Working Paper 01-14
Statistics and Econometrics Series 9
February 2001

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

INTRODUCING MODEL UNCERTAINTY IN TIME SERIES BOOTSTRAP

Andrés M. Alonso, Daniel Peña and Juan Romo*

Abstract

It is common in parametric bootstrap to select the model from the data, and then treat it as it were the true model. Kilian (1998) have shown that ignoring the model uncertainty may seriously undermine the coverage accuracy of bootstrap confidence intervals for impulse response estimates which are closely related with multi-step-ahead prediction intervals. In this paper, we propose different ways of introducing the model selection step in the resampling algorithm. We present a Monte Carlo study comparing the finite sample properties of the proposed method with those of alternative methods in the case of prediction intervals.

Keywords: sieve bootstrap, prediction, time series, model uncertainty.

* Dept. of Statistics and Econometrics, Univ. Carlos III de Madrid, Spain. Alonso, e-mail: alonsoam@est-econ.uc3m.es; Peña: dpena@est-econ.uc3m.es; Romo: romo@est-econ.uc3m.es;
This work was partially supported by the DGES project PB96-0111 and Cátedra de Calidad BBVA. AMS 1991 Subject Classifications. Primary: 62G09; Secondary: 62M10.

1 Introduction

An important question in empirical time series analysis is how to predict the future values of an observed time series on the basis of its recorded past, and more specifically how to calculate prediction intervals. A traditional approach to these questions assumes that the series $\{X_t\}_{t\in\mathbb{Z}}$ follows a linear finite dimension model with a known errors distribution, e.g. a Gaussian autoregressive-moving average ARMA(p,q) model as in Box and Jenkins (1976). In such a case, if the orders p and q are known, a maximum likelihood procedure could be employed for estimating the parameters and then, plug in those estimates in the linear predictors. In addition, some bootstrap approaches have been proposed in order to avoid the use of a specified errors distribution, see v.g. Stine (1987) and Thombs and Schucany (1990) for AR(p) models, and Pascual *et al.* (1998) for ARMA(p,q) models. But, those bootstrap proposals also assume that p and q are known. Alonso *et al.* (2001) show that the AR(∞)-sieve bootstrap provides consistent prediction intervals for a general class of linear models that includes stationary and invertible ARMA processes. This procedure select an approximating autoregressive model AR$(\widehat{p})$ from the data, and then use the selected order as if it were the true order. However, this approach ignores the variability involved in model selection, which can be a considerable part of the overall uncertainty.

In practice, having observed a sample of size n, the model, and particularly p and q are invariably unknown. Thus, we should select a model from the data. Many models selection procedures have been proposed, v.g. the final prediction error (FPE) of Akaike (1969), the Akaike (1973) information criterion (AIC) or its bias-corrected version (AICC) of Hurvich and Tsai (1989) and the Bayesian information criterion of Schwarz (1978), see Bhansali (1993) for a review.

For finite autoregressive models, Massaroto (1990) and Grigoletto (1998), propose to take into account model uncertainty as follows: first, to obtain $\widehat{p}$ by a consistent model selection procedure, then generate bootstrap resamples from the estimated AR$(\widehat{p})$ and to re-estimate in each resample the order by the same method used for $\widehat{p}$. Thus, their prediction intervals consider the sampling variability caused by model selection method. Essentially the same algorithm was suggested by Kilian (1998) in the context of generating impulse response confidence intervals, the so called *endogenous lag order bootstrap*. It is well known that consistent model selection procedures (as the BIC) tend to select more parsimonious orders. In fact, Grigoletto (1998) and Kilian (1998) recommend to use the less parsimonious AIC procedure. Hjorth (1994) suggests the following: first, estimating an AR(p_{max}) from the data, where p_{max} is the greatest order considered, and then proceeding as in the previous approach. Although this last proposal avoids the dependence on $\widehat{p}$, it could be influenced by the high variability of the p_{max} estimated parameters. In Section 3 Monte Carlo simulations reveal that version with $\widehat{p}$ is generally preferable to version with p_{max}.

Since the previous endogenous order bootstrap could be affected by the initial estimated order, we propose in this paper two ways of introducing the sampling variability of the model selection procedure that does not depend, or are less dependent, on $\widehat{p}$. These approaches are: (i) to use an estimator of the distribution of $\widehat{p}$ based on moving block resampling, and (ii) to construct a probability function for p based on the values of the objective function of the above mentioned information criterions (AIC, AICC, or BIC). Once we have an estimated

distribution $\widehat{F}_p$, we generate resamples from the estimated AR(p^*) with the p^* i.i.d. $\widehat{F}_p$, and then we proceed as in standard bootstrap approaches. In the performed Monte Carlo study, the results point out that (*ii*) outperforms the other proposal and the endogenous lag order bootstrap of Kilian (1998).

The remaining of this paper is organized as follows. Section 2 presents the endogenous lag order method of Kilian (1998) applied to the sieve bootstrap prediction intervals and introduces our proposals (*i*) and (*ii*). In Section 3 we present a theoretical justification of the proposed methods and in Section 4 we present the results of a Monte Carlo study comparing the finite sample properties of the proposed methods with those of alternative methods. All proofs are given in an Appendix.

2 Proposed approaches

2.1 The sieve endogenous order bootstrap

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real valued, stationary process with expectation $\mathrm{E}[X_t] = \mu_X$ that admits a one-sided infinite-order autoregressive representation:

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

with coefficients $\{\phi_j\}_{j=0}^{+\infty}$ satisfying $\sum_{j=0}^{+\infty} \phi_j^2 < \infty$. This representation motivates the AR(∞)-sieve bootstrap, that was first proposed by Kreiss (1988) and extensions can be found in Bühlmann (1997). The method proceeds as follows:

1 Given a sample $\{X_1, \ldots, X_n\}$, select the order $\widehat{p}$ of the autoregressive approximation by AICC criterion.

The AICC $= -n\log(\sigma^2) + 2(p+1)n/(n-p-2)$ is a bias-corrected version of AIC which has a more extreme penalty for large-order models to counteract 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. Other advantage of using AICC is that the value of the maximum cut-off p_{max} had virtually no effect on the model chosen by this criterion, while for many of the other criteria increasing the value of p_{max} tends to lead to increased overfitting of the model, see Hurvich and Tsai (1989).

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

3 Compute the residuals:

$$\widehat{\varepsilon}_t = \sum_{j=0}^{\widehat{p}} \widehat{\phi}_j (X_{t-j} - \bar{X}), \qquad \widehat{\phi}_0 = 1,\ t \in (\widehat{p}+1, \ldots, n). \tag{2}$$

4 Define the empirical distribution function of the centred residuals:

$$\widehat{F}_{\tilde{\varepsilon}}(x) = (n - \widehat{p})^{-1} \sum_{t=\widehat{p}+1}^{n} 1_{\{\tilde{\varepsilon}_t \leq x\}}, \tag{3}$$

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

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

6 Define X_t^* by the recursion:

$$\sum_{j=0}^{\widehat{p}} \widehat{\phi}_j (X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{4}$$

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

Up to this step, the resampling plan coincides with the sieve bootstrap, and is valid for bootstrapping some statistics defined as a functional of an m-dimensional distribution function (see details in Section 3.3 of Bühlmann (1997)). In the next step we introduce the endogenous lag order selection:

7 Given the bootstrap replication $\{X_1^*, X_2^*, \ldots, X_n^*\}$, select the order $\widehat{p}^*$ of the autoregressive approximation as in the step 1.

The steps 1 to 7 are not effective for bootstrap prediction, because the algorithm 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 $X_{T-p+1}^* = X_{T-p+1}, \ldots, X_T^* = X_T$.

8 Compute the estimation of the autoregressive coefficients: $(\widehat{\phi}_1^*, \widehat{\phi}_2^*, \ldots, \widehat{\phi}_{\widehat{p}^*}^*)$, as in step 2.

9 Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{\widehat{p}^*} \widehat{\phi}_j^* (X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \tag{5}$$

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 9 B times. The $(1-\alpha)\%$ prediction interval for X_{T+h} is given by

$$\left[Q^*(\alpha/2), Q^*(1-\alpha/2)\right], \tag{6}$$

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

The consistency of the intervals in (6) follows from Lemma 3.2 and Proposition 3.10, Proposition 3.5 and Theorem 3.12.

Notice that, if we omit step 7 and use the $\widehat{p}^* = \widehat{p}$ in step 8 and recursion (5), the resampling plan coincides with the sieve bootstrap prediction algorithm of Alonso *et al.* (2001). Both approaches will be compared in the Monte Carlo study of Section 4.

2.2 The sieve exogenous order bootstrap

In this subsection we present a different way of introducing the sampling variability of the model selection procedure. First, we describe the general algorithm and then we present two possible implementations, based on (i) blockwise resampling with missing values techniques, and (ii) information criteria functions. Let $\{X_t\}_{t\in\mathbb{Z}}$ as in the previous subsection, and let $IC(p)$ be the objective function of some model selection method. Assume that we have a probability distribution estimator $\widehat{F}_p$ of the random variable $\widehat{p} = \text{argmin}_{0\leq p\leq p_{max}} \{IC(p)\}$, i.e., we have estimates of the following probabilities:

$$\Pr\{\widehat{p} = p\}\,, \text{ for } 0 \leq p \leq p_{max}. \tag{7}$$

The characterization of the asymptotic limit of (7) is a standard way of proving the consistency of information criteria in finite autoregressive models, see v.g. Shibata (1976). Two methods for obtaining an approximation of (7) will be presented below and it constitute the first step of the sieve exogenous order bootstrap.

The sieve exogenous order bootstrap modify the previous step 2 as follows:

2a Construct estimators of the coefficients for the $p_{max}+1$ autoregressive models: $(\widehat{\phi}_1^{(p)}, \widehat{\phi}_2^{(p)}, \ldots, \widehat{\phi}_p^{(p)})$ for $1 \leq p \leq p_{max}$, and $\widehat{\phi}_0^{(p)} = 1$ for $0 \leq p \leq p_{max}$.

3a Compute the residuals for the model with $p = \widehat{p}$:

$$\widehat{\varepsilon}_t = \sum_{j=0}^{\widehat{p}} \widehat{\phi}_j^{(\widehat{p})}(X_{t-j} - \bar{X}), \qquad t \in (\widehat{p}+1, \ldots, n). \tag{8}$$

4a Define the empirical distribution functions of the centred residuals:

$$\widehat{F}_{\tilde{\varepsilon}}(x) = (n-\widehat{p})^{-1} \sum_{t=\widehat{p}+1}^{n} 1_{\{\tilde{\varepsilon}_t \leq x\}}, \tag{9}$$

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

The following step (numerated as 4b, in order to maintain the same number of steps as the endogenous approach) is the distinctive part of the proposed algorithm:

4b Draw a random value p^* from $\widehat{F}_p$.

Notice that given the probabilities in (7), the variability of $\widehat{p}$ enters exogenously to the bootstrap algorithm.

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

6a Define X_t^* by the recursion:

$$\sum_{j=0}^{p^*} \widehat{\phi}_j^{(p^*)}(X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{10}$$

where the starting p^* observations are equal to $\bar{X}$.

At this point we might perform the step 7 of the endogenous order bootstrap, but simulation studies (not shown here) point out that there are small differences when we introduce this step in the exogenous approach. By contrary, the computational cost increases when we perform step 7 in each resample.

8a Compute the estimation of the autoregressive coefficients: $(\widehat{\phi}_1^*, \ldots, \widehat{\phi}_{p^*}^*)$, as in step 2.

9a Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{p^*} \widehat{\phi}_j^*(X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \tag{11}$$

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

As before, the bootstrap distribution of X_{T+h}^* is used to approximate the unknown distribution of X_{T+h} given the observed sample, and steps 4b - 9a are repeated B times in order to obtain $\widehat{F}_{X_{T+h}^*}^*$.

Next, we develop two ways of obtaining an estimator or an approximation of the probabilities (7). This constitutes the first step of the sieve exogenous order bootstrap.

(i) **Moving block resampling order distribution.** The moving block jackknife and bootstrap were introduced by Künsch (1989) and independently by Liu and Singh (1992). In the following, we use the presentation of Liu and Singh (1992): Lets $X_1, X_2, \ldots, X_n$ be random variables with the common distribution function F_X, and let T be the parameter of interest and T_n its estimator based on $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Let B_i denote a block of ℓ consecutive observations, i.e. $B_i = (X_i, X_{i+1}, \ldots, X_{i+\ell+1})$ for $i = 1, 2, \ldots, n - \ell + 1$. Then,

- For the moving block jackknife (MBJ), we denote the i-th jackknife statistics by $T_{n,-i}$ which is equal to the estimator $T_{n-\ell}$ evaluated in the reduced sample $\mathbf{X} - B_i$. Then, the following *jackknife histogram* is a distribution estimator of $\tau_n(T_n - T)$:

$$H_N(x) = (n - \ell + 1)^{-1} \sum_{i=1}^{n-\ell+1} 1\left\{\tau_\ell \ell^{-1}(n-\ell)(T_{n,-i} - T_n) \leq x\right\}, \tag{12}$$

where the τ_n is an appropriate normalizing constant.

- For the moving block bootstrap (MBB), we resample k blocks from $\{B_1, B_2, \ldots, B_{n-\ell+1}\}$ with replacement and with equal probability for each block to be drawn. For simplicity, we assume $n = k\ell$, then a bootstrap resample is obtained by joining the blocks together in random order. The bootstrap statistic its defined by T_n^* which is equal to the estimator T_n evaluated in the bootstrap resample. Then, the bootstrap distribution:

$$\Pr^* \{\tau_n(T_n^* - T_n) \leq x\} \tag{13}$$

  is a distribution estimator of $\tau_n(T_n - T)$.

Originally, the MBJ was proposed as a variance estimator, but using similar arguments as for subsampling method of Politis and Romano (1994) it is possible to establish the consistency of (12), see Alonso *et al.* (2000) for linear statistics. Also for linearized statistics the consistency of (13) is proved by Politis and Romano (1992) and Bühlmann and Künsch (1995).

Some works point out that in blockwise bootstrap a smooth transition between blocks could be preferable to random joining, see Carltein *et al.* (1998). Also, Künsch (1989) shows that in MBJ it is better to downweight the block B_i instead of a completed deletion. Those facts motivate the moving block resampling by missing values techniques proposed in Alonso *et al.* (2000).

- In the moving missing block jackknife (M^2BJ), we consider the deleted block B_i as ℓ consecutive missing values and we estimated those observations taking into account the dependence structure of $\{X_t\}_{t\in\mathbb{Z}}$. Then the i-th M^2BJ statistics $\widetilde{T}_{n,-i}$ is equal to the estimator T_n evaluated in $(\mathbf{X} - B_i) \cup \widehat{B}_i$, where $\widehat{B}_i$ denotes the estimate of B_i. Then, the following M^2BJ histogram is a distribution estimator of $\tau_n(T_n - T)$:

$$\widetilde{H}_N(x) = (n - \ell + 1)^{-1} \sum_{i=1}^{n-\ell+1} 1\left\{\tau_\ell \ell^{-1}(n-\ell)(\widetilde{T}_{n,-i} - T_n) \leq x\right\}. \tag{14}$$

Notice that M^2BJ statistics have the computational advantage of using the same functional form as T_n, while for MBJ statistics we should implement the calculations considering that ℓ observations are missing.

- In the moving missing block bootstrap (M^2BB), as in MBB, we have in each resample k blocks $(B_{i_1}^*, B_{i_2}^*, \ldots, B_{i_k}^*)$. Let $\ell = b + m$, and we consider the m last observations in each block as missing values. Thus, we will have km missing observations that will be replaced by their estimates. The M^2BB statistic is defined by $\widetilde{T}_n^*$ which is equal to the estimator T_n evaluated in the resulting resample. Then, the bootstrap distribution

$$\Pr^* \left\{\tau_n(\widetilde{T}_n^* - T_n) \leq x\right\} \tag{15}$$

  is a distribution estimator of $\tau_n(T_n - T)$.

The M^{2}BB scheme resemble to a block joining engine similar to the matched-block bootstrap of Carltein *et al.* (1998).

In our case $T_n = \widehat{p}$, and the computational implementation of M^{2}BJ and M^{2}BB only requires additionally a missing values estimation method. We will use the generalized least square estimators proposed by Peña and Maravall (1991). The consistency of (12)-(15) estimators for this particular statistics is beyond of the scope of this paper.

A problem related with blockwise resampling is the selection of block size ℓ (in ours case, ℓ and k). A general approach to solve this problem was proposed by Hall *et al.* (1995) for MBB distribution and variance estimators. This approach could be easily modified in order to select ℓ and k for M^{2}BJ and M^{2}BB, but it involves a discrete optimization in ℓ and k that could be a computational disadvantage. Since in each bootstrap resample we have to estimate autoregressive models up to order p_{max}, we use in the Monte Carlo study $\ell = 2p_{max}, 3p_{max}$ and $4p_{max}$ (with $p_{max} = n/10$ as recommended Bhansali (1983)) and $k = 1$ in order to have a moderate number of missing values.

(*ii*) **Information criterion function order distribution.** This approach is related to the Bayesian formulation of the prediction problem, see v.g. Kass and Raftery (1995). Assume that $p_{max} + 1$ autoregressive models are being considered, then the posterior probability of model AR(p) is given by

$$\Pr\{\mathrm{AR}(p)|\mathbf{X}\} = \frac{\alpha_p B_{p0}}{\sum_{i=0}^{p_{max}} \alpha_i B_{i0}}, \tag{16}$$

where the B_{p0} are the Bayes factors for AR(p) against AR(0), and $\alpha_p = \Pr\{\mathrm{AR}(p)\}/\Pr\{\mathrm{AR}(0)\}$ is the prior odds for model AR(p) against model AR(0). AR(0) is equivalent to assume that $\{X_t\}_{t\in\mathbb{Z}}$ is a white noise process. Once we have the probabilities (16), we calculate the posterior distribution function of X_{T+h} that takes into account the model uncertainty, by the relationship:

$$F_{X_{T+h}}(x) = \sum_{i=0}^{p_{max}} F^{(p)}_{X_{T+h}}(x) \Pr\{\mathrm{AR}(p)|\mathbf{X}\}, \tag{17}$$

where $F^{(p)}_{X_{T+h}}(x)$ is the distribution function of X_{T+h} calculated assuming that $\{X_t\}_{t\in\mathbb{Z}}$ is an AR(p) process.

Notice that using probabilities (16) in step 4b conduct us to obtain an approximation of the distribution (17).

Since calculating the p_{max} Bayes factors is computationally arduous, here we use the following approximation: $B_{p0} \approx \exp(S_{p0})$ (as recommend Kass and Raftery (1995)), S_{p0} is the Schwarz criterion and its given by

$$S_{p0} = L_p - L_0 - \frac{1}{2} p \log n, \tag{18}$$

where L_0 and L_p are the log-likelihood of model AR(0) and AR(p) evaluated at $\widehat{\phi}_0$ and ($\widehat{\phi}_0$, $\widehat{\phi}_1, \ldots, \widehat{\phi}_p$), respectively. Notice that S_{p0} could be obtained by simple linear transformation of the consistent BIC(p) objective function. Then probabilities (16) could be approximated by:

$$\Pr\{\mathrm{AR}(p)|\mathbf{X}\} = \frac{\alpha_p \exp(-1/2\mathrm{BIC}(p))}{\sum_{i=0}^{p_{max}} \alpha_i \exp(-1/2\mathrm{BIC}(p))}. \tag{19}$$

A word of caution about the approximation of Bayes factor by $\exp(S)$. Its relative error is $O(1)$, i.e. frequently $\exp(S_{p0})/B_{p0} \nrightarrow 1$.

Buckland *et al.* (1997) propose a similar approach in the context of Poisson regression, line transient sampling and survival model, but they do not provide a theoretical justification of the proposed bootstrap methods. Also, Buckland *et al.* (1997) recommend to use AIC(p) instead of BIC(p) in approximation (19).

In the Monte Carlo study of the next section, the sieve exogenous order bootstrap based on (16) with the above approximation perform reasonably well. We present the results with AICC model selection procedure, but additional simulations studies (available on request to the authors) illustrate that the sieve exogenous order bootstrap based on (16) with the BIC procedure perform similarly. The consistency of the intervals (6) using probabilities (19) follows from Lemma 3.2 and Proposition 3.3, Proposition 3.5 and Theorem 3.12.

A related approach was proposed by LeBlanc and Tibshirani (1996) in the cases of regression and classification for combining predictors, but they use as weights (or probabilities) the following expression: $L_k / \sum_{i=1}^{K} L_i$ where L_k is the likelihood for model k, and the K considered models have the same dimension.

3 Consistency results

The asymptotic validity of the proposed intervals 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 procedures have three main steps: (i) obtaining or selecting the p^*, (ii) obtaining the estimates $\widehat{\phi}^*_{p^*}$ in order to have information about the distribution of $\widehat{\phi}_p$, and (iii) computing the future values X^*_{T+h}.

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

> *Assumption A1:* $X_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ $(t \in \mathbb{Z})$ with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ stationary, ergodic and $\mathrm{E}[\varepsilon_t|\mathcal{F}_{t-1}] \equiv 0$, $\mathrm{E}[\varepsilon_t\varepsilon_s|\mathcal{F}_{\min(t,s)-1}] \equiv \delta_{t,s}\sigma^2 < \infty$, $\mathrm{E}[|\varepsilon_t|^s] < \infty$ 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 $r \in \mathbb{N}$.

Notice that Assumptions *A1* and *A2* are satisfied by stationary and invertible ARMA(p,q) processes which have an exponential decay of the coefficients ${\psi_j}_{j=0}^{+\infty}$ (cf. Bühlmann (1997)). We present the theoretical results for two types of linear models: Assumption *A3* $\{X_t\}_{t\in\mathbb{Z}}$ not generate to a finite order AR process, and its complement Assumption *A3*^{c} $\{X_t\}_{t\in\mathbb{Z}}$ is an AR(p_0) process for some finite p_0. Additionally, we impose the following assumption about the autoregressive approximation:

> *Assumption B:* $0 \leq p = p(n) \leq p_{max}(n)$, where $p_{max}(n) \to \infty$, $p_{max}(n) = o(n^{1/2})$ as $n \to \infty$, and $\widehat{\phi}_p = (\widehat{\phi}_{1,n}, \ldots \widehat{\phi}_{p,n})'$ satisfy the empirical Yule Walker equations

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

We present the results for the order selection method $S_n(p) = (n+2p)\widehat{\sigma}^2_{p,n}$, where $\widehat{\sigma}^2_{p,n} = n^{-1}\sum_{t=p_{max}}^{n-1}(X_{t+1} + \widehat{\phi}_{1,n}X_t + \cdots + \widehat{\phi}_{p,n}X_{t+1-p})^2$, proposed by Shibata (1980). This order selection is a version of the final prediction error (FPE), and has a close relation to other asymptotically efficient method like AIC and AICC.

The following two lemmas characterize the asymptotic behavior of the selected order sequence $\{\widehat{p}(n)\}$. Lemma 3.1 is a consequence of Theorem 3.1, Theorem 4.1 and Remark 5.2 of Pötscher (1990), and Lemma 3.2 is the analogous of Corollary 4.1 of Shibata (1980) based on Theorem 3.1 of Karagrigoriou (1997).

Lemma 3.1 *Suppose that assumption A1 with $s = 4$, A2 with $r = 1$, and B with $p_{max} = o(n^{1/2})$ hold. Then, the random sequence $\widehat{p} = \widehat{p}(n)$ that minimize $S_n(p)$, satisfies that*

1. *Under assumption A3^c, $\Pr\{\widehat{p}(n) \geq p_0\} \to 1$, as $n \to \infty$.*
2. *Under assumption A3^c, $\{\widehat{p}(n)\}$ is a divergent sequence.*

Now we introduce some notation used in Lemma 3.2 (cf. Shibata (1980)). Let's denote $L_n(p) = p\sigma^2/n + \|\phi_p - \phi\|_\Gamma$, where $\phi_p = (\phi_{1,n},\ldots,\phi_{p,n})'$ are the theoretical Yule-Walker statistics, and the norm $\|x\|_A = (x'Ax)^{1/2}$. Also, we denote $\{p_0(n)\}$ the non-random sequence that minimizes $L_n(p)$, and $\{p_\varepsilon(n)\}$ is the non-random sequence defined by $\{p_\varepsilon(n)\} = \min\{p : L_n(p)/L_n(p_0(n)) \leq 1+\varepsilon\}$.

Lemma 3.2 *Suppose that assumption A1 with $s = 16$, A2 with $r = 1$, A3, and B with $p_{max} = o(n^{1/2})$ hold. Then, for any $\varepsilon > 0$ the random sequence $\widehat{p} = \widehat{p}(n)$ that minimize $S_n(p)$ satisfy that*

$$\Pr\{\widehat{p}(n) \geq p_\varepsilon(n)\} \to 1, \text{ as } n \to \infty.$$

Now, we use $S_n(p)$ in probabilities (19) instead of AICC(p) or BIC(p). In the following proposition we establish that if $\{X_t\}_{t\in\mathbb{Z}}$ is *A3* and we select a random order $p = p(n)$ having probability function given by (19), then the probability of selecting p in any finite set is zero in comparison with the probability of $p = \widehat{p}$.

Proposition 3.3 *Suppose that assumption A1 with $s = 16$, A2 with $r = 1$, A3, and B with $p_{max} = o(n^{1/2})$ hold. Then, for any $0 < C < +\infty$*

$$\frac{\sum_{c=1}^{C}\Pr_{S_n}\{p = c\}}{\Pr_{S_n}\{p = \widehat{p}\}} \to 0, \quad \textit{in probability}, \tag{20}$$

where $\Pr_{S_n}$ denote the probabilities (19) calculated with S_n.

Analogously, if $\{X_t\}_{t\in\mathbb{Z}}$ verifies Assumption *A3*^{c} we can establish that for any $1 \leq c < p_0$, the probability of selecting p in $\{1,\ldots,c\}$ is zero in comparison with the probability of $p = \widehat{p}$. This is a direct consequence of Lemma 3.1(1) and that $\sigma_c^2 > \sigma^2$ holds for $c < p_0$.

Remark 3.4 The statement of Proposition 3.3 holds for any divergent sequence $\{p(n)\}$ such that $p(n) = o(n^{1/2})$. Also notice that, if $\{X_t\}_{t\in\mathbb{Z}}$ verifies *A3*, (20) implies that for any $0 < C < +\infty$, $lim_{n\to\infty} \Pr_{S_n}\{p \leq C\} = 0$, and similarly, if $\{X_t\}_{t\in\mathbb{Z}}$ verifies *A3*^{c}, for any $0 < c < p_0$, $lim_{n\to\infty} \Pr_{S_n}\{p \leq c\} = 0$.

The following proposition is related with step 8 (8a); it establishes that the analogous bootstrap $\widehat{\phi}^*_{j,n}$ are consistent estimators of the theoretical Yule-Walker statistics $\phi_{j,n}$, defined by the $\Gamma_p\phi_p = -\gamma_p$, assuming that the bootstrap resamples are constructed following an estimated AR($p(n)$) process, where $p(n)$ is a divergent sequence.

Proposition 3.5 *Suppose that assumptions A1 with $s = 4$, A2 with $r > 1$ and B with $p_{max} = o((n/\log(n))^{1/(2r+2)})$ hold. Then*

$$\max_{1\leq j\leq p(n)} |\widehat{\phi}^*_{j,n} - \phi_{j,n}| \xrightarrow{P^*} 0, \text{ in probability.} \tag{21}$$

Remark 3.6 Notice that Lemma 3.2 (Proposition 3.3) and Proposition 3.5 imply that endogenous (exogenous based on probabilities (19)) sieve bootstrap provides consistent estimators of the theoretical Yule-Walker statistics when $\{X_t\}_{t\in\mathbb{Z}}$ verifies *A3*.

Remark 3.7 The statement of Proposition 3.5 holds, when $\{X_t\}_{t\in\mathbb{Z}}$ sarisfies Assumption *A3*^{c}, since the proof of Theorem 3.1 and 3.2 of Bühlmann (1995) can be modified in order to avoid the assumption about $p \to \infty$. Notice that under A3^c we have that $\max_{1\leq j\leq p_{max}} |\widehat{\phi}_{j,n} - \phi_{j,n}| = O_{a.s.}((\log(n)/n)^{1/2})$ and $\phi_{j,n} = \phi_j = 0$ for $j > p_0$. Then, the term $O_{a.s.}(p^{-r})$ does not appears in expression (33) if p is finite.

In the endogenous sieve bootstrap, we re-select an order $\widehat{p}^*$ in each resample that minimizes $S^*_n(p) = (n + 2p)\widehat{\sigma}^*_{p,n}$. Proposition 3.10 establishes that $\widehat{p}^*$ is also a divergent sequence. We will use the following lemma in the proof of Proposition 3.10.

Lemma 3.8 *Suppose that assumption A1 with $s = 16$, A2 with $r > 2$, and B with $p_{max} = o((n/\log(n))^{1/(2r+2)})$ hold. Then,*

$$\max_{1\leq p\leq p(n)} \left|\widehat{\sigma}^{2\,*}_{p,n} - \sigma^2_{p,n}\right| \xrightarrow{P^*} 0, \text{ in probability,} \tag{22}$$

where $\widehat{\sigma}^{2\,*}_{p,n} = n^{-1}\sum_{t=p_{max}}^{n-1}\left(X^*_{t+1} + \widehat{\phi}^*_{1,n}X^*_t + \cdots + \widehat{\phi}^*_{p,n}X^*_{t+1-p}\right)^2$ *and* $\sigma^2_{p,n} = \mathrm{E}\Big[n^{-1}\sum_{t=p_{max}}^{n-1}\left(X_{t+1} + \phi_{1,n}X_t + \cdots + \phi_{p,n}X_{t+1-p}\right)^2\Big]$.

Remark 3.9 The statement of Lemma 3.8 holds, when $\{X_t\}_{t\in\mathbb{Z}}$ verifies Assumption *A3*^{c}, arguments similar to those in Remark 3.7.

Proposition 3.10 *Suppose that assumption A1 with $s = 16$, A2 with $r > 2$, A3, and B with $p_{max} = o((n/\log(n))^{1/(2r+2)})$ hold. Then, the random sequence $\widehat{p}^* = \widehat{p}^*(n)$ that minimizes $S_n^*(p) = (n+2p)\widehat{\sigma}_{p,n}^{2*}$, is a divergent sequence.*

Remark 3.11 Under $A3^c$, by Lemma 3.1(1) and Remark 3.9 we can establish an analogous result to Proposition 3.10, i.e., $\widehat{p}^* = \widehat{p}^*(n)$ that minimizes $S_n^*(p)$ satisfies that $\Pr\{\widehat{p}(n) \geq p_0\} \to 1$, as $n \to \infty$, since now we have that $\sigma_c^2 > \sigma^2$ for $0 < c < p_0$.

The following theorem is related with step 9 (9a), it establishes that bootstrap distribution $F^*_{X^*_{T+h}}\big(x|X_T^* = X_T, \ldots, X^*_{T-p^*(n)} = X_{T-p^*(n)}\big)$ converges to the distribution function $F_{X_{T+h}}\big(x|X_{-\infty}^T\big)$, where $X_{-\infty}^T$ denotes the time series sample path up to time T. We assume that the bootstrap predictions X^*_{T+h} are constructed using the parameters $\widehat{\phi}^*$ estimated in step 8 (8a), and order $p^* = p^*(n)$ which is a divergent sequence. In Theorem 3.12, we use the notation of Thombs and Schucany (1990) which left the last observation X_T fixed and a sample of size n is written as $(X_{T-n+1}, \ldots, X_T)$.

Theorem 3.12 *Suppose that assumptions A1 with $s = 16$, A2 with $r > 2$ and B with $p = o((n/\log n)^{1/(2r+2)})$ hold. Then*

$$X^*_{T+h} \xrightarrow{d} X_{T+h}, \text{ in probability.} \tag{23}$$

Remark 3.13 Under $A3^c$, by Lemma 3.1(1), and Remark 3.7 we can establish an analogous result to Theorem 3.12 since $\sum_{j=p(n)^*+1}^{+\infty} |\phi_j|$ in expressions (56) and (61) are trivially $o_P(1)$. Notice that these results generalize the approaches of Masarotto (1990), Grigolleto (1998) and Kilian (1998), since here we only need an over-consistent order selection method for obtaining (23).

4 Simulations results

We compare the different sieve bootstrap approaches described in the previous section for the following models:

Model 1: $X_t = 0.75X_{t-1} - 0.5X_{t-2} + \varepsilon_t$

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

Model 1 was considered by Cao *et al.* (1997) and Model 2 by Pascual *et al.* (1998). As in those papers we used the following error distributions F_ε: 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 $F_1 \sim \mathcal{N}(-1, 1)$ and $F_2 \sim \mathcal{N}(9, 1)$. We take sample sizes $n = 50$, 100, and 200, leads $h = 1$ to $h = 5$, and nominal coverage $1 - \alpha = 0.8$ and 0.95.

To compare the different prediction intervals, we use their mean coverage and length, and the proportions of observations lying out to the left and to the right of the interval. 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 by (6) based on $B = 1000$ bootstrap resamples.

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 and M $\in$ {S, EnS1, EnS2, ExS1, ExS2}.

In steps 1 and 2 we obtain the "theoretical" and bootstrap interval lengths using $L_T = X^{\lceil R(1-\alpha/2)\rceil}_{T+h} - X^{\lceil R\alpha/2\rceil}_{T+h}$, and $L_M = Q^*_M(1-\alpha/2) - Q^*_M(\alpha/2)$. Finally, steps 1 to 3 are repeated $S = 1000$ times to obtain $C_{M,i}$, $L_{M,i}$ with $i = 1, \ldots, S$, and we calculate the estimates:

$$\begin{array}{lcl} \bar{C}_M & = & S^{-1}\sum C_{M,i} \\ SE(\bar{C}_M) & = & \left(S^{-1}(S-1)^{-1}\sum(C_{M,i} - \bar{C_M})^2\right)^{1/2} \\ \bar{L}_M & = & S^{-1}\sum L_{M,i} \\ SE(\bar{L}_M) & = & \left(S^{-1}(S-1)^{-1}\sum(L_{M,i} - \bar{L}_M)^2\right)^{1/2}. \end{array} \tag{24}$$

The different sieve bootstrap are denoted by:

S corresponds to the sieve bootstrap without introducing model uncertainty.

EnS1 the endogenous sieve bootstrap using $\widehat{p}$ in steps 2 - 6.

EnS2 the endogenous sieve bootstrap using p_{max} in steps 2 - 6.

ExS1 the exogenous sieve bootstrap using the moving missing block bootstrap in step 1a.

ExS2 the exogenous sieve bootstrap using the AICC information criterion probabilities (19) in step 1a.

In tables 1-3 and 4-6, we present the results for Model 1 and Model 2, using the three sample sizes and error distributions, nominal coverage 95%, and lead times $h = 1$ and 5. For ExS1, we report the results with $\ell = 3p_{max}$. The other possible combinations of parameters are available on request to the authors.

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

For Model 1 with gaussian errors, methods EnS1, ExS1 and ExS2 have a better performance than S in terms of mean coverage and length for all sample sizes and lead times. Method EnS2, which corresponds to Hjorth's proposal, has lower coverage than method S, revealing that not all ways of introducing model uncertainty produce the correct effect.

For Model 1 with exponential or with contaminated errors, similar results are observed, although in these cases also the EnS2 obtain a higher coverage than S. In terms of mean coverage, the ExS2 generally outperforms the other sieve bootstraps.

===> Tables 4 - 6 about here<===

We obtain similar results for Model 2. Notice that in this case, the sieve approach never uses the correct model. We observe that for $h = 1$ and all error distributions S method is outperformed by all sieve approaches that include model variability. But, again for $h = 3$, the Hjorth's proposal often has lower coverage than method S.

5 Conclusion

It has been shown by Masarotto (1990) and Grigolleto (1998) that if the order of the AR is unknown but finite it can obtained prediction intervals by bootstrap incorporating the sampling variability of $\widehat{p}$ with better coverage probabilities than those produced by standard bootstrap procedures. Their approaches could be affected by the selected order $\widehat{p}$. In this paper we have proposed two alternative methods that are less dependent on the initial selected order. We have shown that, for general linear models, if we use an AR approximation, we can derive a bootstrap procedure for building prediction intervals that has the two following properties: first, the procedure is consistent, that is, it generates as prediction a random variable that converges in conditional distribution to the variable we are interested in forecasting; second, Monte Carlo simulations show that the proposed procedure provides in general better coverage results than previous methods in general cases.

Appendix

Proof of Proposition 3.3: First, note that $\max_{1\leq c\leq C}|\widehat{\sigma}^2_{c,n} - \sigma^2_c| \xrightarrow{P} 0$ as $n \to \infty$, where σ^2_c is the c-step ahead error prediction variance. From Lemma 3.2 we have that $\widehat{p} \to \infty$ in probability, therefore $\widehat{\sigma}^2_{\widehat{p},n} \xrightarrow{P} \sigma^2$ as $n \to \infty$. In the other hand, $\sigma^2_1 \geq \cdots \geq \sigma^2_C > \sigma^2$, where the last inequality follows from assumption *A3*. Then, for a sufficiently large n and for all $1 \leq c \leq C$, we have that and $\widehat{\sigma}^2_{\widehat{p},n} < \widehat{\sigma}^2_{c,n}$.

By (19) we have,

$$\frac{\sum_{c=1}^{C} \Pr_{S_n}\{p = c\}}{\Pr_{S_n}\{p = \widehat{p}\}} = \frac{\sum_{c=1}^{C} \exp\left(-(n+2c)\widehat{\sigma}^2_{c,n}\right)}{\exp(-(n+2\widehat{p})\widehat{\sigma}^2_{\widehat{p},n})}. \tag{25}$$

Let's analyze a generic term in (25),

$$\frac{\exp\left(-(n+2c)\widehat{\sigma}^2_{c,n}\right)}{\exp\left(-(n+2\widehat{p})\widehat{\sigma}^2_{\widehat{p},n}\right)} = \exp\left(n\left(\widehat{\sigma}^2_{\widehat{p},n} - \widehat{\sigma}^2_{c,n}\right) + 2\widehat{p}\widehat{\sigma}^2_{\widehat{p},n} - 2c\widehat{\sigma}^2_{c,n}\right), \tag{26}$$

where the term $n\left(\widehat{\sigma}^2_{\widehat{p},n} - \widehat{\sigma}^2_{c,n}\right) = O_P(n)$ and goes to $-\infty$ in probability, as $n \to \infty$; and the other terms are $o_P(n^{1/2}) + O_P(1)$ and go to $+\infty$ in probability, as $n \to \infty$. Of course, the first term dominates the second one and expression (26) has limit equal to 0.■

Proof of Proposition 3.5: The vector $\widehat{\phi}^*_p = (\widehat{\phi}^*_{1,n}, \ldots, \widehat{\phi}^*_{p,n})'$ is defined by the bootstrap empirical Yule-Walker equations:

$$\widehat{\Gamma}_p^* \widehat{\phi}_p^* = -\widehat{\gamma}_p^*, \tag{27}$$

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

$$\begin{aligned} \left\| \widehat{\phi}_p^* - \phi_p \right\|_\infty &= \left\| (\Gamma_p^{-1} - \widehat{\Gamma}_p^{*\,-1}) \widehat{\gamma}_p^* + \Gamma_p^{-1} (\gamma_p - \widehat{\gamma}_p^*) \right\|_\infty \\ &\leq \|\Gamma_p^{-1} - \widehat{\Gamma}_p^{*\,-1}\|_{row} \left\| \widehat{\gamma}_p^* \right\|_\infty + \left\| \Gamma_p^{-1} \right\|_{row} \left\| \gamma_p - \widehat{\gamma}_p^* \right\|_\infty , \end{aligned} \tag{28}$$

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 assumption *A1* and *A2*, we have that $\|\Gamma_p\|_{row}$ and $\|\Gamma_p^{-1}\|_{row}$ are uniformly bounded in p. Since $\Gamma_p^{-1} - \widehat{\Gamma}_p^{*\,-1} = \Gamma_p^{-1} \left(\widehat{\Gamma}_p^* - \Gamma_p \right) \widehat{\Gamma}_p^{*\,-1}$, and $\|\widehat{\Gamma}_p^* - \Gamma_p\|_{row} \leq |\widehat{\gamma}_0^* - \gamma_0| + 2\|\widehat{\gamma}_p^* - \gamma_p\|_1$, we can concentrate our attention on this last term.

Since,

$$\|\widehat{\gamma}_p^* - \gamma_p\|_\infty \leq \|\widehat{\gamma}_p^* - \gamma_p\|_1 \leq p^{1/2} \|\widehat{\gamma}_p^* - \gamma_p\|_2, \tag{29}$$

to get convergence to zero in (28), it is enough to consider the last term in (29).

Now,

$$\begin{aligned} \|\widehat{\gamma}_p^* - \gamma_p\|_2^2 &= \textstyle\sum_{k=1}^p (\widehat{R}^*(k) - R(k))^2 \\ &\leq 2 \textstyle\sum_{k=1}^p (\widehat{R}^*(k) - \mathrm{E}^*[\widehat{R}^*(k)])^2 + 2 \sum_{k=1}^p (\mathrm{E}^*[\widehat{R}^*(k)] - R(k))^2 \\ &= 2(S_1 + S_2). \end{aligned} \tag{30}$$

We have that $S_2 = O_P((n/\log(n))^{-(2r-1)/(2r+2)})$, since

$$S_2 = \textstyle\sum_{k=1}^p \left(\mathrm{E}^*[\varepsilon_1^{*\,2}] \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \widehat{\psi}_{i,n} \widehat{\psi}_{j,n} \delta_{i+k,j} - \mathrm{E}[\varepsilon_1^2] \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \psi_i \psi_j \delta_{i+k,j} \right)^2 \tag{31}$$

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

Now,

$$\begin{aligned} S_2 &= \textstyle\sum_{k=1}^p \Big(\mathrm{E}^*[\varepsilon_1^{*\,2}] \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} (\widehat{\psi}_{i,n} \widehat{\psi}_{j,n} - \psi_i \psi_j) \delta_{i+k,j} \\ &\qquad + (\mathrm{E}^*[\varepsilon_1^{*\,2}] - \mathrm{E}[\varepsilon_1^2]) \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \psi_i \psi_j \delta_{i+k,j} \Big)^2 \\ &\leq 2 \textstyle\sum_{k=1}^p \left(\mathrm{E}^*[\varepsilon_1^{*\,2}] \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} (\widehat{\psi}_{i,n} \widehat{\psi}_{j,n} - \psi_i \psi_j) \delta_{i+k,j} \right)^2 \\ &\quad + 2 \textstyle\sum_{k=1}^p \left((\mathrm{E}^*[\varepsilon_1^{*\,2}] - \mathrm{E}[\varepsilon_1^2]) \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \psi_i \psi_j \delta_{i+k,j} \right)^2 = 2(I_1 + I_2) \end{aligned} \tag{32}$$

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

$$\sup_{i \in \mathbb{N}} |\widehat{\psi}_{i,n} - \psi_i| = O_{a.s.}((\log(n)/n)^{1/2}) + O_{a.s.}(p^{-r}) \tag{33}$$

and

$$\sup_{n\geq n_1}\sum_{i=0}^{+\infty} i^r|\widehat{\psi}_{i,n}| = O_{a.s.}(1), \tag{34}$$

where n_1 is a random variable.
Using the above results, we have that $I_1 = O_P((n/\log(n))^{-(2r-1)/(2r+2)})$, since

$$\begin{aligned} I_1 &= \mathrm{E}^*[\varepsilon_1^{*2}]^2 \sum_{k=1}^p \Big(\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}|\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \psi_i\psi_j|\delta_{i+k,j}\Big)^2 \\ &\leq \mathrm{E}^*[\varepsilon_1^{*2}]^2 \sum_{k=1}^p \Big(\sum_{i,j=0}^{+\infty}|\widehat{\psi}_{i,n}\psi_j - \psi_i\psi_j|\delta_{i+k,j} + \sum_{i,j=0}^{+\infty}|\widehat{\psi}_{i,n}\widehat{\psi}_{j,n} - \widehat{\psi}_{i,n}\psi_j|\delta_{i+k,j}\Big)^2 \\ &= \mathrm{E}^*[\varepsilon_1^{*2}]^2 \sum_{k=1}^p \Big(\sum_{i=0}^{+\infty}|\widehat{\psi}_{i,n} - \psi_i||\psi_{i+k}| + \sum_{j=0}^{+\infty}|\widehat{\psi}_{j,n} - \psi_j||\widehat{\psi}_{j-k,n}|\Big)^2 \\ &= O_P(p)(O_{a.s.}((\log(n)/n)^{1/2}) + O_{a.s.}(p^{-r}))^2 \\ &= O_P((n/\log n)^{-(2r-1)/(2r+2)}). \end{aligned} \tag{35}$$

Under assumptions *A1* and *B* of this proposition, we can establish a stronger conclusion than in Lemma 5.3 of Bühlmann (1997), in fact

$$\mathrm{E}^*[\varepsilon_t^{*2}] - \mathrm{E}[\varepsilon_t^2] = o_P((\log(n)/n)^{1/2}p) = O_P((n/\log n)^{-r/(2r+2)}). \tag{36}$$

Therefore, $I_2 = o_P((n/\log(n))^{-(2r-1)/(2r+2)})$.

For the other term in (30), we have $S_1 = O_P(n^{-1}(n/\log(n))^{1/(2r+2)})$, since

$$\begin{aligned} S_1 &= \sum_{k=1}^p \Big(n^{-1}\sum_{t=1}^{n-k}\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\varepsilon^*_{t-i}\varepsilon^*_{t+k-j} \\ &\qquad - \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\mathrm{E}^*[\varepsilon_1^{*2}]\delta_{i+k,j}\Big)^2 \\ &= \sum_{k=1}^p n^{-2}\sum_{t,s=1}^{n-k}\sum_{i,j,h,l=0}^{+\infty}\widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\widehat{\psi}_{h,n}\widehat{\psi}_{l,n}\Big(\varepsilon^*_{t-i}\varepsilon^*_{t+k-j} - \mathrm{E}^*[\varepsilon_1^{*2}]\delta_{i+k,j}\Big) \\ &\qquad \Big(\varepsilon^*_{s-h}\varepsilon^*_{s+k-l} - \mathrm{E}^*[\varepsilon_1^{*2}]\delta_{h+k,l}\Big). \end{aligned} \tag{37}$$

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

$$\begin{aligned} \mathrm{E}^*[S_1] &= \sum_{k=1}^p n^{-2}\sum_{t,s=1}^{n-k}\sum_{i,j,h,l=0}^{+\infty}\widehat{\psi}_{i,n}\widehat{\psi}_{j,n}\widehat{\psi}_{h,n}\widehat{\psi}_{l,n}\Big(\mathrm{E}^*[\varepsilon^*_{t-i}\varepsilon^*_{t+k-j}\varepsilon^*_{s-h}\varepsilon^*_{s+k-l}] \\ &\qquad -\mathrm{E}^*[\varepsilon_1^{*2}]^2\delta_{i+k,j}\delta_{h+k,l}\Big). \end{aligned} \tag{38}$$

Notice that

$$\mathrm{E}^*[\varepsilon^*_{t-i}\varepsilon^*_{t+k-j}\varepsilon^*_{s-h}\varepsilon^*_{s+k-l}] = \begin{cases} \mathrm{E}^*[\varepsilon_1^{*4}], & \text{if } t-i = t+k-j = s-h = s+k-l \\ \mathrm{E}^*[\varepsilon_1^{*2}]^2, & \text{if two pairs of different indexes} \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

then,

$$\mathrm{E}^*[\varepsilon^*_{t-i}\varepsilon^*_{t+k-j}\varepsilon^*_{s-h}\varepsilon^*_{s+k-l}] - \mathrm{E}^*[\varepsilon_1^{*2}]^2\delta_{i+k,j}\delta_{h+k,l} = \tag{40}$$

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

Using (34) 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 (38) 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, $\mathrm{E}^*[S_1] = O_P(pn^{-1}) = O_P(n^{-1}(n/\log n)^{1/(2r+2)})$.

Finally, we have

$$p^{1/2}\|\widehat{\gamma}_p^* - \gamma_p\|_2 = O_P((n/\log(n))^{-(r-1)/(2r+2)}) \tag{41}$$

and the assumption *A2* with $r > 1$ concludes the proof.■

Proof of Lemma 3.8: We have that

$$\begin{aligned} \widehat{\sigma}^{2\,*}_{p,n} &= n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\left(X^*_{t+1} + \widehat{\phi}_p^{*\prime}\mathbf{X}^*_{t,p}\right)^2 \\ &= n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\left(X^*_{t+1} + \phi_p'\mathbf{X}^*_{t,p} + \left(\widehat{\phi}_p^* - \phi_p^*\right)'\mathbf{X}^*_{t,p}\right)^2 \\ &= n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\left(X^*_{t+1} + \phi_p'\mathbf{X}^*_{t,p}\right)^2 + 2n^{-1}\sum_{t=p_{max}}^{n-1}\left(X^*_{t+1} + \phi_p'\mathbf{X}^*_{t,p}\right)\left(\widehat{\phi}_p^* - \phi_p^*\right)'\mathbf{X}^*_{t,p} \\ &\qquad + n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\left(\left(\widehat{\phi}_p^* - \phi_p^*\right)'\mathbf{X}^*_{t,p}\right)^2 = S_1 + S_2 + S_3, \end{aligned} \tag{42}$$

where $\mathbf{X}^*_{t,c} = (X^*_t, \ldots, X^*_{t+1-p})'$. To establish (22), we proof that S_1 goes to $\sigma_{p,n}$, and S_2 and S_3 are asymptotically negligible uniformly in p.

Using the proof of Proposition 3.5, we obtain that $S_3 = O_P((n/\log(n))^{-(r-2)/(2r+2)})$, since,

$$\begin{aligned} S_3 &= \left(\widehat{\phi}_p^* - \phi_p^*\right)' n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\mathbf{X}^*_{t,p}\mathbf{X}^{*\prime}_{t,p}\left(\widehat{\phi}_p^* - \phi_p^*\right) \\ &\leq \left\|\widehat{\phi}_p^* - \phi_p^*\right\|_2^2 \left\|n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\mathbf{X}^*_{t,p}\mathbf{X}^{*\prime}_{t,p}\right\|_{spec} \\ &= O_P((n/\log(n))^{-(r-2)/(2r+2)})O_P(1), \end{aligned} \tag{43}$$

and this bound does not depend on p.

To prove that S_1 goes to $\sigma_{p,n}$, we first obtain that $\mathrm{E}^*[S_1] - \sigma^2_{p,n} = O_P((n/\log n)^{-(r-1)/(2r+2)})$. We have,

$$\begin{aligned} \sigma^2_{p,n} &= \mathrm{E}\left[n^{-1}\textstyle\sum_{t=p_{max}}^{n-1} X^2_{t+1} + 2\phi_p' X_{t+1}\mathbf{X}_{t,p} + \phi_p'\mathbf{X}_{t,p}\mathbf{X}'_{t,p}\phi_p\right] \\ &= n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\mathrm{E}\Big[\sum_{i,j=0}^{+\infty}\psi_i\psi_j\varepsilon_{t+1-i}\varepsilon_{t+1-j} \\ &\qquad\qquad + 2\phi_p'\left(\sum_{i,j=0}^{+\infty}\psi_i\psi_j\varepsilon_{t+1-i}\varepsilon_{t+1-j-h}\right)_{1\leq h\leq p} \\ &\qquad\qquad + \phi_p'\left(\sum_{i,j=0}^{+\infty}\psi_i\psi_j\varepsilon_{t+1-i-h}\varepsilon_{t+1-j-l}\right)_{1\leq h,l\leq p}\phi_p\Big] \\ &= n^{-1}\textstyle\sum_{t=p_{max}}^{n-1}\Big(\sum_{i,j=0}^{+\infty}\psi_i\psi_j\mathrm{E}[\varepsilon_t^2]\delta_{i,j} + 2\phi_p'\Big(\sum_{i,j=0}^{+\infty}\psi_i\psi_j\mathrm{E}[\varepsilon_t^2]\delta_{i,j+h}\Big)_{1\leq h\leq p} \\ &\qquad\qquad + \phi_p'\left(\sum_{i,j=0}^{+\infty}\psi_i\psi_j\mathrm{E}[\varepsilon_t^2]\delta_{i+h,j+l}\right)_{1\leq h,l\leq p}\phi_p\Big), \end{aligned} \tag{44}$$

where $(.)_{1\leq h\leq p}$ denotes a $p\times 1$ vector and $(.)_{1\leq h,l\leq p}$ a $p\times p$ matrix.

Analogously for $\mathrm{E}^*[S_1]$,

$$\begin{aligned}\mathrm{E}^*[S_1] &= n^{-1}\sum_{t=p_{max}}^{n-1}\Big(\sum_{i,j=0}^{+\infty}\widehat{\psi}_i\widehat{\psi}_j\mathrm{E}^*[\varepsilon_t^{*2}]\delta_{i,j}+2\phi_p'\Big(\sum_{i,j=0}^{+\infty}\widehat{\psi}_i\widehat{\psi_j}\mathrm{E}^*[\varepsilon_t^{*2}]\delta_{i,j+h}\Big)_{1\leq h\leq p}\\ &\quad+\phi_p'\Big(\sum_{i,j=0}^{+\infty}\widehat{\psi}_i\widehat{\psi}_j\mathrm{E}^*[\varepsilon_t^{*2}]\delta_{i+h,j+l}\Big)_{1\leq h,l\leq p}\phi_p\Big).\end{aligned} \tag{45}$$

Then $|\mathrm{E}^*[S_1]-\sigma^2_{p,n}|$ depends on $\mathrm{E}^*[\varepsilon_t^{*2}]-\mathrm{E}[\varepsilon_t^2]=O_P((n/\log n)^{-r/(2r+2)})$ and some sums similar to $\sum_{i=0}^{+\infty}|\widehat{\psi}_i-\psi_i||\psi_{i+h}|=O_{a.s.}((\log(n)/n)^{1/2})+O_{a.s.}(p^{-r}))$, (see proof of Proposition 3.5).

Only remains to prove that $\mathrm{var}^*[S_1]$ goes to 0.

$$\begin{aligned}\mathrm{E}^*[S_1]^2 &= n^{-2}\sum_{t,s=p_{max}}^{n-1}\Big(\mathrm{E}^*[X_{t+1}^{*2}]+2\phi_p\mathrm{E}^*[X_{t+1}^*\mathbf{X}_{t,p}^*]+\phi_p'\mathrm{E}^*[\mathbf{X}_{t,p}^*\mathbf{X}_{t,p}^{*'}]\phi_p\Big)\\ &\quad\Big(\mathrm{E}^*[X_{s+1}^{*2}]+2\phi_p\mathrm{E}^*[X_{s+1}^*\mathbf{X}_{s,p}^*]+\phi_p'\mathrm{E}^*[\mathbf{X}_{s,p}^*\mathbf{X}_{s,p}^{*'}]\phi_p\Big)\end{aligned} \tag{46}$$

and

$$\begin{aligned}\mathrm{E}^*[S_1^2] &= n^{-2}\sum_{t,s=p_{max}}^{n-1}\mathrm{E}^*\Big[\Big(X_{t+1}^{*2}+2\phi_pX_{t+1}^*\mathbf{X}_{t,p}^*+\phi_p'\mathbf{X}_{t,p}^*\mathbf{X}_{t,p}^{*'}\phi_p\Big)\\ &\quad\Big(X_{s+1}^{*2}+2\phi_pX_{s+1}^*\mathbf{X}_{s,p}^*+\phi_p'\mathbf{X}_{s,p}^*\mathbf{X}_{s,p}^{*'}\phi_p\Big)\Big].\end{aligned} \tag{47}$$

Notice that $\mathrm{E}^*[S_1^2]-\mathrm{E}^*[S_1]^2$ depend on sums similar to:

$$\begin{aligned}&n^{-2}\sum_{t,s=p_{max}}^{n-1}\mathrm{E}^*[X_{t+1-i}^*X_{t+1-j}^*X_{s+1-h}^*X_{s+1-l}^*]-\mathrm{E}^*[X_{t+1-i}^*X_{t+1-j}^*]\mathrm{E}^*[X_{s+1-h}^*X_{s+1-l}^*]\\ &=n^{-2}\sum_{t,s=p_{max}}^{n-1}\sum_{i',j',h',l'=0}^{+\infty}\widehat{\psi}_{i'}\widehat{\psi}_{j'}\widehat{\psi}_{h'}\widehat{\psi}_{l'}\Big(\mathrm{E}^*[\varepsilon_{t+1-i-i'}\varepsilon_{t+1-j-j'}\varepsilon_{s+1-h-h'}\varepsilon_{s+1-l-l'}]\\ &\quad-\mathrm{E}^*[\varepsilon_{t+1-i-i'}\varepsilon_{t+1-j-j'}]\mathrm{E}^*[\varepsilon_{s+1-h-h'}\varepsilon_{s+1-l-l'}]\Big),\end{aligned} \tag{48}$$

which is $O_P\big(n^{-1}(n/\log n)^{1/(2r+2)}\big)$ by using similar arguments as in (38).
Finally, by the Cauchy-Schwarz inequality we have that $S_2=O_P\big(n^{-1/2}(n/\log n)^{-(r-3)/(4r+4)}\big)$. Then,

$$\big|\widehat{\sigma}^{2\,*}_{p,n}-\sigma^2_{p,n}\big|=O_P\big((n/\log n)^{-(r-2)/(2r+2)}\big), \tag{49}$$

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

Proof of Proposition 3.10: Suppose that there exits a constant C, $0<C<+\infty$, such that $\lim_{n\to\infty}\Pr^*\{\widehat{p}^*(n)\leq C\}>0$. This is equivalent to

$$\lim_{n\to\infty}\Pr^*\big\{\exists p'=p'(n)\leq C: S_n^*(p')\leq S_n^*(p)\big\}>0. \tag{50}$$

From Lemma 3.2 we have that $\widehat{p}$ is a divergent sequence, i.e. for any $0<C<+\infty$ we have that $\Pr\{\widehat{p}>C\}\to 1$. Then, (50) implies

$$\begin{gathered}\lim_{n\to\infty}\Pr^*\{\exists p'\leq C: S_n^*(p')\leq S_n^*(\widehat{p})\}>0\\ \Updownarrow\\ \lim_{n\to\infty}\Pr^*\Big\{\exists p'\leq C:\tfrac{n+2p'}{n}(\widehat{\sigma}^{2\,*}_{p',n}-\sigma^2_{p',n})-\tfrac{n+2\widehat{p}}{n}(\widehat{\sigma}^{2\,*}_{\widehat{p},n}-\sigma^2_{\widehat{p},n})\leq\tfrac{n+2\widehat{p}}{n}\sigma^2_{\widehat{p},n}-\tfrac{n+2p'}{n}\sigma^2_{p',n}\Big\}>0.\end{gathered} \tag{51}$$

By using (22) from Lemma 3.8, we have for all $\varepsilon > 0$ that

$$\lim_{n\to\infty} \Pr^* \left\{ \exists p' \leq C : -\varepsilon < \frac{n+2\widehat{p}}{n}\sigma^2_{\widehat{p},n} - \frac{n+2p'}{n}\sigma^2_{p',n} \right\} > 0. \tag{52}$$

By assumption *A3*, we have that $\sigma^2_C > \sigma^2$, and note that $\sigma^2_{\widehat{p},n} \xrightarrow{P} \sigma^2$, and $\sigma^2_C \leq \liminf \sigma^2_{p',n} \leq \limsup \sigma^2_{p',n} \leq \sigma^2_1$. Choosing a sufficiently small ε in order to get a contradiction with the above relation (52).■

Proof of Theorem 3.12: We can write X_{T+h} and X^*_{T+h} as:

$$X_{T+h} = -\sum_{j=1}^{+\infty} \phi_j X_{T+h-j} + \varepsilon_{T+h} \tag{53}$$

and

$$X^*_{T+h} = -\sum_{j=1}^{+\infty} \widehat{\phi}^*_{j,n} X^*_{T+h-j} + \varepsilon^*_{T+h}, \tag{54}$$

where $\widehat{\phi}^*_{j,n}$ denote the estimates of ϕ_j from a resample of size n, $\widehat{\phi}^*_{j,n} = 0$ for $j > p^*(n)$, and $X^*_t = X_t$ for $t \leq T$. For simplicity of notation we present the proof for $h = 1$.

From Lemma 5.4 of Bühlmann (1997), we have $\varepsilon^*_{T+1} \xrightarrow{d} \varepsilon_{T+1}$, in probability. Then, by the Slutzky lemma only remains to prove that the difference of the first terms in X^*_{T+1} and X_{T+1} goes to 0 in probability:

$$\begin{aligned} -\sum_{j=1}^{+\infty} (\widehat{\phi}^*_{j,n} - \phi_j) X_{T+1-j} &= -\sum_{j=1}^{p^*(n)} (\widehat{\phi}^*_{j,n} - \phi_j) X_{T+1-j} + \sum_{j=p^*(n)+1}^{+\infty} \phi_j X_{T+1-j} \\ &= S_1 + S_2 \end{aligned} \tag{55}$$

First, we have $S_2 = o_P((n/\log n)^{-r/(2r+2)})$, since

$$\mathrm{E}[|S_2|] \leq \mathrm{E}[|X_t|] \sum_{j=p(n)^*+1}^{+\infty} |\phi_j| = o(p^*(n)^{-r}) \tag{56}$$

and second, we establish that $S_1 = O_P((n/\log n)^{-r/(2r+2)})$. We have that

$$|S_1| \leq |\sum_{j=1}^{p^*(n)} (\widehat{\phi}^*_{j,n} - \phi_{j,n}) X_{T+1-j}| + |\sum_{j=1}^{p^*(n)} (\phi_{j,n} - \phi_j) X_{T+1-j}| = I_1 + I_2 \tag{57}$$

where $\phi_{p^*} = (\phi_{1,n}, \ldots, \phi_{p^*,n})^t$ are the theoretical Yule-Walker statistics.

For I_1 we use the following result in the proof of Proposition 3.5:

$$\max_{1\leq j\leq p(n)} |\widehat{\phi}^*_{j,n} - \phi_{j,n}| = O_P((n/\log n)^{-(r-1)/(2r+2)}). \tag{58}$$

Therefore,

$$\begin{aligned} I_1 &\leq \left(\sum_{j=1}^{p^*(n)}(\widehat{\phi}^*_{j,n}-\phi_{j,n})^2\right)^{1/2}\left(\sum_{j=1}^{p^*(n)} X^2_{T+1-j}\right)^{1/2} \\ &\leq p^*(n)^{1/2}\max_{1\leq j\leq p}|\widehat{\phi}^*_{j,n}-\phi_{j,n}|O_P(p^*(n)^{1/2}) = O_P(p^*(n)(\log n/n)^{1/2}) \\ &= O_P((n/\log n)^{-r/(2r+2)}). \end{aligned} \tag{59}$$

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

$$\sum_{j=0}^{p^*(n)}|\phi_{j,n}-\phi_j| \leq c\sum_{j=p^*(n)+1}^{+\infty}|\phi_j| \tag{60}$$

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

$$\mathrm{E}[I_2] \leq \mathrm{E}[|X_t|]\sum_{j=1}^{p^*(n)}|\phi_{j,n}-\phi_j| = o(p^*(n)^{-r}). \tag{61}$$

Finally,

$$-\sum_{j=1}^{+\infty}\widehat{\phi}_{j,n}X_{T+h-j} = \sum_{j=1}^{+\infty}\phi_j X_{T+1-j} + O_P((n/\log(n))^{-r/(2r+2)}) \tag{62}$$

Then, $X^*_{T+1} \xrightarrow{d} X^*_{T+1}$ in probability.

For general h, it is clear that we can write the difference of first terms in X_{T+h} and X^*_{T+h} as a sum of a continuous function $f(\phi_1,\ldots,\phi_{h-1},\widehat{\phi}^*_{1,n},\ldots,\widehat{\phi}^*_{h-1,n})$ (S_1+S_2), and a term similar to S_1+S_2. The second terms in X_{T+h} and X^*_{T+h} are a linear combination of the corresponding (and independent) errors $(\varepsilon_{T+1},\ldots,\varepsilon_{T+h},\varepsilon^*_{T+1},\ldots,\varepsilon^*_{T+h})$.■

References

[1] Akaike, H. (1969) Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, **21**, 243-247.

[2] 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.

[3] Alonso, A.M., Peña, D. and Romo, J. (2000) Resampling time series by missing values techniques, Working Paper 00-42, Universidad Carlos III de Madrid.

[4] Alonso, A.M., Peña, D. and Romo, J. (2001) Forecasting time series with sieve bootstrap, *J. Statist. Plann. Inference*, in press.

[5] Bhansali, R.J. (1983) A simulation study of autoregressive and window estimators of the inverse autocorrelation function, *Appl. Statist.*, **32** 141-149.

[6] Bhansali, R.J. (1993) Order selection for linear time series models: A review. In: *Developments in Time Series Analysis*, T. Subba Rao, ed., Chapman and Hall, London, 50-66.

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

[8] Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model selection: An integral part of inference, *Biometrics*, **53**, 603-618.

[9] Bühlmann, P. (1997) Sieve bootstrap for time series, *Bernoulli*, **8**, 123-148.

[10] Bühlmann, P. and Künsch, H. (1995) The blockwise bootstrap for general parameters of stationary time series, *Scand. J. Statist.*, **22**, 35-54.

[11] 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. Simulation Comput.*, **26**, 961-978.

[12] Carlstein, E., Do, K., Hall, P., Hesterberg, T. and Künsch, H. R. (1996) Matched-block bootstrap for dependent data, *Bernoulli*, **4**, 305-328.

[13] Grigoletto, M. (1998) Bootstrap prediction intervals for autoregresions: some alternatives, *Intern. J. Forecasting*, **14**, 447-456.

[14] Hall, P., Horowitz, J.L. and Jing, B.Y. (1995) On blocking rules for the bootstrap with dependent data, *Biometrika*, **82**, 561-574.

[15] Hjorth, J.S.U. (1994) *Computer Intensive Statistical Methods: Validation, Model Selection, and Bootstrap*, Chapman and Hall, London.

[16] Hurvich, C.M. and Tsai, C.L. (1989) Regression and time series model selection in small samples, *Biometrika*, **76**, 297-307.

[17] Karagrigoriou, A. (1997) Asymptotic efficiency of the order selection of a nongaussian AR process, *Statist. Sinica*, **7**, 407-423.

[18] Kass, R.E. and Raftery, A.E. (1995) Bayes Factors, *J. Amer. Statist. Assoc.*, **90**, 773-795.

[19] Kilian, L. (1998) Accounting for lag uncertainty in autoregressions: The endogenous lag order bootstrap algorithhm, *J. Time Ser. Anal.*, **19**, 531-548.

[20] Kreiss, J.-P. (1988) *Asymptotical Inference of Stochastic Processes*, Habilitationsschrift, Universität Hamburg.

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

[22] LeBlanc, M. and Tibshirani, R. (1996) Combining estimates in regresion and classification, *J. Amer. Statist. Assoc.*, **91**, 1641-1650.

[23] Liu, R.Y. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture week dependence, In: *Exploring the Limits of Bootstrap*, R. Lepage and L. Billard eds., 225-248, Wiley, New York.

[24] Masarotto, G. (1990) Bootstrap prediction intervals for autoregressions, *Intern. J. Forecasting*, **6**, 229-239.

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

[26] Peña, D. and Maravall, A. (1991) Interpolation, outliers, and inverse autocorrelations, *Comm. Statist. Theory Methods*, **20**, 3175-3186.

[27] Politis, D.N. and Romano, J.F. (1992) A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, *Ann. Statist.*, **20**, 1985-2007.

[28] Politis, D.N. and Romano, J.F. (1994) Large sample confidence regions based on subsamples under minimal assumptions, *Ann. Statist.*, **22**, 2031-2050.

[29] Pötscher, B.M. (1990) Estimation of autoregressive moving-average order given an infinite number of models and approximation of spectral densities, *J. Time Ser. Anal.*, **11**, 165-179.

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

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

[32] Schwarz, G. (1978) Estimating the dimension of a model, *Ann. Statist.*, **6**, 461-464.

[33] Shibata, R. (1976) Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.

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

Table 1: Simulation results for Model 1, with Gaussian Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	3.93
1	50	S	92.27 (0.13)	3.92/3.81	3.83 (0.02)
		EnS1	92.59 (0.12)	3.75/3.66	3.87 (0.02)
		EnS2	92.12 (0.13)	3.99/3.89	3.81 (0.02)
		ExS1	92.56 (0.12)	3.79/3.65	3.85 (0.02)
		ExS2	92.74 (0.12)	3.69/3.57	3.88 (0.02)
	100	S	93.53 (0.09)	3.17/3.30	3.88 (0.01)
		EnS1	93.83 (0.08)	3.00/3.18	3.92 (0.01)
		EnS2	93.21 (0.09)	3.33/3.46	3.85 (0.01)
		ExS1	93.77 (0.09)	3.07/3.16	3.92 (0.01)
		ExS2	93.96 (0.08)	2.97/3.07	3.96 (0.01)
	200	S	94.28 (0.06)	2.96/2.75	3.91 (0.01)
		EnS1	94.46 (0.06)	2.88/2.67	3.93 (0.01)
		EnS2	93.67 (0.07)	3.25/3.08	3.84 (0.01)
		ExS1	94.52 (0.06)	2.81/2.68	3.94 (0.01)
		ExS2	94.68 (0.06)	2.75/2.57	3.99 (0.01)
h	n	Theoretical	95%	2.50% / 2.50%	5.20
5	50	S	92.01 (0.12)	4.03/3.96	4.86 (0.02)
		EnS1	92.18 (0.12)	3.91/3.92	4.88 (0.02)
		EnS2	91.62 (0.13)	4.19/4.19	4.81 (0.02)
		ExS1	92.12 (0.13)	3.92/3.96	4.87 (0.02)
		ExS2	92.29 (0.12)	3.82/3.89	4.91 (0.02)
	100	S	93.47 (0.09)	3.27/3.26	5.02 (0.02)
		EnS1	93.66 (0.08)	3.19/3.15	5.05 (0.02)
		EnS2	93.11 (0.09)	3.48/3.41	4.97 (0.02)
		ExS1	93.71 (0.08)	3.17/3.12	5.06 (0.02)
		ExS2	93.93 (0.08)	3.08/2.99	5.13 (0.02)
	200	S	94.21 (0.06)	2.92/2.87	5.13 (0.01)
		EnS1	94.36 (0.06)	2.83/2.81	5.15 (0.01)
		EnS2	93.62 (0.07)	3.21/3.17	5.03 (0.01)
		ExS1	94.43 (0.06)	2.79/2.78	5.17 (0.01)
		ExS2	94.66 (0.06)	2.67/2.66	5.20 (0.01)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).

Table 2: Simulation results for Model 1, with Exponential Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	3.68
1	50	S	92.75 (0.21)	2.93/4.32	3.70 (0.03)
		EnS1	93.42 (0.18)	2.36/4.23	3.76 (0.03)
		EnS2	93.40 (0.18)	2.16/4.45	3.74 (0.03)
		ExS1	93.22 (0.18)	2.60/4.18	3.75 (0.03)
		ExS2	93.31 (0.20)	2.45/4.24	3.79 (0.03)
	100	S	93.81 (0.15)	3.01/3.18	3.78 (0.02)
		EnS1	94.66 (0.12)	2.18/3.16	3.82 (0.02)
		EnS2	94.74 (0.11)	1.70/3.56	3.80 (0.02)
		ExS1	94.49 (0.12)	2.37/3.14	3.82 (0.02)
		ExS2	94.58 (0.13)	2.30/3.12	3.82 (0.02)
	200	S	94.47 (0.13)	1.72/2.81	3.75 (0.01)
		EnS1	95.11 (0.11)	2.07/2.82	3.78 (0.02)
		EnS2	95.40 (0.10)	1.37/3.23	3.82 (0.02)
		ExS1	95.14 (0.11)	2.07/2.79	3.81 (0.02)
		ExS2	95.23 (0.11)	1.99/2.77	3.83 (0.02)
h	n	Theoretical	95%	2.50% / 2.50%	5.20
5	50	S	91.94 (0.16)	3.73/4.33	4.89 (0.04)
		EnS1	92.15 (0.16)	3.53/4.32	4.93 (0.04)
		EnS2	91.98 (0.16)	3.42/4.60	4.86 (0.04)
		ExS1	92.00 (0.16)	3.68/4.32	4.91 (0.04)
		ExS2	92.31 (0.16)	3.41/4.28	4.94 (0.04)
	100	S	93.43 (0.11)	3.20/3.37	5.06 (0.03)
		EnS1	93.74 (0.11)	2.92/3.33	5.10 (0.03)
		EnS2	93.54 (0.11)	2.67/3.79	5.01 (0.03)
		ExS1	93.70 (0.11)	2.93/3.36	5.09 (0.03)
		ExS2	94.08 (0.10)	2.60/3.32	5.18 (0.03)
	200	S	94.28 (0.08)	2.80/2.92	5.16 (0.02)
		EnS1	94.48 (0.08)	2.64/2.88	5.18 (0.02)
		EnS2	94.26 (0.08)	2.32/3.41	5.07 (0.02)
		ExS1	94.54 (0.08)	2.55/2.91	5.19 (0.02)
		ExS2	94.89 (0.08)	2.22/2.89	5.21 (0.02)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).

Table 3: Simulation results for Model 1, with Contaminated Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	12.60
1	50	S	92.92 (0.17)	2.85/4.23	12.53 (0.06)
		EnS1	93.48 (0.15)	2.35/4.17	12.67 (0.06)
		EnS2	93.49 (0.15)	1.84/4.67	12.65 (0.07)
		ExS1	93.27 (0.14)	2.54/4.19	12.63 (0.06)
		ExS2	93.66 (0.14)	2.27/4.06	12.76 (0.06)
	100	S	93.97 (0.10)	2.45/3.58	12.69 (0.03)
		EnS1	94.46 (0.10)	2.04/3.50	12.81 (0.03)
		EnS2	93.96 (0.11)	1.28/4.76	12.78 (0.04)
		ExS1	94.31 (0.11)	2.15/3.54	12.80 (0.03)
		ExS2	94.68 (0.11)	1.94/3.37	12.83 (0.03)
	200	S	94.35 (0.09)	2.42/3.23	12.70 (0.02)
		EnS1	94.81 (0.07)	1.99/3.20	12.78 (0.02)
		EnS2	94.41 (0.08)	0.89/4.70	12.87 (0.02)
		ExS1	94.94 (0.07)	1.89/3.17	12.85 (0.02)
		ExS2	95.30 (0.08)	1.65/3.05	13.14 (0.02)
h	n	Theoretical	95%	2.50% / 2.50%	15.75
5	50	S	92.81 (0.17)	3.59/3.61	15.66 (0.08)
		EnS1	93.12 (0.17)	3.35/3.53	15.81 (0.08)
		EnS2	92.84 (0.18)	3.20/3.96	15.68 (0.08)
		ExS1	92.96 (0.17)	3.49/3.55	15.72 (0.08)
		ExS2	93.38 (0.16)	3.20/3.42	15.94 (0.08)
	100	S	93.82 (0.11)	3.14/3.04	15.82 (0.05)
		EnS1	94.14 (0.11)	2.87/2.99	15.94 (0.05)
		EnS2	93.89 (0.12)	2.61/3.51	15.87 (0.06)
		ExS1	94.20 (0.11)	2.86/2.94	16.01 (0.06)
		ExS2	94.68 (0.11)	2.51/2.82	16.32 (0.06)
	200	S	94.28 (0.09)	2.97/2.75	15.85 (0.04)
		EnS1	94.57 (0.08)	2.75/2.68	15.96 (0.04)
		EnS2	94.33 (0.09)	2.38/3.30	15.90 (0.04)
		ExS1	94.65 (0.08)	2.69/2.66	16.02 (0.04)
		ExS2	95.06 (0.08)	2.31/2.53	16.04 (0.04)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).

Table 4: Simulation results for Model 2, with Gaussian Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	3.93
1	50	S	91.30 (0.19)	4.05/4.65	3.96 (0.02)
		EnS1	91.69 (0.19)	3.89/4.43	4.03 (0.02)
		EnS2	91.77 (0.18)	3.90/4.33	3.96 (0.02)
		ExS1	91.67 (0.18)	3.93/4.40	4.00 (0.02)
		ExS2	92.00 (0.18)	3.77/4.23	4.00 (0.02)
	100	S	93.00 (0.11)	3.58/3.43	3.93 (0.01)
		EnS1	93.26 (0.11)	3.42/3.32	3.97 (0.01)
		EnS2	93.02 (0.11)	3.53/3.45	3.90 (0.01)
		ExS1	93.36 (0.11)	3.33/3.31	3.96 (0.01)
		ExS2	93.59 (0.11)	3.24/3.17	3.99 (0.01)
	200	S	93.77 (0.07)	3.07/3.16	3.91 (0.01)
		EnS1	93.90 (0.07)	3.01/3.08	3.93 (0.01)
		EnS2	93.50 (0.08)	3.24/3.26	3.86 (0.01)
		ExS1	94.00 (0.07)	2.97/3.03	3.93 (0.01)
		ExS2	94.37 (0.07)	2.82/2.80	3.98 (0.01)
h	n	Theoretical	95%	2.50% / 2.50%	4.94
5	50	S	91.69 (0.13)	4.08/4.23	4.63 (0.02)
		EnS1	91.88 (0.13)	3.99/4.13	4.64 (0.02)
		EnS2	91.54 (0.13)	4.17/4.29	4.59 (0.02)
		ExS1	91.75 (0.13)	4.05/4.20	4.62 (0.02)
		ExS2	91.93 (0.13)	3.99/4.08	4.63 (0.02)
	100	S	93.03 (0.09)	3.51/3.46	4.75 (0.01)
		EnS1	93.20 (0.09)	3.45/3.35	4.77 (0.01)
		EnS2	92.83 (0.10)	3.60/3.57	4.71 (0.01)
		ExS1	93.21 (0.09)	3.38/3.42	4.76 (0.01)
		ExS2	93.30 (0.09)	3.33/3.37	4.76 (0.01)
	200	S	93.63 (0.07)	3.17/3.20	4.80 (0.01)
		EnS1	93.82 (0.07)	3.08/3.09	4.83 (0.01)
		EnS2	93.32 (0.07)	3.34/3.34	4.75 (0.01)
		ExS1	93.80 (0.07)	3.10/3.10	4.82 (0.01)
		ExS2	94.08 (0.06)	2.95/2.97	4.87 (0.01)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).

Table 5: Simulation results for Model 2, with Exponential Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	3.68
1	50	S	92.00 (0.28)	3.75/4.25	3.96 (0.03)
		EnS1	92.53 (0.24)	3.28/4.19	4.01 (0.03)
		EnS2	92.58 (0.25)	3.26/4.16	3.96 (0.03)
		ExS1	92.87 (0.23)	3.09/4.04	4.02 (0.03)
		ExS2	92.91 (0.22)	3.00/4.09	4.01 (0.03)
	100	S	93.09 (0.22)	3.44/3.46	3.91 (0.02)
		EnS1	93.64 (0.20)	2.98/3.38	3.97 (0.02)
		EnS2	93.93 (0.17)	2.52/3.55	3.90 (0.02)
		ExS1	93.99 (0.18)	2.65/3.36	3.96 (0.02)
		ExS2	94.40 (0.16)	2.24/3.36	3.99 (0.02)
	200	S	93.98 (0.19)	1.91/3.11	3.86 (0.02)
		EnS1	94.40 (0.16)	2.45/3.15	3.88 (0.02)
		EnS2	94.63 (0.13)	1.95/3.43	3.83 (0.02)
		ExS1	94.67 (0.15)	2.23/3.09	3.89 (0.02)
		ExS2	95.03 (0.12)	1.93/3.05	3.93 (0.02)
h	n	Theoretical	95%	2.50% / 2.50%	4.86
5	50	S	91.90 (0.15)	3.35/4.74	4.65 (0.03)
		EnS1	92.12 (0.15)	3.15/4.72	4.66 (0.03)
		EnS2	91.82 (0.15)	3.48/4.70	4.63 (0.03)
		ExS1	92.07 (0.15)	3.26/4.67	4.66 (0.03)
		ExS2	92.19 (0.15)	3.18/4.63	4.66 (0.03)
	100	S	92.89 (0.11)	3.37/3.74	4.76 (0.02)
		EnS1	93.09 (0.11)	3.21/3.70	4.80 (0.03)
		EnS2	92.86 (0.11)	3.25/3.90	4.73 (0.03)
		ExS1	93.08 (0.11)	3.18/3.74	4.77 (0.03)
		ExS2	93.29 (0.10)	3.02/3.69	4.80 (0.03)
	200	S	93.18 (0.09)	3.32/3.30	4.81 (0.02)
		EnS1	93.59 (0.08)	3.14/3.27	4.84 (0.02)
		EnS2	93.13 (0.08)	3.18/3.59	4.76 (0.02)
		ExS1	93.65 (0.08)	3.10/3.25	4.84 (0.02)
		ExS2	93.84 (0.07)	2.90/3.26	4.87 (0.02)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).

Table 6: Simulation results for Model 2, with Contaminated Errors.

Lag	Sample	Method	$\bar{C}_M$ (se)	Cov. (b./a.)	$\bar{L}_M$ (se)
h	n	Theoretical	95%	2.50% / 2.50%	12.60
1	50	S	92.38 (0.28)	2.88/4.74	13.08 (0.07)
		EnS1	93.11 (0.24)	2.24/4.65	13.27 (0.08)
		EnS2	92.88 (0.21)	2.55/4.58	13.07 (0.07)
		ExS1	93.27 (0.23)	2.25/4.48	13.28 (0.07)
		ExS2	93.27 (0.24)	2.17/4.56	13.27 (0.07)
	100	S	93.41 (0.18)	2.35/4.24	13.10 (0.04)
		EnS1	93.97 (0.16)	1.86/4.17	13.23 (0.04)
		EnS2	93.68 (0.16)	1.65/4.66	13.00 (0.04)
		ExS1	94.11 (0.16)	1.75/4.14	13.22 (0.04)
		ExS2	94.40 (0.15)	1.46/4.14	13.23 (0.04)
	200	S	94.17 (0.12)	2.06/3.77	13.02 (0.02)
		EnS1	94.67 (0.10)	1.70/3.63	13.14 (0.02)
		EnS2	94.15 (0.10)	1.24/4.60	12.97 (0.02)
		ExS1	94.86 (0.10)	1.54/3.60	13.14 (0.02)
		ExS2	95.09 (0.09)	1.45/3.46	13.16 (0.02)
h	n	Theoretical	95%	2.50% / 2.50%	14.86
5	50	S	93.82 (0.17)	2.24/3.94	15.28 (0.09)
		EnS1	93.97 (0.17)	2.11/3.92	15.28 (0.08)
		EnS2	93.70 (0.17)	2.34/3.96	15.13 (0.08)
		ExS1	93.91 (0.17)	2.12/3.98	15.24 (0.09)
		ExS2	93.96 (0.17)	2.05/3.99	15.24 (0.09)
	100	S	94.62 (0.11)	2.09/3.29	15.35 (0.05)
		EnS1	94.83 (0.10)	1.93/3.24	15.41 (0.05)
		EnS2	94.45 (0.11)	1.98/3.57	15.19 (0.06)
		ExS1	94.91 (0.10)	1.92/3.17	15.43 (0.05)
		ExS2	95.02 (0.10)	1.82/3.16	15.43 (0.05)
	200	S	94.69 (0.08)	2.41/2.89	15.23 (0.04)
		EnS1	94.99 (0.07)	2.20/2.81	15.35 (0.04)
		EnS2	94.61 (0.08)	2.08/3.31	15.19 (0.04)
		ExS1	95.10 (0.07)	2.11/2.79	15.38 (0.04)
		ExS2	95.18 (0.07)	2.03/2.69	15.44 (0.04)

NOTE: Standard error (se) are in parentheses. $\bar{C}_M$, $\bar{L}_M$ and se's are computed from (24).