FORECASTING TIME SERIES WITH SIEVE BOOTSTRAP

Andrés M. Alonso
Daniel Peña
Juan Romo

00-07

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-7
Statistics and Econometrics Series 1
February 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

FORECASTING TIME SERIES WITH SIEVE BOOTSTRAP

Andres M. Alonso, Daniel Peña and Juan Romo*

Abstract

In this paper we consider bootstrap methods for constructing nonparametric prediction intervals for a general class of linear processes. Our approach uses the sieve bootstrap procedure of Bühlmann (1997) based on residual resampling from an autoregressive approximation to the given process. We show that the sieve bootstrap provides consistent estimators of the conditional distribution of future values given the observed data, assuming that the order of the autoregressive approximation increases with the sample size at a suitable rate and some restrictions about polynomial decay of the coefficients $\{\psi_j\}_{j=0}^{+\infty}$ of the process MA(∞) representation. We present a Monte Carlo study comparing the finite sample properties of the sieve bootstrap with those of alternative methods. Finally, we illustrate the performance of the proposed method with real data examples.

Key Words
Sieve bootstrap; prediction intervals; time series; linear processes.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain, e-mail: alonsoam@est-econ.uc3m.es. *AMS (1990) Classification*: 62M10, 62M20, 62G09.

1. Introduction

When studying a time series, one of the main goals is the estimation of forecast confidence intervals based on an observed trajectory of the process. The traditional approach of finding prediction intervals for a linear time series assumes that the distribution of the error process is known. Thus, these prediction intervals could be adversely affected by departures from the true underlying distribution. For example, using a Monte Carlo study, Thombs and Schucany (1990) have shown that the standard (Gaussian) Box Jenkins method performs poorly given a skewed bimodal skewed error distribution.

Some bootstrap approaches have been proposed as a distribution free alternative to compute prediction intervals. Stine (1986) proposes a bootstrap method to estimate the prediction mean squared error of the estimated linear predictor of an AR(p) where p is known, assuming that the error distribution is symmetric and with finite moments. Also, for an AR(p) process with known p, and relaxing the assumptions of Stine (1987), Thombs and Schucany (1990) propose a first backward and then forward bootstrap method to find the h steps-aheads prediction intervals. Cao *et al.* (1997) study a conditional bootstrap method alternative to Thombs and Schucany's proposal, which is computationally much faster. Massarotto (1990) and Grigoletto (1998) propose a bootstrap method for AR(p) processes with finite unknown p, assuming that some consistent estimator $\widehat{p}$ is available. Pascual *et al.* (1998) generalize the conditional bootstrap approach of Cao *et al.* (1997) to ARMA(p,q) processes with known p and q and they also include the parameter estimation variability.

This paper considers bootstrap methods to construct nonparametric prediction intervals for a more general class of linear processes than those previously studied. The class of linear processes considered can be written as a one-sided infinite-order moving average process

$$X_t - \mu_X = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}, \qquad \psi_0 = 1,\ t \in \mathbb{Z}, \tag{1}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated random variables with $\mathrm{E}[\varepsilon_t] = 0$ and with at most a polynomial decay of the coefficients $\{\psi_j\}_{j=0}^{+\infty}$. This class includes the stationary and invertible ARMA(p,q) processes, since they have an exponential decay of $\{\psi_j\}_{j=0}^{+\infty}$. Our approach uses the sieve bootstrap procedure of Bühlmann (1997) based on residual resampling from a sequence of approximating autoregressive models for $\{X_t\}_{t\in\mathbb{Z}}$ with order $p = p(n)$ that increases as a function of the sample size n. This sieve bootstrap has a nice nonparametric property, being model-free within the considered class of linear processes. Thus, the proposed bootstrap prediction intervals could be applied to this more general class of linear models without specifying a finite dimensional model as in previous bootstrap proposals.

The paper is organized as follows. Section 2 introduces the sieve bootstrap for estimating forecast intervals. Section 3 establishes that the sieve bootstrap provides consistent estimators of the conditional distribution of future values X_{T+h} ($h > 0$) given the observed data. We also justify the introduction of the variability due to parameter estimation. A simulation study shows that the average coverage is better when intervals are constructed incorporating the parameter uncertainty, particularly for small sample sizes. Section 4 presents a Monte Carlo study comparing the finite sample properties of the sieve bootstrap with those of alternative methods. The simulation experiments include some nonlinear models in order to illustrate where the sieve bootstrap forecast procedure breaks down. Finally, in Section 5 the performance of the proposed method is illustrated with real data examples.

2. Sieve bootstrap forecast intervals

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real valued, stationary process with expectation $\mathrm{E}[X_t] = \mu_X$ that admits a MA(∞) representation as in (1) with $\sum_{j=0}^{+\infty}\psi_j^2 < \infty$. Under the additional assumption of invertibility we can represent $\{X_t\}_{t\in\mathbb{Z}}$ as a one-sided infinite-order autoregressive process

$$\sum_{j=0}^{+\infty}\phi_j(X_{t-j}-\mu_X) = \varepsilon_t, \qquad \phi_0 = 1,\ t\in\mathbb{Z}, \tag{2}$$

with coefficients $\{\phi_j\}_{j=0}^{+\infty}$ satisfying $\sum_{j=0}^{+\infty}\phi_j^2 < \infty$. This AR($\infty$) representation motivates Bühlmann's sieve bootstrap. The method proceeds as follows:

1. Given a sample $\{X_1, ..., X_n\}$, select the order $p = p(n)$ of the autoregressive approximation by AICC criterion: AICC $= -n\log(\sigma^2) + 2(p+1)n/(n-p-2)$, (cf. Section 9.3 of Brockwell and Davis (1991)).

The AICC criterion is a bias-corrected version of AIC (Akaike (1973)), and it has a more extreme penalty for large-order models which counteracts the overfitting nature of AIC. Other order selection criteria (such as BIC) could be used, but we prefer AICC assuming the view that the true model is complex and not of finite dimension, and also because the AICC is asymptotically efficient for autoregressive models, i.e., it chooses an AR model which achieves the optimal rate of convergence of the mean-square prediction error.

2. Construct some estimators of the autoregressive coefficients: $\widehat{\phi}_p = (\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_p)^t$. Following Bühlmann (1997) we take the Yule-Walker estimates.

3. Compute the residuals:

$$\widehat{\varepsilon}_t = \sum_{j=0}^{p}\widehat{\phi}_j(X_{t-j}-\bar{X}); \qquad \widehat{\phi}_0 = 1,\ t\in(p+1,\ldots,n). \tag{3}$$

4. Define the empirical distribution function of the centred residuals:

$$\widehat{F}_{\tilde{\varepsilon}}(x) = (n-p)^{-1} \sum_{t=p+1}^{n} 1_{\{\tilde{\varepsilon}_t \leq x\}}, \tag{4}$$

where $\tilde{\varepsilon}_t = \widehat{\varepsilon}_t - \widehat{\varepsilon}^{(\cdot)}$ and $\widehat{\varepsilon}^{(\cdot)} = (n-p)^{-1} \sum_{t=p+1}^{n} \widehat{\varepsilon}_t$.

5. Draw a resample ε_t^* of i.i.d. observations from $\widehat{F}_{\tilde{\varepsilon}}$.

6. Define X_t^* by the recursion:

$$\sum_{j=0}^{p} \widehat{\phi}_j (X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{5}$$

where the starting p observations are equals to $\bar{X}$.

In practice we generate an AR(p) resample using (5) with sample size equal to $n + 100$ and then discard the first 100 observations. For autoregressive models, other authors fix the first p observations equal to 0 or draw them with equal probability from all the $n - p + 1$ possible blocks of consecutive observations of the original series. Asymptotically, the effect of starting values is negligible (cf. Kreiss and Franke (1992)).

Up to this step, the resampling plan coincides with the sieve bootstrap, and is valid for bootstrapping some statistics defined as a functional of a m-dimensional distribution function (see details in Section 3.3 of Bühlmann (1997)). However, it is not effective for bootstrap prediction, because it does not replicate the conditional distribution of X_{T+h} given the observed data. But, if we proceed as do Cao *et al.* (1997) fixing the last p observations we can obtain resamples of the future values X_{T+h}^* given $X_{T-p+1}^* = X_{T-p+1}, \ldots, X_T^* = X_T$.

7. Compute the estimation of the autoregressive coefficients: $\widehat{\phi}_p^* = (\widehat{\phi}_1^*, \ldots, \widehat{\phi}_p^*)^t$, as in step 2.

8. Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{p} \widehat{\phi}_j^* (X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \tag{6}$$

where $h > 0$, and $X_t^* = X_t$, for $t \leq T$.

Finally, $F_{X_{T+h}^*}^*$ the bootstrap distribution of X_{T+h}^* is used to approximate the unknown distribution of X_{T+h} given the observed sample. As usual, a

Monte Carlo estimate $\widehat{F}^*_{X^*_{T+h}}$ is obtained by repeating the steps 5 to 8 B times. The $(1-\alpha)\%$ prediction interval for X_{T+h} is given by

$$[Q^*(\alpha/2), Q^*(1-\alpha/2)], \tag{7}$$

where $Q^*(\cdot) = \widehat{F}^{*-1}_{X^*_{T+h}}(\cdot)$ are the quantiles of the estimated bootstrap distribution.

Notice that, if we omit step 7 and use the $\widehat{\phi}_j$ in recursion (6), our resampling plan is similar thouth more general to the conditional bootstrap of Cao *et al.* (1997). Both approaches will be compared in the Monte Carlo study of Section 4. It is possible to modify the previous algorithm in order to incorporate the variability caused by AICC model selection adding a step where we select a new order p' from the resample obtained with recursion (5) and use it in the subsequent steps.

3. Asymptotic results

The asymptotic validity of the proposed intervals (7) depends on the limiting behavior of the distribution $F^*_{X^*_{T+h}}$, and it is sufficient to establish convergence in the conditional distribution of the bootstrap version X^*_{T+h} to X_{T+h}. Notice that the proposed bootstrap procedure has two main parts: (i) obtaining the estimates $\widehat{\phi}^*_p$ in order to have information about the distribution of $\widehat{\phi}_p$, and (ii) computing of the future values X^*_{T+h}. First, in Proposition 1, we prove the convergence in probability of $\widehat{\phi}^*_p$ to $\widehat{\phi}_p$, and in Theorem 1 we prove the large-sample validity of a conditional sieve approach.

We now consider the precise assumptions about the stationary process $\{X_t\}_{t\in\mathbb{Z}}$ required to prove our results:

> *Assumption A1:* $X_t - \mu_X = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ $(t \in \mathbb{Z})$ with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ stationary, ergodic and $\mathrm{E}[\varepsilon_t|\mathcal{F}_{t-1}] \equiv 0$, $\mathrm{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] \equiv \sigma^2 < \infty$, $\mathrm{E}[|\varepsilon_t|^s] < \infty$ for some $s \geq 4$, and $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\varepsilon_s\}_{s=-\infty}^{t-1}$.
> *Assumption A2:* $\Psi(z)$ is bounded away from zero for $|z| \leq 1$, and $\sum_{j=0}^{+\infty} j^r |\psi_j| < \infty$ for some $r \in \mathbb{N}$.

Additionally, we impose the following assumption about the autoregressive approximation:

> *Assumption B:* $p = p(n) \to \infty$, $p(n) = o(n)$ $(n \to \infty)$, and the $\widehat{\phi}_p = (\widehat{\phi}_{1,n}, \ldots \widehat{\phi}_{p,n})^t$ satisfy the empirical Yule Walker equations
>
> $$\widehat{\Gamma}_p \widehat{\phi}_p = -\widehat{\gamma}_p \tag{8}$$
>
> where $\widehat{\Gamma}_p = [\widehat{R}(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p = (\widehat{R}(1), \ldots, \widehat{R}(p))^t$, and $\widehat{R}(j) = n^{-1}\sum_{t=1}^{n-|j|}(X_t - \bar{X})(X_{t+|j|} - \bar{X})$

Our first result, proved in Appendix A, is analogous to Theorem 3.1 (a) of Thombs and Schucany (1990) about a finite autoregression. It also generalizes (for these particular statistics) Theorem 3.3 of Bühlmann (1997), since we are considering an increasing-size vector of statistics.

Proposition 1 *Suppose that assumptions A1 with $s = 4$, A2 with $r > 2$ and B with $p = o((n/\log(n))^{1/(2r+2)})$ hold. Then*

$$\max_{1\leq j\leq p(n)} |\widehat{\phi}_j^* - \widehat{\phi}_j| \xrightarrow{P^*} 0, \text{ in probability.} \tag{9}$$

The following theorem, proof in Appendix B, gives the consistency of the conditional sieve bootstrap, i.e., without step 7 in our proposed algorithm, as in Cao *et al.* (1997).

Theorem 1 *Suppose that assumptions A1 with $s = 4$, A2 with $r = 1$ and B with $p = o((n/\log(n))^{1/4})$ hold . Then*

$$X^*_{T+h} \xrightarrow{d} X_{T+h}, \text{ in probability.} \tag{10}$$

In Corollary 1, we state that the above conclusion holds for the "complete" sieve bootstrap, but under more restrictive assumptions.

Corollary 1 *Suppose that assumptions A1 with $s = 4$, A2 with $r > 2$ and B with $p = o((n/\log(n))^{1/4})$ hold. Then*

$$X^*_{T+h} \xrightarrow{d} X_{T+h}, \text{ in probability.} \tag{11}$$

4. Simulation results

First, we compare the sieve bootstrap approaches with the proposal (PRR) of Pascual *et al.* (1998). We use two sieve bootstrap prediction methods: (CS) a conditional sieve, i.e., omitting step 7 in the algorithm of Section 2, and (VS), the complete algorithm. We report the results for the following model considered in Pascual *et al.* (1998):

Model 1: $X_t = \varepsilon_t - 0.3\varepsilon_{t-1} + 0.7\varepsilon_{t-2}$.

Since method PRR use the correct model, we could interpret their results as benchmarks. In practice, having observed a sample of size n, the model, and particularly p and q, is invariably unknown. We also include the results for PRR with "wrong" but plausible model, a MA(1) model, under the heading PRR*.

The error distributions F_ε considered are the standard normal, a shifted exponential distribution with zero mean and scale parameter equal to one, and a contaminated distribution 0.9 F_1 + 0.1 F_2 with $F_1 \sim N(-1,1)$ and $F_2 \sim N(9,1)$. We take sample sizes $n = 25$, 50, and 100, leads $h = 1$, 2 and 3, and nominal coverages $1-\alpha = 0.8$ and 0.95.

To compare the different prediction intervals, we use their mean coverage and length, the proportions of observations lying out to the left and to the right of the interval and a combined measure of coverage and length. These quantities are estimated as follows:

1. For a combination of model, sample size and error distribution, simulate a series, and generate $R = 1000$ future values X_{T+h}.

2. For each bootstrap procedure obtain the $(1-\alpha)\%$ prediction interval $[Q^*_M(\alpha/2), Q^*_M(1-\alpha/2)]$ based on $B = 1000$ bootstrap resamples (M $\in$ { CS, VS, PRR, PRR* }).

3. The coverage for each method is estimated as $C_M = \#\{Q^*_M(\alpha/2) \leq X^r_{T+h} \leq Q^*_M(1-\alpha/2)\}/R$, where X^r_{T+h} with $r = 1,\ldots,R$, are the R future values generated in first step.

In steps 1 and 2 we obtain the "theoretical" and bootstrap interval lengths using $L_T = X_{T+h}^{\lceil R(1-\alpha/2)\rceil} - X_{T+h}^{\lceil R\alpha/2\rceil}$, and $L_M = Q^*_M(1-\alpha/2) - Q^*_M(\alpha/2)$. Finally, steps 1 to 3 are repeated $S = 200$ times to obtain $C_{M,i}$, $L_{M,i}$ with $i = 1,\ldots,S$, and we calculate the estimates:

$$\begin{aligned}
\bar{C}_M &= S^{-1}\sum C_{M,i} \\
SE(\bar{C}_M) &= \left(S^{-1}(S-1)^{-1}\sum(C_{M,i}-\bar{C}_M)^2\right)^{1/2} \\
\bar{L}_M &= S^{-1}\sum L_{M,i} \\
SE(\bar{L}_M) &= \left(S^{-1}(S-1)^{-1}\sum(L_{M,i}-\bar{L}_M)^2\right)^{1/2} \\
CQ_M &= |1-\bar{C}_M/\bar{C}_T| + |1-\bar{L}_M/\bar{L}_T|,
\end{aligned} \tag{12}$$

where $\bar{L}_T = S^{-1}\sum L_{T,i}$ is the estimated "true" mean interval length, $\bar{C}_T = (1-\alpha)\%$ is the nominal coverage, and $|\cdot|$ denotes the absolute value. Sometimes, when we compare two methods M_1 and M_2, the mean coverage $\bar{C}_{M_1}$ is closer to the nominal value than $\bar{C}_{M_2}$ but the corresponding mean length $\bar{L}_{M_1}$ is greater than $\bar{L}_{M_2}$, and (the important case) greater than $\bar{L}_T$. The combined measure CQ_M will be used to compare the methods in such situation.

The results for Model 1 are presented in Tables 1-3, using the three sample sizes and error distributions, nominal coverage 95%, and lead times $h = 1$ and 3. The other possible combinations of parameters are available on request to the authors. Essentially, similar results are obtained in all cases.

===> Tables 1 - 3 around here <===

For Model 1, method PRR have a better performance than CS and VS in terms of mean coverage and length, as expected. Notice that in this case, the sieve approach never uses the correct model. When comparing the sieve results with the "incorrect" model results PRR*, we observe that in terms of coverage PRR* have a generally closer mean coverage to the nominal value, but in terms of mean length this method tend to over-estimate the nominal lengths. This is clear when we use the CQ_M statistic. In terms of CQ_M, for lead time $h = 1$ the sieve bootstrap has better results than PRR*, but for $h = 3$ the opposite is observed. Bhansali (1996) shows that the AIC type criterion is not efficient for multistep prediction and proposes a modification that selects the order of the approximating autoregression depending of the lead time.

In the three tables, it is observed the better behavior of VS with respect to the conditional approach CS. Also, as expected, PRR, CS and VS improve coverage with the sample size.

As mentioned in the Introduction, the proposed method is applicable to linear models other than ARMA's. In Tables 4 and 5, we present the results for the following models:

Model 2: X_t is a Gaussian process with autocovariance generating function equal to $G(z) = \sum_{k=-\infty}^{+\infty} \gamma_k z^k$, where $\gamma_k = 1/(|k|+1)^3$.

Model 3: X_t is a Gaussian ARFIMA(0, d, 0) process, with $d = 0.5$.

===> Tables 4 and 5 around here <===

Both models are simulated using the Cholesky decomposition of the autocovariance matrix (cf. Beran (1990)). Note that Model 2 satisfies Assumption *A2* with $r = 1$ and CS is asymptotically valid, but the fractional process does not satisfy *A2* with any $r \geq 0$. However, sieve bootstrap methods perform reasonably well in Model 3 and similary for Model 2. The mean coverage and length tend to the nominal values as the sample size grows, and also here VS outperforms CS.

We now consider the following two SETAR(2,1,1) models in order to evaluate the robustness of sieve bootstrap procedure to departures of linearity:

Model 4: $X_t = (1.0 - 0.9X_{t-1} + \varepsilon_t)1_{\{X_{t-1}\leq 0\}} + (-1.0 - 0.9X_{t-1} + \varepsilon_t)1_{\{X_{t-1}>0\}}$, where ε_t are i.i.d N(0,1)

Model 5: X_t as in Model 5, where ε_t are i.i.d uniform (-1,1)

Models similar to 4 and 5 are considered by Tong (1983) to illustrate the cyclical behavior of threshold autoregressive models. In particular, Model 5 showss a cyclical movement between intervals $(-20, 0]$ and $(0, 20)$, which is less sharply defined in Model 4.

===> Tables 6 and 7 around here <===

Conditional sieve intervals CS obtain a mean coverage close to the nominal value when sample size is 50 and 100, but the mean length is generally bigger than the nominal length; this is more clear for lead time $h = 3$. In terms of CQ_M, the results for Model 5 are poorer than for Model 4, indicating that the cyclical pattern is not representable by a linear model. The VS has worse results than CS in all cases. Both methods perform poorly when sample size is 25. Note also the asymmetric proportions of below and above coverages.

5. Real data examples

In this section we illustrate the performance of sieve bootstrap procedures in three real data sets consisting of series F of Box and Jenkins (1976), which is modelled as an AR(2), Wolf's sunspot data, which is known to exhibit asymmetric cyclic behavior and it is well represented by a threshold autoregressive model, and the Nile river data, which presents a long-memory behavior. In the three cases, we compute the 1-step and multistep ahead forecasts intervals for the last ten available observations by using the sieve bootstrap (VS) and the Box-Jenkins methodology (BJ). The nominal coverage was fixed to 90%. The computations were implemented in Splus and the code is available on request to the authors.

Example 1: Series F of Box and Jenkins (1976) consists of the yields from 70 consecutive batches of a chemical process. Figures 1 and 2 show the 1-step and multistep prediction intervals, for the Box-Jenkins approach using Gaussian maximum likelihood estimates for an AR(2) process (see page 239 in Box and Jenkins (1976)). The upper limits of the intervals are very similar, but in lower limits is observed a downward shift of bootstrap limits revealing the asymmetric distribution of the residuals. The estimated residual skewness is -0.459 with a 95% confidence interval (-1.130, 0.006) and the residual kurtosis is 3.295 with a 95% confidence interval (2.435, 5.235), both constructed by BCA bootstrap method.

===> Figures 1 and 2 around here <===

Example 2: The Wolf annual sunspot index in the period 1700 - 1979. We consider a short series consisting of the period 1769 - 1869 which is the series E of Box and Jenkins (1976) identified as an AR(2) or an AR(3) process, and the complete series 1700 - 1979 analyzed by Tong (1983) with a threshold model. Priestley (1989) reviews different linear and nonlinear models considered for these series.

Figures 3 and 4 show the 1-step and multistep prediction intervals for series E. For the BJ intervals we use an AR(2) model (the AR(3) model provided similar results). Before estimating the model, the effects of several outliers have been removed from the original series using the SCA package (see Liu and Hudak (1992)). In Figure 3, we observe a upward shift of the upper and lower bootstrap limits, and in Figure 4 it is observed in lower bootstrap limits. For multistep prediction the VS intervals are narrower than BJ intervals. The estimated residual skewness is 0.546 with a 95% confidence interval (0.176, 0.939) and the residual kurtosis is 2.849 with a 95% confidence interval (2.287, 4.109).

===> Figures 3 and 4 around here <===

Figures 5 and 6 show the results for the longer series using an AR(9) in the BJ intervals. The shape and length of prediction intervals are similar revealing that the AR(9) residuals follow a Gaussian distribution. As in Priestley (1989), an ARMA(6,6) model was considered, but for 1-step prediction four observations lye out of the intervals and the multistep intervals are wider than for the AR(9) model. The estimated residual skewness is 0.748 with a 95% confidence interval (-0.236, 1.699) and the residual kurtosis is 7.553 with a 95% confidence interval (4.042, 10.984). For the this series, the detection of outliers is more complicated than for the previous one, since once you have corrected some outliers effects, then some others outliers appears. This fact could indicate the inadequacy of a linear model for this series.

===> Figures 5 and 6 around here <===

Example 3: The yearly minimal water levels of the Nile river for the years 622 - 1281, measured at the Roda Gauge near Cairo, which is typically modelled as a long-memory process (cf. Beran (1994)). For the BJ intervals we use an AR(7) model selected by AIC. The results for both methods are similar, only a little upward shift of the bootstrap intervals is observed. The estimated residual skewness is 0.245 with a 95% confidence interval (0.095, 0.4065) and the residual kurtosis is 3.078 with a 95% confidence interval (2.816, 3.363), which is not a big departure from the Gaussian distribution.

===> Figures 7 and 8 around here <===

Acknowledgements

We would like to thank Mike Wiper for carefully reading that greatly improved the paper. This research was partially supported by the Dirección General de Educación Superior project DGES PB96-0111.

Appendix A

We use the following lemma of Bühlmann (1997):

Lemma 1 *Let assumptions A1 with* $s = \max\{2w, 4\}$*, A2 with* $r = 0$ *and B with* $p = o((n/\log(n))^{1/2})$ *hold. Then*

$$E^*[\varepsilon_t^{*2w}] = E[\varepsilon_t^{2w}] + o_P(1). \tag{13}$$

Proof of Proposition 1: The vector $\widehat{\phi}_p^*$ is defined by the bootstrap empirical Yule Walker equations:

$$\widehat{\Gamma}_p^* \widehat{\phi}_p^* = -\widehat{\gamma}_p^*, \tag{14}$$

where $\widehat{\Gamma}_p^* = [\widehat{R}^*(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p^* = (\widehat{R}^*(1), \ldots, \widehat{R}^*(p))^t$, and $\widehat{R}^*(j) = n^{-1} \sum_{t=1}^{n-|j|}(X_t^* - \bar{X}^*)(X_{t+|j|}^* - \bar{X}^*)$. Then

$$\begin{aligned} \|\widehat{\phi}_p^* - \widehat{\phi}_p\|_\infty &= \| \left(\widehat{\Gamma}_p^{-1} - \widehat{\Gamma}_p^{*-1}\right) \widehat{\gamma}_p^* + \widehat{\Gamma}_p^{-1} \left(\widehat{\gamma}_p - \widehat{\gamma}_p^*\right) \|_\infty \\ &\leq \|\widehat{\Gamma}_p^{*-1} - \widehat{\Gamma}_p^{-1}\|_{row} \|\widehat{\gamma}_p^*\|_\infty + \|\widehat{\Gamma}_p^{-1}\|_{row} \|\widehat{\gamma}_p^* - \widehat{\gamma}_p\|_\infty, \end{aligned} \tag{15}$$

where $\|x\|_\infty = \max_{1 \leq i \leq p} |x_i|$, and $\|X\|_{row} = \max_{1 \leq i \leq p} \sum_{j=1}^p |X_{i,j}|$.

From Theorem 2.1 of Hannan and Kavalieris (1986), we have that $\|\widehat{\Gamma}_p\|_{row}$ and $\|\widehat{\Gamma}_p^{-1}\|_{row}$ are bounded. Since $\widehat{\Gamma}_p^{-1} - \widehat{\Gamma}_p^{*-1} = \widehat{\Gamma}_p^{-1}\left(\widehat{\Gamma}_p^* - \widehat{\Gamma}_p\right)\widehat{\Gamma}_p^{*-1}$, and $\|\widehat{\Gamma}_p^* - \widehat{\Gamma}_p\|_{row} \leq |\widehat{\gamma}_0^* - \widehat{\gamma}_0| + 2\|\widehat{\gamma}_p^* - \widehat{\gamma}_p\|_1$, we can concentrate our attention on this last term:

$$\|\widehat{\gamma}_p^* - \widehat{\gamma}_p\|_1 \leq \|\widehat{\gamma}_p^* - \gamma_p\|_1 + \|\gamma_p - \widehat{\gamma}_p\|_1. \tag{16}$$

From Theorem 3 of An *et al.* (1982), the second summand is $O_{a.s.}((n/\log(n))^{-r/(2r+2)})$.

Since

$$\|\widehat{\gamma}_p^* - \gamma_p\|_\infty \leq \|\widehat{\gamma}_p^* - \gamma_p\|_1 \leq p^{1/2}\|\widehat{\gamma}_p^* - \gamma_p\|_2, \tag{17}$$

to get convergence to zero in (15), it is enough to consider the last term in (17).

$$\begin{aligned}
\|\widehat{\gamma}_p^* - \gamma_p\|_2^2 &= \textstyle\sum_{k=1}^{p}(\widehat{R}^*(k) - R(k))^2 \\
&\le \textstyle\sum_{k=1}^{p}(\widehat{R}^*(k) - \mathrm{E}^*[\widehat{R}^*(k)])^2 + \sum_{k=1}^{p}(\mathrm{E}^*[\widehat{R}^*(k)] - R(k))^2 \\
&= S_1 + S_2
\end{aligned} \tag{18}$$

But $S_2 = O_P((n/\log(n))^{-(2r-3)/(2r+2)})$, since

$$S_2 = \textstyle\sum_{k=1}^{p}\left(\mathrm{E}^*[\varepsilon_1^{*2}]\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\delta_{i+k,j} - \mathrm{E}[\varepsilon_1^2]\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\psi_i\psi_j\delta_{i+k,j}\right)^2 \tag{19}$$

where $\delta_{i,j} = 1$ if $i = j$, and 0 otherwise, and $\widehat{\Psi}(z) = \sum_{i=0}^{+\infty}\psi_{i,n}z^i = \widehat{\Phi}(z)^{-1}$ which is well defined because $\widehat{\Phi}(z)$ is always causal (cf. Brockwell and Davis (1991)). Now,

$$\begin{aligned}
S_2 &= \textstyle\sum_{k=1}^{p}\Big(\mathrm{E}^*[\varepsilon_1^{*2}]\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}(\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \psi_i\psi_j)\delta_{i+k,j} \\
&\qquad\qquad + (\mathrm{E}^*[\varepsilon_1^{*2}] - \mathrm{E}[\varepsilon_1^2])\textstyle\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\psi_i\psi_j\delta_{i+k,j}\Big)^2 \\
&\le 2\textstyle\sum_{k=1}^{p}\left(\mathrm{E}^*[\varepsilon_1^{*2}]\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}(\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \psi_i\psi_j)\delta_{i+k,j}\right)^2 \\
&\quad + 2\textstyle\sum_{k=1}^{p}\left((\mathrm{E}^*[\varepsilon_1^{*2}] - \mathrm{E}[\varepsilon_1^2])\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\psi_i\psi_j\delta_{i+k,j}\right)^2 = I_1 + I_2
\end{aligned} \tag{20}$$

Theorem 3.1 and 3.2 of Bühlmann (1995) establishes the following results:

$$\sup_{i\in\mathbb{N}}|\widehat{\psi}_{i,n} - \psi_i| = O_{a.s.}((\log(n)/n)^{1/2}) + O_{a.s.}(p^{-r}) \tag{21}$$

and

$$\sup_{n\ge n_1}\sum_{i=0}^{+\infty} i^r|\widehat{\psi}_{i,n}| = O_{a.s.}(1), \tag{22}$$

where n_1 is a random variable.

Using the above results, we have that $I_1 = O_P((n/\log(n))^{-(2r-3)/(2r+2)})$, since

$$\begin{aligned}
I_1 &\le 2\mathrm{E}^*[\varepsilon_1^{*2}]^2 p\left(\textstyle\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}|\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \psi_i\psi_j|\right)^2 \\
&\le 2\mathrm{E}^*[\varepsilon_1^{*2}]^2 p\left(\textstyle\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}|\widehat{\psi}_{i,n}\psi_j - \psi_i\psi_j| + \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}|\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \widehat{\psi}_{i,n}\psi_j|\right)^2 \\
&= O_P(p)(O_{a.s.}((\log(n)/n)^{1/2}p) + O_{a.s.}(p^{-r+1}) + o_{a.s.}(p^{-r}))^2 \\
&= O_P((n/\log(n))^{-(2r-3)/(2r+2)})
\end{aligned} \tag{23}$$

Under assumptions A1 and B of this proposition, we can establish a stronger conclusion in Lemma 1, in fact

$$\mathrm{E}^*[\varepsilon_t^{*2}] - \mathrm{E}[\varepsilon_t^2] = o_P((\log(n)/n)^{1/2}p). \tag{24}$$

Therefore,

$$I_2 = o_P((\log(n)/n)p^3) = o_P((n/\log(n))^{-(2r-1)/(2r+2)}). \tag{25}$$

For the other term in (18), we have $S_1 = O_P(n^{-1}(n/\log(n))^{1/(2r+2)})$, since

$$\begin{aligned} S_1 &= \sum_{k=1}^{p} \Big(n^{-1} \sum_{t=1}^{n-k} \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\varepsilon_{t-i}^*\varepsilon_{t+k-j}^* \\ &\qquad - \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\mathrm{E}^*[\varepsilon_1^{*2}]\delta_{i+k,j}\Big)^2 \\ &= \sum_{k=1}^{p} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j=0}^{+\infty} \sum_{h,l=0}^{+\infty} \widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\widehat{\psi}_{h,n}\widehat{\psi}_{l,n} \left(\varepsilon_{t-i}^*\varepsilon_{t+k-j}^* - \mathrm{E}^*[\varepsilon_1^{*2}]\delta_{i+k,j}\right) \\ &\qquad \left(\varepsilon_{s-h}^*\varepsilon_{s+k-l}^* - \mathrm{E}^*[\varepsilon_1^{*2}]\delta_{h+k,l}\right). \end{aligned} \tag{26}$$

Taking E^* in the above expression, we have,

$$\begin{aligned} \mathrm{E}^*[S_1] &= \sum_{k=1}^{p} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j=0}^{+\infty} \sum_{h,l=0}^{+\infty} \widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\widehat{\psi}_{h,n}\widehat{\psi}_{l,n}\big(\mathrm{E}^*[\varepsilon_{t-i}^*\varepsilon_{t+k-j}^*\varepsilon_{s-h}^*\varepsilon_{s+k-l}^*] \\ &\qquad -\mathrm{E}^*[\varepsilon_1^{*2}]^2\delta_{i+k,j}\delta_{h+k,l}\big). \end{aligned} \tag{27}$$

Notice that

$$\mathrm{E}^*[\varepsilon_{t-i}^*\varepsilon_{t+k-j}^*\varepsilon_{s-h}^*\varepsilon_{s+k-l}^*] = \begin{cases} \mathrm{E}^*[\varepsilon_1^{*4}] & \text{if } t-i=t+k-j=s-h=s+k-l \\ \mathrm{E}^*[\varepsilon_1^{*2}]^2 & \text{if two pairs of different indices} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

and

$$\mathrm{E}^*[\varepsilon_{t-i}^*\varepsilon_{t+k-j}^*\varepsilon_{s-h}^*\varepsilon_{s+k-l}^*] - \mathrm{E}^*[\varepsilon_1^{*2}]^2\delta_{i+k,j}\delta_{h+k,l} \tag{29}$$

$$= \begin{cases} \mathrm{E}^*[\varepsilon_1^{*4}] - \mathrm{E}^*[\varepsilon_1^{*2}]^2 & \text{if } t-i=t+k-j=s-h=s+k-l \\ 0 & \text{if } t-i=t+k-j\neq s-h=s+k-l \\ \mathrm{E}^*[\varepsilon_1^{*2}]^2 & \text{if } t-i=s-h\neq t+k-j=s+k-l \\ & \text{or } t-i=s+k-l\neq s-h=t+k-j \\ 0 & \text{otherwise.} \end{cases}$$

Because of Theorem 3.2 of Bühlmann (1995), we have for some random variable n_1 that $\sup_{n\geq n_1} \sum_{i,j,h,l=0}^{+\infty} \widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\widehat{\psi}_{h,n}\widehat{\psi}_{l,n} = O_{a.s.}(1)$. On the other hand, in (27) when we fix the indices i , j, h and l, the sum $\sum_{t,s=1}^{n-k}(\cdot)$ includes at most $n-k$ nonzero summands. Then, $\mathrm{E}^*[S_1] = O_P(pn^{-1})$.

Finally, we have

$$p^{1/2}\|\widehat{\gamma}_p^* - \gamma_p\|_2 = O_P((n/\log(n))^{-(r-2)/(r+1)}) \tag{30}$$

and the assumption A2 with $r > 2$ concludes the proof.■

Appendix B

We use the following lemma of Bühlmann (1997):

Lemma 2 *Let assumptions A1 with $s = 4$, A2 with $r = 1$ and B with $p = o((n/\log(n))^{1/2})$. Then*

$$\varepsilon_t^* \xrightarrow{d} \varepsilon_t, \text{ in probability.} \tag{31}$$

Proof of Theorem 1: We can write X_{T+h} andX_{T+h}^* as:

$$X_{T+h} = -\sum_{j=1}^{+\infty} \phi_j X_{T+h-j} + \varepsilon_{T+h} \tag{32}$$

$$X_{T+h}^* = -\sum_{j=1}^{+\infty} \widehat{\phi}_{j,n} X_{T+h-j}^* + \varepsilon_{T+h}^* \tag{33}$$

where $\widehat{\phi}_{j,n}$ denote the estimates of ϕ_j with a sample of size n: $(X_{T-n+1}, \ldots, X_T)$, $\widehat{\phi}_{j,n} = 0$ for $j > p(n)$, and $X_t^* = X_t$ for $t \leq T$.

For simplicity of notation we prove the theorem for $h = 1$ and $h = 2$:

$$\begin{aligned} X_{T+1}^* - X_{T+1} &= -\sum_{j=1}^{+\infty}(\widehat{\phi}_{j,n} - \phi_j)X_{T+1-j} + \varepsilon_{T+1}^* - \varepsilon_{T+1} \\ &= -\sum_{j=1}^{p(n)}(\widehat{\phi}_{j,n} - \phi_j)X_{T+1-j} + \sum_{j=p(n)+1}^{+\infty} \phi_j X_{T+1-j} + \varepsilon_{T+1}^* - \varepsilon_{T+1} \\ &= S_{1,1} + S_{2,1} + S_{3,1} \end{aligned} \tag{34}$$

From Lemma 2, we have $S_{3,1} \xrightarrow{d} 0$, in probability. Also,

$S_{2,1} = o_P((n/\log(n))^{-r/(2r+2)})$, since

$$\mathrm{E}[|S_{2,1}|] \leq \mathrm{E}[|X_t|] \sum_{j=p(n)+1}^{+\infty} |\phi_j| = o(p^{-r}) \tag{35}$$

and third, we establish that $S_{1,1} = O_P((n/\log(n))^{-r/(2r+2)})$. We have that

$$|S_{1,1}| \leq |\sum_{j=1}^{p(n)}(\widehat{\phi}_{j,n} - \phi_{j,n})X_{T+1-j}| + |\sum_{j=1}^{p(n)}(\phi_{j,n} - \phi_j)X_{T+1-j}| = I_1 + I_2 \quad (36)$$

where $\phi_p = (\phi_{1,n}, \ldots, \phi_{p,n})^t$ are defined by the theoretical Yule-Walker equation $\Gamma_p\phi_p = -\gamma_p$.

For I_1 we use the result in Theorem 2.1 of Hannan and Kavalieris (1986):

$$\max_{1\leq j\leq p} |\widehat{\phi}_{j,n} - \phi_{j,n}| = O_{a.s.}((\log(n)/n)^{1/2}). \quad (37)$$

Therefore,

$$\begin{aligned} I_1 &\leq \left(\sum_{j=1}^{p(n)}(\widehat{\phi}_{j,n} - \phi_{j,n})^2\right)^{1/2}\left(\sum_{j=1}^{p(n)} X_{T+1-j}^2\right)^{1/2} \\ &\leq p(n)^{1/2}\max_{1\leq j\leq p}|\widehat{\phi}_{j,n} - \phi_{j,n}|O_P(p(n)^{1/2}) = O_P(p(n)(\log(n)/n)^{1/2}) \\ &= O_P((n/\log(n))^{-r/(2r+2)}). \end{aligned} \quad (38)$$

For I_2 we use the extended Baxter inequality (cf. Hannan and Deistler (1988)):

$$\sum_{j=0}^{+\infty}|\phi_{j,n} - \phi_j| \leq c\sum_{j=p(n)+1}^{+\infty}|\phi_j| \quad (39)$$

where c is a constant depending on the true structure. Therefore,

$$\mathrm{E}[I_2] \leq \mathrm{E}[|X_t|]\sum_{j=1}^{p(n)}|\phi_{j,n} - \phi_j| = o(p^{-r}). \quad (40)$$

Finally, $X_{T+1}^* - X_{T+1} = S_{1,1} + S_{2,1} + S_{3,1} \xrightarrow{d} 0$ in probability.

For $h = 2$, we can write X_{T+2} and X_{T+1}^* as:

$$\begin{aligned} X_{T+2} &= -\sum_{j=1}^{+\infty}\phi_j X_{T+2-j} + \varepsilon_{T+2} \\ &= \phi_1\sum_{j=1}^{+\infty}\phi_j X_{T+1-j} - \sum_{j=2}^{+\infty}\phi_j X_{T+2-j} + \varepsilon_{T+2} - \phi_1\varepsilon_{T+1} \\ &= \sum_{j=1}^{+\infty}(\phi_1\phi_j - \phi_{j+1})X_{T+1-j} + \varepsilon_{T+2} - \phi_1\varepsilon_{T+1} \end{aligned} \quad (41)$$

$$X_{T+2}^* = \sum_{j=1}^{p(n)}(\widehat{\phi}_{1,n}\widehat{\phi}_{j,n} - \widehat{\phi}_{j+1,n})X_{T+1-j} + \varepsilon_{T+2}^* - \widehat{\phi}_{1,n}\varepsilon_{T+1}^* \quad (42)$$

As in the case $h = 1$, we have:

$$\begin{aligned} X^*_{T+2} - X_{T+2} &= \sum_{j=1}^{p(n)} (\widehat{\phi}_{1,n}\widehat{\phi}_{j,n} - \widehat{\phi}_{j+1,n} - \phi_1\phi_j + \phi_{j+1}) X_{T+1-j} \\ &\quad - \sum_{j=p(n)+1}^{\infty} (\phi_1\phi_j - \phi_{j+1}) X_{T+1-j} \\ &\quad + \varepsilon^*_{T+2} - \widehat{\phi}_{1,n}\varepsilon^*_{T+1} - \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1} \\ &= S_{1,2} - S_{2,2} + S_{3,2}. \end{aligned} \tag{43}$$

From Lemma 2, and the independence of ε^*_{T+1} and ε^*_{T+2}, we have $S_{3,2} \xrightarrow{d} 0$ in probability. Moreover,

$$S_{2,2} = O_P((n/\log(n))^{-r/(2r+2)}), \tag{44}$$

since

$$S_{2,2} = \phi_1 S_{2,1} + \sum_{j=p(n)+1}^{\infty} \phi_{j+1} X_{T+1-j} \tag{45}$$

and for the last summand, we proceed as for $S_{2,1}$.

Also, $S_{1,2} = O_P((n/\log(n))^{-r/(2r+2)})$, since

$$\begin{aligned} S_{1,2} &= \widehat{\phi}_{1,n} \sum_{j=1}^{p(n)} (\widehat{\phi}_{j,n} - \phi_j) X_{T+1-j} + (\widehat{\phi}_{1,n} - \phi_1) \sum_{j=1}^{p(n)} \phi_j X_{T+1-j} \\ &\quad - \sum_{j=1}^{p(n)} (\widehat{\phi}_{j+1,n} - \phi_{j+1}) X_{T+1-j} \\ &= -\widehat{\phi}_{1,n} S_{1,1} + I_3 + I_4 \end{aligned} \tag{46}$$

and $I_3 = O_P(n^{-1/2})$ and with I_4 we proceed as with $S_{1,1}$.

Finally, $X^*_{T+2} - X_{T+2} = S_{1,2} + S_{2,2} + S_{3,2} \xrightarrow{d} 0$ in probability.

For general h, it is clear that we could write $X^*_{T+h} - X_{T+h}$ as a sum of some function $f(\phi_1, \ldots, \phi_{h-1}, \widehat{\phi}_{1,n}, \ldots, \widehat{\phi}_{h-1,n})(S_{1,1} + S_{2,1})$, a term similar to $S_{1,1} + S_{2,1}$, and a "linear" combination of the corresponding (and independent) errors $(\varepsilon_{T+1}, \ldots, \varepsilon_{T+h}, \varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+h})$.■

Bibliography

An, H.-Z., Chen, Z.-G. and Hannan, E.J. (1982) Autocorrelation, autoregression and autoregressive approximations, *Ann. Statist.*, **10**, 926-936.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, (B.N. Petrov, and F. Csaki, Eds.), Akademiai Kiado, Budapest, 267-281.

Beran, J. (1994) *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Bhansali, R.J. (1996) Asymptotically efficient autoregressive models selection for multistep prediction, *Ann. Inst. Statist. Math.*, **48**, 577-602.

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

Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*, Springer-Verlag, New York.

Bühlmann, P. (1995) Moving-average representation of autoregressive approximations, *Stochastic Process. Appl.*, **60**, 331-342.

Bühlmann, P. (1997) Sieve bootstrap for time series, *Bernoulli*, **3**, 123-148.

Cao, R., Febrero-Bande, M., González-Manteiga, W., Prada-Sánchez, J.M. and García-Jurado, I. (1997) Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes, *Comm. Statist. A*, **26**, 961-978.

Grigoletto, M. (1998) Bootstrap prediction intervals for autoregressions: some alternatives, *International J. Forecasting*, **14**, 447-456.

Hannan, E.J. and Deistler, M. (1988) *The Statistical Theory of Linear Systems*, John Wile & Sons, Inc., New York.

Hannan, E.J. and Kavalieris, L. (1986) Regressions, autoregression models, *J. Time Ser. Anal.*, **7**, 27-49.

Kreiss, J-P. and Franke, J. (1992) Bootstrapping stationary autoregressive moving-average models, *J. Time Ser. Anal.*, **13**, 297-317.

Liu, L-M. and Hubak, G.B. (1992) *Forecasting and Time Series Analysis Using the SCA Statistical System*, Scientific Computing Associates Corp., Illinois.

Masarotto, G. (1990) Bootstrap prediction intervals for autoregressions, *International J. Forecasting*, **6**, 229-239.

Pascual, L., Romo, J. and Ruiz, E. (1998) Bootstrap predictive inference for ARIMA processes, Working Paper 98-86, Universidad Carlos III de Madrid.

Priestley, M.B. (1989) *Spectral Analysis and Time Series*, Vol. 1, Academic Press Inc., San Diego.

Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear model, *Ann. Statist.*, **8**, 147-164.

Stine, R.A. (1987) Estimating properties of autoregressive forecasts, *J. Amer. Statist. Assoc.*, **82**, 1072-1078.

Thombs, L.A. and Schucany, W.R. (1990) Bootstrap prediction intervals for autoregression, *J. Amer. Statist. Assoc.*, **85**, 486-492.

Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*, Springer Verlag Inc., New York.

Figure 1: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for 1-step ahead prediction in series F of Box and Jenkins (1976).

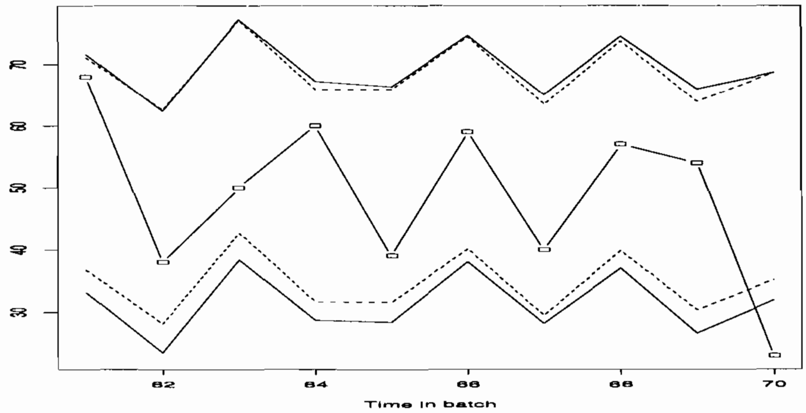

Figure 2: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for multistep ahead prediction in series F of Box and Jenkins (1976).

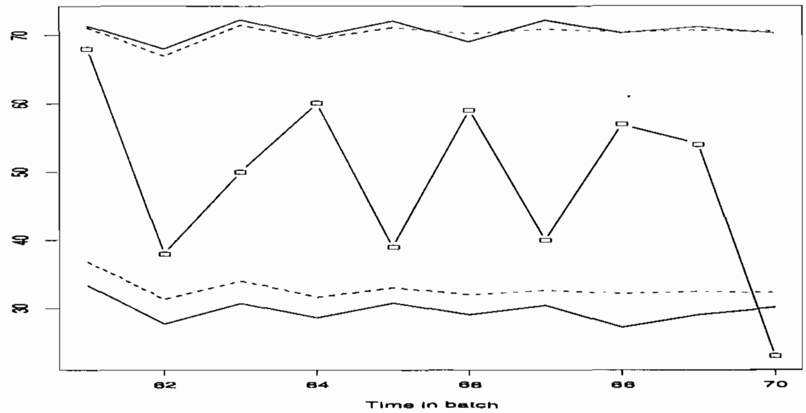

Figure 3: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for 1-step ahead prediction in series E of Box and Jenkins (1976).

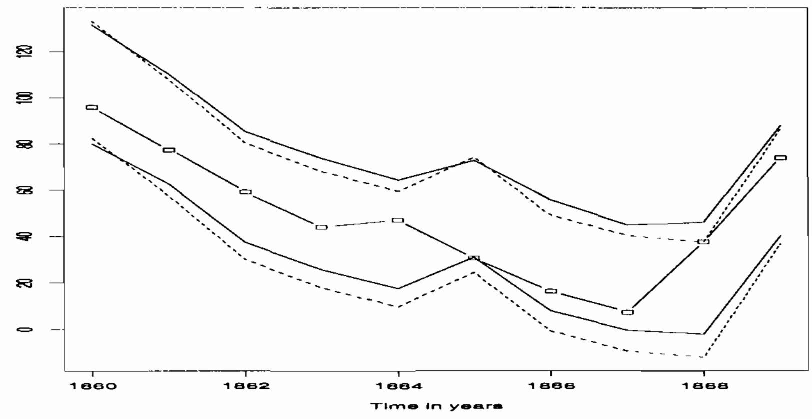

Figure 4: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for multistep ahead prediction in series E of Box and Jenkins (1976).

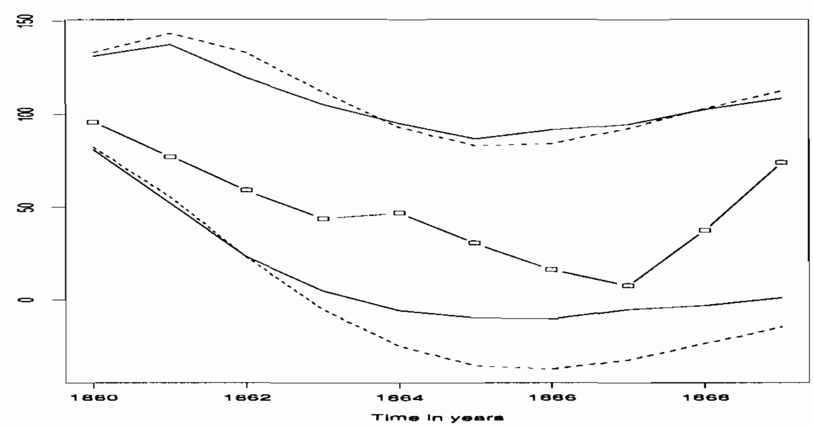

Figure 5: Observed data (-□-), Box Jenkins (−−), and VS intervals (—) for 1-step ahead prediction in sunspots series.

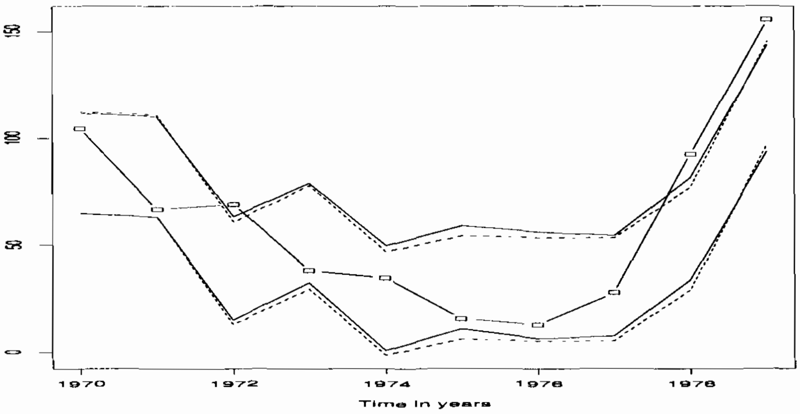

Figure 6: Observed data (-□-), Box Jenkins (−−), and VS intervals (—) for multistep ahead prediction in sunspots series.

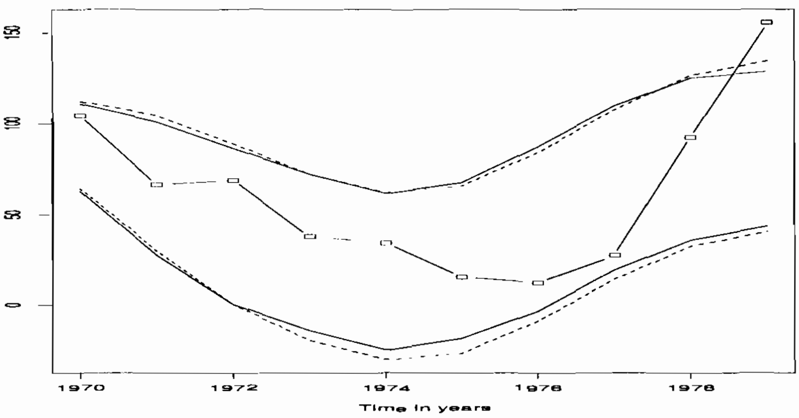

Figure 7: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for 1-step ahead prediction in Nile river series.

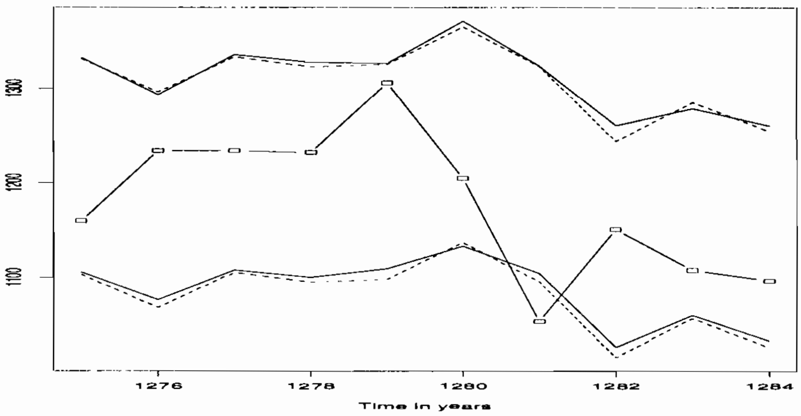

Figure 8: Observed data (-□-), Box Jenkins (– –), and VS intervals (—) for multistep ahead prediction in Nile river series.

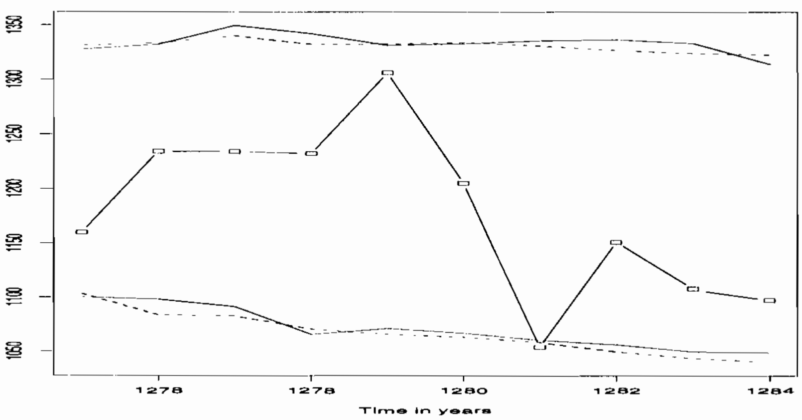

Table 1

Simulation results for Model 1, with Gaussian Errors.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	3.93	0.00
1	25	CS	87.87(0.69)	6.92/5.21	3.96(0.05)	0.08
		VS	88.91(0.61)	6.35/4.74	4.04(0.05)	0.09
		PRR	90.42(0.53)	5.18/4.40	3.89(0.05)	0.06
		PRR*	91.06(0.68)	5.09/3.85	4.64(0.06)	0.22
	50	CS	89.26(0.47)	4.69/6.05	3.74(0.04)	0.10
		VS	91.12(0.45)	3.76/5.12	3.93(0.04)	0.05
		PRR	92.23(0.36)	3.65/4.12	3.87(0.04)	0.04
		PRR*	92.15(0.64)	3.32/4.53	4.68(0.05)	0.23
	100	CS	92.02(0.34)	4.23/3.76	3.83(0.03)	0.06
		VS	92.88(0.30)	3.80/3.32	3.92(0.03)	0.02
		PRR	93.96(0.18)	3.12/2.92	3.93(0.03)	0.01
		PRR*	93.36(0.49)	3.68/2.97	4.71(0.04)	0.22
k	n	Theoretical	95%	2.50% /2.50%	4.92	0.00
3	25	CS	88.63(0.49)	6.18/5.19	4.42(0.06)	0.17
		VS	89.47(0.41)	5.67/4.86	4.47(0.06)	0.15
		PRR	92.10(0.36)	4.06/3.85	4.78(0.06)	0.06
		PRR*	92.00(0.38)	4.14/3.86	4.84(0.07)	0.05
	50	CS	89.59(0.39)	4.90/5.51	4.45(0.04)	0.15
		VS	90.77(0.33)	4.37/4.87	4.57(0.04)	0.11
		PRR	93.31(0.23)	3.31/3.38	4.84(0.04)	0.03
		PRR*	92.95(0.25)	3.47/3.58	4.82(0.04)	0.04
	100	CS	91.68(0.27)	4.19/4.13	4.61(0.03)	0.10
		VS	92.34(0.24)	3.95/3.71	4.68(0.03)	0.08
		PRR	94.15(0.17)	2.92/2.93	4.87(0.03)	0.02
		PRR*	93.50(0.20)	3.38/3.11	4.80(0.04)	0.04

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 2

Simulation results for Model 1, with Exponential Errors.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	3.67	0.00
1	25	CS	86.87(1.05)	7.35/5.78	3.96(0.10)	0.16
		VS	89.12(0.87)	5.20/5.67	4.06(0.10)	0.17
		PRR	92.33(0.67)	2.53/5.14	3.97(0.09)	0.11
		PRR*	91.37(0.78)	3.47/5.16	4.53(0.10)	0.27
	50	CS	89.08(0.87)	6.03/4.90	3.66(0.07)	0.07
		VS	93.34(0.51)	2.36/4.30	3.91(0.07)	0.09
		PRR	94.84(0.32)	0.97/4.19	3.75(0.07)	0.03
		PRR*	94.22(0.51)	2.40/3.38	4.46(0.07)	0.23
	100	CS	91.38(0.75)	4.75/3.88	3.71(0.05)	0.06
		VS	93.54(0.56)	2.81/3.65	3.88(0.05)	0.08
		PRR	95.17(0.31)	1.46/3.37	3.73(0.05)	0.03
		PRR*	93.81(0.54)	3.22/2.97	4.52(0.05)	0.25
k	n	Theoretical	95%	2.50% /2.50%	4.84	0.00
3	25	CS	88.88(0.57)	4.83/6.29	4.40(0.11)	0.15
		VS	89.95(0.51)	3.93/6.12	4.49(0.11)	0.13
		PRR	92.45(0.41)	2.15/5.40	4.76(0.11)	0.04
		PRR*	92.75(0.42)	2.12/5.13	4.93(0.12)	0.04
	50	CS	90.17(0.47)	4.40/5.43	4.43(0.07)	0.13
		VS	91.78(0.39)	3.11/5.11	4.58(0.08)	0.08
		PRR	93.89(0.29)	2.13/3.98	4.84(0.08)	0.02
		PRR*	93.27(0.30)	2.65/4.08	4.71(0.07)	0.04
	100	CS	91.78(0.43)	4.04/4.18	4.56(0.05)	0.09
		VS	92.78(0.32)	3.17/4.05	4.66(0.06)	0.06
		PRR	94.48(0.17)	2.27/3.25	4.81(0.05)	0.01
		PRR*	93.79(0.19)	2.87/3.33	4.72(0.05)	0.04

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 3

Simulation results for Model 1, with Contaminated Errors

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	12.58	0.00
1	25	CS	87.67(1.12)	6.89/5.44	12.48(0.25)	0.09
		VS	89.71(0.95)	4.93/5.36	12.73(0.26)	0.07
		PRR	90.94(0.65)	3.31/5.76	12.36(0.24)	0.06
		PRR*	92.31(0.56)	2.93/4.75	14.10(0.31)	0.15
	50	CS	89.42(0.75)	5.26/5.32	12.25(0.16)	0.08
		VS	91.93(0.52)	3.02/5.05	12.72(0.18)	0.04
		PRR	93.04(0.33)	2.20/4.76	12.45(0.15)	0.03
		PRR*	93.96(0.37)	2.17/3.86	14.02(0.18)	0.13
	100	CS	91.83(0.56)	3.33/4.84	12.55(0.08)	0.04
		VS	93.48(0.34)	1.85/4.67	12.99(0.09)	0.05
		PRR	94.13(0.22)	2.14/3.73	12.61(0.07)	0.01
		PRR*	93.73(0.47)	2.98/3.29	14.03(0.13)	0.13
k	n	Theoretical	95%	2.50% /2.50%	14.81	0.00
3	25	CS	89.20(0.76)	4.45/6.35	14.04(0.30)	0.11
		VS	90.84(0.61)	3.04/6.12	14.31(0.30)	0.08
		PRR	91.69(0.60)	2.53/5.78	14.63(0.30)	0.05
		PRR*	92.66(0.58)	2.11/5.23	15.57(0.34)	0.08
	50	CS	90.98(0.52)	4.50/4.52	14.32(0.18)	0.07
		VS	92.18(0.43)	3.25/4.56	14.58(0.19)	0.04
		PRR	93.95(0.30)	2.34/3.71	15.15(0.18)	0.04
		PRR*	93.64(0.30)	2.66/3.69	15.15(0.19)	0.04
	100	CS	92.32(0.31)	3.56/4.13	14.50(0.12)	0.05
		VS	93.63(0.27)	2.73/3.64	15.00(0.13)	0.03
		PRR	94.53(0.19)	2.39/3.08	14.97(0.10)	0.02
		PRR*	93.51(0.21)	3.40/3.09	14.77(0.13)	0.02

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 4

Simulation Results for Model 2.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	3.89	0.00
1	25	CS	89.39(0.49)	5.71/4.90	3.61(0.05)	0.13
		VS	89.64(0.47)	5.61/4.75	3.62(0.05)	0.12
	50	CS	91.50(0.36)	4.09/4.40	3.70(0.04)	0.08
		VS	92.19(0.31)	3.71/4.10	3.76(0.04)	0.06
	100	CS	93.24(0.23)	3.35/3.40	3.82(0.03)	0.04
		VS	93.50(0.21)	3.17/3.33	3.84(0.03)	0.03
k	n	Theoretical	95%	2.50% /2.50%	3.91	0.00
3	25	CS	89.99(0.46)	5.37/4.65	3.64(0.05)	0.12
		VS	90.09(0.44)	5.35/4.56	3.65(0.05)	0.12
	50	CS	92.07(0.29)	4.01/3.93	3.76(0.04)	0.07
		VS	92.23(0.29)	3.79/3.97	3.79(0.04)	0.06
	100	CS	93.57(0.22)	3.24/3.19	3.87(0.03)	0.03
		VS	93.69(0.20)	3.11/3.20	3.88(0.03)	0.02

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 5

Simulation results for Model 3.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	2.73	0.00
1	25	CS	87.93(0.51)	6.55/5.51	2.45(0.03)	0.18
		VS	88.29(0.49)	6.46/5.25	2.48(0.03)	0.16
	50	CS	90.29(0.42)	4.61/5.09	2.57(0.03)	0.10
		VS	91.27(0.38)	4.09/4.64	2.65(0.03)	0.06
	100	CS	92.56(0.23)	3.78/3.66	2.65(0.02)	0.05
		VS	93.01(0.21)	3.54/3.45	2.68(0.02)	0.04
k	n	Theoretical	95%	2.50% /2.50%	3.05	0.00
3	25	CS	86.64(0.52)	7.13/6.23	2.57(0.03)	0.24
		VS	86.20(0.50)	7.47/6.34	2.54(0.03)	0.26
	50	CS	90.21(0.37)	4.85/4.95	2.82(0.03)	0.13
		VS	90.03(0.36)	4.81/5.15	2.81(0.03)	0.13
	100	CS	92.55(0.25)	3.78/3.67	2.96(0.02)	0.05
		VS	92.58(0.24)	3.74/3.68	2.96(0.02)	0.05

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 6

Simulation results for Model 4.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	3.91	0.00
1	25	CS	78.07(0.97)	4.08/ 17.85	4.27(0.05)	0.27
		VS	64.53(1.04)	6.62/ 28.86	5.77(0.07)	0.79
	50	CS	91.81(0.35)	5.21/2.97	3.87(0.04)	0.04
		VS	90.79(0.38)	7.22/1.99	4.59(0.04)	0.22
	100	CS	94.05(0.23)	3.78/2.18	4.01(0.03)	0.03
		VS	93.70(0.24)	4.94/1.36	4.20(0.03)	0.09
k	n	Theoretical	95%	2.50% /2.50%	6.19	0.00
3	25	CS	79.19(0.80)	3.63/ 17.17	6.90(0.07)	0.28
		VS	59.84(0.90)	7.45/ 32.71	9.65(0.11)	0.93
	50	CS	93.57(0.31)	4.68/1.75	6.71(0.05)	0.11
		VS	87.55(0.44)	9.95/2.49	8.32(0.07)	0.43
	100	CS	95.19(0.19)	3.27/1.55	6.77(0.04)	0.10
		VS	93.32(0.27)	5.59/1.09	7.40(0.05)	0.22

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).

Table 7

Simulation results for Model 5.

Lag	Sample size	Method	$\bar{C}_M$ (se)	Cov. (below/above)	$\bar{L}_M$ (se)	CQ_M
k	n	Theoretical	95%	2.50% /2.50%	1.90	0.00
1	25	CS	56.09(0.67)	0.00/ 43.91	2.54(0.02)	0.75
		VS	40.60(0.69)	0.00/ 59.40	4.66(0.04)	2.03
	50	CS	93.22(0.41)	6.72/0.07	2.18(0.01)	0.17
		VS	83.65(0.45)	16.35/0.00	3.14(0.02)	0.77
	100	CS	94.21(0.38)	5.61/0.18	2.05(0.01)	0.09
		VS	88.51(0.42)	11.49/0.00	2.42(0.01)	0.34
k	n	Theoretical	95%	2.50% /2.50%	3.50	0.00
3	25	CS	62.66(1.01)	0.00/ 37.34	4.56(0.04)	0.64
		VS	33.14(0.85)	0.00/ 66.86	8.14(0.07)	1.97
	50	CS	91.93(0.36)	8.02/0.05	4.00(0.02)	0.17
		VS	78.94(0.60)	21.06/0.00	6.32(0.05)	0.98
	100	CS	94.82(0.25)	4.75/0.43	3.94(0.01)	0.13
		VS	88.98(0.47)	11.02/0.00	4.92(0.03)	0.47

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$, CQ_M and se's are computed from (12).