interval (the nominal values for the length with 1, 2 or 3 lags are 5.15, 6.8128 and 6.9334).

- Generally speaking the lengths of the intervals are similar with the only exception of SCB which produces larger intervals than the others. This feature is not surprising (at least with respect to CB) since. $\mathrm{Var}^*(\hat{a}_h^*) = \mathrm{Var}^*(a^*) + h^2 \int t^2 K(t)\, dt$. For the particular choice of the triangular kernel, used in the simulation study, we have $\mathrm{Var}^*(\hat{a}_h^*) = \mathrm{Var}^*(a^*) + h^2/6$. Typically, along the simulation study, the contribution of the smoothing term in the previous expression is not larger than 2% of the whole bootstrap variance.

- Although the bandwidth choice is an important problem when using the SCB method. the proposal made here seems to work properly independently of the error distribution chosen.

TABLE II

	TS	TS	CB	CB	SCB	SCB
	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
$b = 200$	23.10	27.93	0.03	0.03	0.06	0.10
$b = 500$	26.67	32.49	0.06	0.07	0.10	0.15
$b = 1000$	29.75	34.29	0.13	0.13	0.18	0.22

CPU times, in seconds, (on a SUN SPARC Station 10, model 30) for computing one prediction interval according to the methods TS, CB and SCB based on a series of $n = 50, 100$ observations and using $B = 200, 500, 1000$ bootstrap replications.

- The unsmoothed conditional bootstrap method (CB) is slightly worse than the Box-Jenkins approach in the normal case but better for the exponential and the mixture distribution. In comparison with TS, the performance of CB is worse. However this loss is not so large and the method may deserve some further attention in view of its enormous gain in CPU time. To illustrate this, Table II shows the CPU times (in seconds) for computing one prediction interval according to the methods TS, CB and SCB based on a series of $n = 50, 100$ observations and using $B = 200, 500, 1000$ bootstrap replications. The program was run on a SUN SPARC Station 10, model 30.

- Similar considerations can be made with respect to the methods presented in the paper by Breidt et al (1995) in the sense that for each bootstrap procedure one could define a faster method that only incorporates forward resampling.

ACKNOWLEDGEMENTS

Partially supported by the grants PB91-0794, PB94-0494, XUGA20701B92 and XUGA207003B93. The comments coming from an Associate Editor and an anonymous Referee improved this paper substantially. We thank them.

BIBLIOGRAPHY

Azzalini, A. (1981). "A note on the estimation of the distribution function and quantiles by the kernel method", *Biometrika*, 68, 326-328.

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*, San Francisco, Holden-Day.

Breidt, F.J., Davis, R.A. and Dunsmuir, W.T.M. (1995). "Improved bootstrap prediction intervals for autoregressions", *Journal of Time Series Analysis*, 16, 177-200.

Cao, R., Cuevas, A. and González-Manteiga, W. (1994). "A comparative study of several smoothing methods in density estimation", *Computational Statistics and Data Analysis*, 17, 153-176.

De Angelis, D. and Young, G.A. (1992). "Smoothing the bootstrap ", *International Statistical Review*, 60, 45-56.

Findley, D.F. (1986). "On bootstrap estimates of forecast mean square errors for autoregressive processes", in D.M. Allen, ed., *7th Symposium on the Interface*, Amsterdam, North-Holland, pp. 11-17.

García-Jurado, I., González-Manteiga, W., Prada-Sánchez, J.M., Febrero-Bande, M. and Cao, R. (1995). "Predicting using Box-Jenkins, nonparametric and bootstrap techniques", *Technometrics*, 37, 303-310.

Hall, P. and Marron, J.S. (1987). "Estimation of integrated squared density derivatives", *Statistics and Probability Letters*, 6, 109-115.

Parzen, E. (1962). "On estimating a probability density function and mode", *Annals of Mathematical Statistics*, 33, 1065-1076.

Stine, R.A. (1987). "Estimating properties of autoregressive forecasts", *Journal of the American Statistical Association*, 82, 1072-1078.

Sheather, S.J. and Jones, M.C. (1991) "A reliable data-based selection method for kernel density estimation", *Journal of the Royal Statistical Society, Series B, 53, 683-690.*

Thombs, L.A. and Schucany, W.R. (1990). "Bootstrap prediction intervals for autoregression", *Journal of the American Statistical Association*, 85, 486-492.

Wang, S. (1995). "Optimizing the smoothed bootstrap", *Annals of the Institute of Statistical Mathematics*, 47, 65-80.

Received July, 1995; Revised September, 1996.