Statistica Sinica **14**(2004), 155-174

INTRODUCING MODEL UNCERTAINTY IN TIME SERIES BOOTSTRAP

Andrés M. Alonso[1], Daniel Peña[2] and Juan Romo[2]

[1] *Universidad Autonóma de Madrid and* [2] *Universidad Carlos III de Madrid*

Abstract: It is common in a parametric bootstrap to select the model from the data, and then treat it as it were the true model. Kilian (1998) illustrates 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.

Key words and phrases: Model uncertainty, prediction, sieve bootstrap, time series.

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 this question assumes that the series $\{X_t\}_{t\in\mathbb{Z}}$ follows a linear finite dimension model with a known error 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 to plug in the linear predictors. In addition, some bootstrap approaches have been proposed to avoid the use of a specified error distribution, see e.g., Stine (1987) and Thombs and Schucany (1990) for AR(p) models, and Pascual, Romo and Ruiz (2001) for ARMA(p, q) models. But those bootstrap proposals assume that p and q are known. Alonso, Peña and Romo (2002) 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 selects an approximating autoregressive model AR$(\widehat{p})$ from the data, and then uses the selected order as if it were the true order; however, this approach ignores the variability involved in model selection that 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 unknown. Thus, we should select a model from the data. Many model

selection procedures have been proposed, e.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, 1991) 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 use of 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 proceed 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 4, Monte Carlo simulations reveal that the version with $\widehat{p}$ is generally preferable to the one with $p_{\max}$.

Since the previous endogenous order bootstrap could be affected by the initial estimated order, we propose a way of introducing the sampling variability of the model selection procedure that does not depend, or is less dependent, on $\widehat{p}$. This approach is 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 our Monte Carlo study, the results show that the proposed approach outperforms the endogenous lag order bootstrap of Kilian (1998).

The remainder of the paper is organized as follows. In Section 2 we extend the endogenous lag order method of Kilian (1998) applied to sieve bootstrap prediction intervals and we introduce the proposal based on information criteria. 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 method 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 $\mathrm{E}[X_t] =$

μ_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, first proposed by Kreiss (1988, 1992) and extended by Paparoditis (1996), Bühlmann (1997) and Inoue and Kilian (2002). The method proceeds as follows.

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

The AICC criterion of $-n \log(\sigma^2) + 2(p+1)n/(n-p-2)$ is a bias-corrected version of AIC that has a more extreme penalty for large-order models to counteract the overfitting nature of AIC (see, Hurvich and Tsai (1989, 1991)). An advantage of using AICC (in finite AR processes) is that the value of the maximum cut-off $p_{\max}$ has virtually no effect on the model chosen by this criterion, while for many of the other criteria increasing the value of $p_{\max}$ leads to increased overfitting of the model, see Hurvich and Tsai (1989). 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.

2. Construct some estimators of the autoregressive coefficients $(\widehat{\phi}_1, \ldots, \widehat{\phi}_{\widehat{p}})$. Following Kreiss (1992) and Bühlmann (1997) we take the Yule-Walker estimates.
3. Compute the residuals $\widehat{\varepsilon}_t = \sum_{j=0}^{\widehat{p}} \widehat{\phi}_j (X_{t-j} - \bar{X})$, $\widehat{\phi}_0 = 1,\ t \in (\widehat{p}+1, \ldots, n)$.
4. Define the empirical distribution function of the centred residuals $\widehat{F}_{\widetilde{\varepsilon}}(x) = (n-\widehat{p})^{-1} \sum_{t=\widehat{p}+1}^{n} 1_{\{\widetilde{\varepsilon}_t \le x\}}$, 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 $n - \widehat{p}$ i.i.d. observations from $\widehat{F}_{\widetilde{\varepsilon}}$.
6. Define X_t^* by the recursion:

$$\sum_{j=0}^{\widehat{p}} \widehat{\phi}_j (X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{2}$$

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

In practice we generate an AR($\widehat{p}$) resample using (2) with sample size equal to $n + 100$ and then discard the first 100 observations.

Up to this step, the resampling plan coincides with the sieve bootstrap, and is valid for bootstrapping statistics defined as functionals 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 $\{X_1^*, \ldots, X_n^*\}$, select the order $\widehat{p}^*$ of the autoregressive approximation as in 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, Febrero-Bande, González-Manteiga, Prada-Sánchez and García-Jurado (1997) in fixing the last p observations, in our case the last $\widehat{p}^*$ observations, we can obtain resamples of the future values X_{T+h}^* given $X_{T-p^*+1}^* = X_{T-p^*+1}, \ldots, X_T^* = X_T$.

The next step is performed in order to introduce the variability associated to the estimation of the parameters in the predictions (3). Notice that the algorithm with steps 1-6 and 8 is similar to the procedure proposed by Paparoditis (1996), who estimate the distribution of a vector of autoregressive and moving average parameters, which is more than we need in our case. We only need consistent estimators, $\widehat{\phi}_j^*$, of the autoregressive parameters (see Proposition 2 below).

8. Compute the estimation of the autoregressive coefficients $(\widehat{\phi}_1^*, \ldots, \widehat{\phi}_{\widehat{p}^*}^*)$ as in step 2.
9. Compute future bootstrap observations by the recursion

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{\widehat{p}^*} \widehat{\phi}_j^* (X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \tag{3}$$

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

Finally, the bootstrap distribution, $F_{X_{T+h}^*}^*$, 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

$$\left[Q^*(\alpha/2), Q^*(1-\alpha/2)\right], \tag{4}$$

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

The consistency of the intervals in (4) follows from Lemma 2, Proposition 2, Proposition 3 and Theorem 1, presented in Section 3.

Notice that, if we omit step 7 and use $\widehat{p}^* = \widehat{p}$ in step 8 and the recursion (3), the resampling plan coincides with the sieve bootstrap prediction algorithm of Alonso, Peña and Romo (2002). The two approaches will be compared in the Monte Carlo study of Section 4.

Notice that the previous algorithm constructs two types of bootstrap replicates based on (forward) residuals $\widehat{\varepsilon}_t$. The first type has n observations and they are used to obtain bootstrap estimates of the model parameters. The second

type of replicates fixes the $\widehat{p}^*$ bootstrap observations, $(X^*_{T-\widehat{p}^*+1}, \ldots, X^*_T)$, equal to the last $\widehat{p}^*$ original observations, $(X_{T-\widehat{p}^*+1}, \ldots, X_T)$, which are then used to compute the bootstrap forecast through (3). This allows us to obtain predictions conditional on the observed data. The second type of replicates includes the fix $\widehat{p}^*$ plus the new h observations.

This procedure differs from the proposal of Thombs and Schucany (1990) which fixes the last p bootstrap observations, i.e., $X^*_{T-p+1} = X_{T-p+1}, \ldots, X^*_T = X_T$, and obtains the bootstrap observations $X^*_1, \ldots, X^*_{T-p}$ using backward residuals and the backward representation of finite AR models. As before, bootstrap forecasts are computed using (3) and the forward residuals. In Thombs and Schucany's proposal, a replicate has $n + h$ observations. Two main criticisms apply to Thombs and Schucany's procedure: (i) if the forward errors are non-Gaussian then the backward errors are not independent, so it is necessary to modify the original Thombs and Schucany procedure (see, Breidt, Davis and Dunsmuir (1995)); (ii) its use is restricted to models with finite backward representation (i.e., it excludes models with nontrivial moving average components).

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 a possible implementation based on information criteria. Let $\{X_t\}_{t\in\mathbb{Z}}$ be 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 $\widehat{p} = \operatorname{argmin}_{0\le p\le p_{\max}} \{IC(p)\}$, i.e., we have estimates of

$$\operatorname{pr}\{\widehat{p} = p\}, \text{ for } 0 \le p \le p_{\max}. \tag{5}$$

The characterization of the asymptotic limit of (5) is a standard way of studying the consistency of information criteria in finite autoregressive models, see, e.g., Theorem 1 of Shibata (1976). A method for obtaining an approximation of (5) will be presented below and it constitutes the first step of the sieve exogenous order bootstrap.

The sieve exogenous order bootstrap modifies the previous steps as follows:

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

3a. Compute the residuals for the model with $p = \widehat{p}$ as $\widehat{\varepsilon}_t = \sum_{j=0}^{\widehat{p}} \widehat{\phi}_j^{(\widehat{p})}(X_{t-j} - \bar{X})$, $t \in (\widehat{p}+1, \ldots, n)$.

4a. Define the empirical distribution functions of the centred residuals $\widehat{F}_{\tilde{\varepsilon}}(x) = (n-\widehat{p})^{-1} \sum_{t=\widehat{p}+1}^{n} 1_{\{\tilde{\varepsilon}_t \le x\}}$, where $\tilde{\varepsilon}_t = \widehat{\varepsilon}_t - \widehat{\varepsilon}^{(\cdot)}$, and $\widehat{\varepsilon}^{(\cdot)} = (n-\widehat{p})^{-1} \sum_{t=\widehat{p}+1}^{n} \widehat{\varepsilon}_t$.

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

Notice that, given the probabilities in (5), the variability of $\widehat{p}$ enters exogenously to the bootstrap algorithm.

5a. Draw a resample ε_t^* of $n - p^*$ i.i.d. observations from $\widehat{F}_{\widetilde{\varepsilon}}$.
6a. Define X_t^* by the recursion:

$$\sum_{j=0}^{p^*} \widehat{\phi}_j^{(p^*)}(X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{6}$$

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

At this point, in the endogenous order bootstrap, we perform step 7, i.e., we select the order $\widehat{p}^*$. Note that this is not necessary in the exogenous approach. Moreover, we reduce the computational cost by skipping step 7 with respect to the endogenous order bootstrap.

7a. Estimate the autoregressive coefficients $(\widehat{\phi}_1^*, \ldots, \widehat{\phi}_{p^*}^*)$ as in step 2.
8a. Compute the future bootstrap observations by the recursion $X_{T+h}^* - \bar{X} = -\sum_{j=1}^{p^*} \widehat{\phi}_j^*(X_{T+h-j}^* - \bar{X}) + \varepsilon_{T+h}^*$, 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 - 8a are repeated B times in order to obtain $\widehat{F}_{X_{T+h}^*}^*$.

Next, we develop a way of obtaining an estimator or an approximation of the probabilities (5). This constitutes the first step of the sieve exogenous order bootstrap.

Information criterion function order distribution. This approach is related to the Bayesian way of accounting for model uncertainty, see, e.g., Section 6 of Kass and Raftery (1995). Assume that $p_{\max} + 1$ autoregressive models are being considered and a sample $\mathbf{X}$ of size n, has been observed. Then the posterior probability of model AR(p) is

$$\operatorname{pr}\{X \sim \mathrm{AR}(p)|\mathbf{X}\} = \frac{\operatorname{pr}\{\mathbf{X}|X \sim \mathrm{AR}(p)\}\operatorname{pr}\{X \sim \mathrm{AR}(p)\}}{\sum_{i=0}^{p_{\max}} \operatorname{pr}\{\mathbf{X}|X \sim \mathrm{AR}(i)\}\operatorname{pr}\{X \sim \mathrm{AR}(i)\}}, \tag{7}$$

where $X \sim \mathrm{AR}(p)$ denotes that X follows an AR(p) model. Particularly, $X \sim \mathrm{AR}(0)$ is equivalent to assuming that $\{X_t\}_{t\in\mathbb{Z}}$ is a white noise process. If we divide the numerator and denominator of (7) by $\operatorname{pr}\{\mathbf{X}|X \sim \mathrm{AR}(0)\}\operatorname{pr}\{X \sim \mathrm{AR}(0)\}$, we obtain

$$\operatorname{pr}\{X \sim \mathrm{AR}(p)|\mathbf{X}\} = \frac{\alpha_p B_{p0}}{\sum_{i=0}^{p_{\max}} \alpha_i B_{i0}}, \tag{8}$$

where $B_{p0} = \text{pr}\{\mathbf{X}|X \sim \text{AR}(p)\}/\text{pr}\{\mathbf{X}|X \sim \text{AR}(0)\}$ is called the Bayes factors for AR(p) against AR(0), and $\alpha_p = \text{pr}\{X \sim \text{AR}(p)\}/\text{pr}\{X \sim \text{AR}(0)\}$ is called the prior odds for model AR(p) against model AR(0); here $B_{00} = \alpha_0 = 1$.

Once we have (8), we calculate the posterior distribution function of X_{T+h} that takes into account the model uncertainty as

$$F_{X_{T+h}}(x) = \sum_{p=0}^{p_{\max}} F^{(p)}_{X_{T+h}}(x)\text{pr}\{X \sim \text{AR}(p)|\mathbf{X}\}, \tag{9}$$

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 (8) in step 4b allows us to obtain an approximation to (9). On the other hand, in the endogenous order bootstrap, we obtain an approximation of $F_{X_{T+h}}(x) = \sum_{p=0}^{p_{\max}} F^{(p)}_{X_{T+h}}(x)\text{pr}\{\widehat{p}^* = p|\mathbf{X}^* \sim \text{AR}(\widehat{p})\}$.

Since calculating the $p_{\max}$ Bayes factors involved in (8) is computationally arduous, and we do not have a priori information about the underlying model, here we use the approximation $B_{p0} \approx \exp(S_{p0})$ (as recommend Kass and Raftery (1995)), where S_{p0} is the Schwarz criterion, given by

$$S_{p0} = L_p - L_0 - \frac{1}{2}p\log n, \tag{10}$$

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 a simple linear transformation of the consistent BIC(p) objective function. Then (8) could be approximated by:

$$\text{pr}\{\text{AR}(p)|\mathbf{X}\} \approx \frac{\alpha_p \exp(-1/2\text{BIC}(p))}{\sum_{i=0}^{p_{\max}} \alpha_i \exp(-1/2\text{BIC}(i))}, \tag{11}$$

where $\text{BIC}(i) = n\log(2\pi\widehat{\sigma}_i^2) + i\log n$ and $\widehat{\sigma}_i^2$ is an estimate of the residual variance in the AR(i) model. The first summand in BIC(i) is $-2L_p$. In our case, we consider that the $p_{\max}+1$ models are equally probable a priori, so $\alpha_i = 1$ for $0 \le i \le p_{\max}$.

A word of caution about the approximation of Bayes factor, B_{p0}, by $\exp(S_{p0})$. It is based on the following result: $(S_{p0} - \log B_{p0})/\log B_{p0} \to 0$. Thus, S_{p0} may be viewed as a rough approximation to the logarithm of the Bayes factor (see, e.g., Kass and Wasermann (1995)). On the other hand, the quality of this approximation is not important in our asymptotic results. Notice that the exogenous bootstrap can be interpreted as a procedure that selects an order p^* taking into account the estimated weights (probabilities) $\text{pr}\{X \sim \text{AR}(p)|\mathbf{X}\}$ defined in (11),

and then builds resamples following an AR(p^*) model. To establish the asymptotic validity of the exogenous bootstrap we only need that estimated weights have a convenient limit, see our Proposition 1 and Remark 1.

Buckland, Burnham and Augustin (1997) propose a similar approach in the context of Poisson regression, line transient sampling and survival models, but they do not provide a theoretical justification of the proposed bootstrap methods. Also, they recommend using AIC(p) instead of BIC(p) at (11).

In the Monte Carlo study of Section 4, the sieve exogenous order bootstrap based on (8), with the above approximation, performs reasonably well. We present the results with the AICC model selection procedure, but additional simulations studies (available on request to the authors) illustrate that the sieve exogenous order bootstrap based on (8) with the BIC procedure performs similarly.

The consistency of the intervals at (4) using (11) follows from Lemma 2, Proposition 1, Proposition 2 and Theorem 1 in Section 3.

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) $L_k/\sum_{i=1}^{K} L_i$ 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 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 $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ a sequence of independent random variables with $\mathrm{E}[\varepsilon_t] \equiv 0$, and $\mathrm{E}[|\varepsilon_t|^s] < \infty$ for some $s \geq 4$.

Assumption A2. $\Psi(z) = \sum_{j=0}^{+\infty} \psi_j z^j$ 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)).

Since the model selection procedures have a different asymptotic behaviour for finite and for (nontrivial) infinite autoregressive processes (see, e.g., Theorems

3.1 and 4.1 of Pötscher (1990)), we present the theoretical results for two types of linear models.

Assumption A3. $\{X_t\}_{t\in\mathbb{Z}}$ does not degenerate to a finite order AR process.

Assumption A4. $\{X_t\}_{t\in\mathbb{Z}}$ is an AR(p_0) process for some finite p_0.

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 $\widehat{\phi}_p = (\widehat{\phi}_{1,n}, \dots \widehat{\phi}_{p,n})'$ satisfy the empirical Yule Walker equations $\widehat{\Gamma}_p\widehat{\phi}_p = -\widehat{\gamma}_p$, where $\widehat{\Gamma}_p = [\widehat{R}(i-j)]_{1\leq i,j\leq p}$, $\widehat{\gamma}_p = (\widehat{R}(1), \dots, \widehat{R}(p))^t$, and $\widehat{R}(j) = n^{-1}\sum_{t=1}^{n-|j|}(X_t - \bar{X})(X_{t+|j|} - \bar{X})$.

Notice that our Assumption B differs from Assumption B in Bühlmann (1997) and from Assumption 1(d) in Inoue and Kilian (2002), since in our case we restrict the behaviour of $p_{\max}(n)$ instead of restricting $p(n)$.

Assumption C. The model selection criterion has convergence rate C_n, i.e., it satisfies the following condition:

$$C_n^{-1} \max_{0\leq p\leq p_{\max}(n)} |\mathrm{pr}\,\{\widehat{p} = p\} - \pi_p| = O(1), \tag{12}$$

where $\widehat{p} = \mathrm{argmin}_{0\leq p\leq p_{\max}}\{IC(p)\}$ is the selected order, $\pi_p = \lim_{n\to\infty}\mathrm{pr}\,\{\widehat{p} = p\}$ and C_n is a non-random sequence.

Zhang (1993) establishes that AIC and BIC have convergence rate $O(n^{-1/2})$ in the case of linear regression models and i.i.d. errors. In Lemma 3, Proposition 2 and 3 and Theorem 1 we suppose the model selection criterion has convergence rate $C_n = O((n/\log n)^{-\delta})$ for some $\delta > 0$.

We present the results for the order selection method proposed by Shibata (1980):

$$S_n(p) = (n+2p)\widehat{\sigma}_{p,n}^2, \tag{13}$$

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

The following two lemmas characterize the asymptotic behavior of the selected order sequence $\{\widehat{p}(n)\}$. Lemma 1 is a consequence of Theorem 3.1, Theorem 4.1 and Remark 5.2 of Pötscher (1990), and Lemma 2 is the analogous to Corollary 4.1 of Shibata (1980), based on Theorem 3.1 of Karagrigoriou (1997).

Lemma 1. *Suppose that A1 with $s = 4$, A2 with $r = 1$, and B hold. Then, the random sequence $\widehat{p} = \widehat{p}(n)$ that minimizes $S_n(p)$, satisfies*
1. *under Assumption A3, $\{\widehat{p}(n)\}$ is a divergent sequence;*
2. *under Assumption A4, $\mathrm{pr}\,\{\widehat{p}(n) \geq p_0\} \to 1$ as $n \to \infty$.*

Let $L_n(p) = p\sigma^2/n + \|\boldsymbol{\phi}_p - \boldsymbol{\phi}\|_\Gamma$, where $\boldsymbol{\phi}_p = (\phi_{1,n}, \ldots, \phi_{p,n})'$ are the theoretical Yule-Walker statistics (the index n in $\phi_{i,n}$ is used to differentiate $\phi_{i,n}$ and the i-th autoregressive parameter ϕ_i), let $\|x\|_A = (x'Ax)^{1/2}$, let Γ be the infinite dimensional covariance matrix and $\boldsymbol{\phi}$ the infinite dimensional vector of autoregressive parameters. Also, denote by $\{p_0(n)\}$ the non-random sequence that minimizes $L_n(p)$, and let $\{p_\varepsilon(n)\}$ be the non-random sequence defined by $\{p_\varepsilon(n)\} = \min\{p : L_n(p)/L_n(p_0(n)) \leq 1 + \varepsilon\}$.

Lemma 2. *Suppose that A1 with $s = 16$, A2 with $r = 1$, A3 and B hold. Then, for any $\varepsilon > 0$, the random sequence $\widehat{p} = \widehat{p}(n)$ that minimizes $S_n(p)$ satisfies*

$$\operatorname{pr}\{\widehat{p}(n) \geq p_\varepsilon(n)\} \to 1, \text{ as } n \to \infty. \tag{14}$$

Now we use $S_n(p)$ in (11) instead of BIC(p). In the following proposition we establish that if $\{X_t\}_{t\in\mathbb{Z}}$ is $A3$ and we select a random order $p = p(n)$ taking into account the probability function defined by (11), then the probability of selecting p in any finite set is zero in comparison with the probability of $p = \widehat{p}$.

Proposition 1. *Suppose that A1 with $s = 16$, A2 with $r = 1$, A3 and B hold. Then for any positive integer C, $0 < C < +\infty$, we have*

$$\frac{\sum_{c=1}^{C} \operatorname{pr}_{S_n}\{p = c\}}{\operatorname{pr}_{S_n}\{p = \widehat{p}\}} \to 0 \quad \textit{in probability}, \tag{15}$$

where pr_{S_n} *denotes* (11) *calculated with S_n, i.e.,* $\operatorname{pr}_{S_n}\{p = c\} = \{\exp(-0.5(n+2c)\widehat{\sigma}^2_{c,n})\}/\{\sum_{p=0}^{p_{\max}} \exp(-0.5(n+2p)\widehat{\sigma}^2_{p,n})\}$.

Remark 1. Analogously, if $\{X_t\}_{t\in\mathbb{Z}}$ satisfies A4, 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 1(2) and that $\sigma_c^2 > \sigma^2$ holds for $c < p_0$, where σ_c^2 is the c-step ahead error prediction variance.

Remark 2. Proposition 1 holds for any divergent sequence $\{p(n)\}$ such that $p(n) = o(n^{1/2})$. Also notice that, if $\{X_t\}_{t\in\mathbb{Z}}$ satisfies $A3$, (15) implies that for any $0 < C < +\infty$, $\operatorname{pr}_{S_n}\{p \leq C\} \xrightarrow{P} 0$, and similarly if $\{X_t\}_{t\in\mathbb{Z}}$ verifies $A4$, for any $0 < c < p_0$, $\operatorname{pr}_{S_n}\{p \leq c\} \xrightarrow{P} 0$ as $n \to \infty$.

Proposition 2. *Suppose that A1 with $s = 16$, A2 with $r > 2$, A3, B with $p_{\max} = o((n/\log(n))^{1/(2r+2)})$ and C with $C_n = o((n/\log(n))^{-k/(2r+2)})$, for some $k > 3$, hold. Additionally, we suppose the random sequence $\{p(n)\}$ satisfies* (14). *Then* $\max_{1\leq j\leq p_{\max}(n)} |\widehat{\phi}^*_{j,n} - \phi_{j,n}| \longrightarrow 0$ *in probability.*

Remark 3. Notice that Lemma 2 (Proposition 1) and Proposition 2 imply that the endogenous (exogenous based on (11)) sieve bootstrap provides consistent estimators of the theoretical Yule-Walker statistics when $\{X_t\}_{t\in\mathbb{Z}}$ satisfies $A3$.

Remark 4. The statement of Proposition 2 holds when $\{X_t\}_{t\in\mathbb{Z}}$ satisfies *A*4, since the proof of Theorem 3.1 and 3.2 of Bühlmann (1995) can be modified in order to avoid the assumption that $p \to \infty$. Notice that under *A*4 we have $\max_{1\leq j\leq p_{\max}} |\widehat{\phi}_{j,n} - \phi_{j,n}| = O_{a.s.}((\log(n)/n)^{1/2})$ and $\phi_{j,n} = \phi_j = 0$ for $j > p_0$.

Remark 5. Theorem 3.1 and 3.3 of Paparoditis (1996) and Theorem 1 of Inoue and Kilian (2002) establish a stronger result than our Proposition 2 (both papers deal with the distribution of the increasing size vector of statistics $\widehat{\boldsymbol{\phi}}_p$), but they impose the condition $\sum_{j=1}^{\infty} |\phi_j|(1+\eta)^j < \infty$ for some $\eta > 0$, whereas *A*1 implies $\sum_{j=1}^{\infty} j^r|\phi_j| < \infty$ with $r > 0$.

Lemma 3. *Suppose that A*1 *with* $s = 16$, *A*2 *with* $r > 2$, *B with* $p_{\max} = o((n/\log(n))^{1/(2r+2)})$ *and C with* $C_n = o((n/\log(n))^{-k/(2r+2)})$, *for some* $k > 3$, *hold. Additionally, we suppose the random sequence* $\{\widehat{p}(n)\}$ *satisfies* (14). *Then*

$$\max_{1\leq p\leq p_{\max}(n)} \left|\widehat{\sigma}^{2\,*}_{p,n} - \sigma^2_{p,n}\right| \longrightarrow 0 \text{ in probability}, \tag{16}$$

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

Remark 6. Lemma 3 holds when $\{X_t\}_{t\in Z}$ satisfies *A*4, following arguments similar to those in Remark 4.

Proposition 3. *Suppose that A*1 *with* $s = 16$, *A*2 *with* $r > 2$, *A*3, *B with* $p_{\max} = o((n/\log(n))^{1/(2r+2)})$ *and C with* $C_n = o((n/\log(n))^{-k/(2r+2)})$, *for some* $k > 3$, *hold. Additionally, we suppose the random sequence* $\{\widehat{p}(n)\}$ *satisfies* (14). *Then the random sequence* $\widehat{p}^* = \widehat{p}^*(n)$ *that minimizes* $S^*_n(p) = (n+2p)\widehat{\sigma}^{2\,*}_{p,n}$ *is a divergent sequence.*

Remark 7. Under *A*4, by Lemma 1(2) and Remark 6, we can establish an analogous result to Proposition 3, i.e., $\widehat{p}^* = \widehat{p}^*(n)$ that minimizes $S^*_n(p)$ satisfies $\mathrm{pr}\{\widehat{p}(n) \geq p_0\} \to 1$ as $n \to \infty$, since now we have that $\sigma^2_c > \sigma^2$ for $0 < c < p_0$.

Theorem 1. *Suppose that A*1 *with* $s = 16$, *A*2 *with* $r > 2$, *B with* $p_{\max} = o((n/\log(n))^{1/(2r+2)})$ *and C with* $C_n = o((n/\log(n))^{-k/(2r+2)})$, *for some* $k > 3$, *hold. Additionally, we suppose the random sequence* $\{\widehat{p}(n)\}$ *satisfies* (14). *Then*

$$X^*_{T+h}\big|_{\left(X^*_T = X_T, \ldots, X^*_{T-p^*(n)} = X_{T-p^*(n)}\right)} \xrightarrow{d^*} X_{T+h}\big|_{X^T_{-\infty}}, \text{ in probability}. \tag{17}$$

Remark 8. Under *A*4, by Lemma 1(2) and Remark 4, we can establish an analogous result to Theorem 1 since $\sum_{j=p(n)^*+1}^{+\infty} |\phi_j|$ is trivially $o_P(1)$. Notice that these results generalize the approaches of Masarotto (1990), Grigoletto (1998) and Kilian (1998), since here we only need an order selection method that satisfies $\mathrm{pr}\{\widehat{p} \geq p_0\} \to 1$ to obtain (17).

4. Simulations Results

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

Model 1: $(1 + 0.7B - 0.2B^2)X_t = \varepsilon_t$.

Model 2: $X_t = (1 + 0.7B - 0.2B^2)\varepsilon_t$.

As in Cao et al. (1997) and Pascual, Romo and Ruiz (2001), 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 took sample sizes $n = 25$, 50 and 100, leads $h = 1$ to $h = 5$, and nominal coverage $1 - \alpha = 0.95$.

To compare the different prediction intervals, we use mean coverage and length, and the proportions of observations lying 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 $R = 1,000$ future values X_{T+h}.
2. For each bootstrap procedure obtain the $(1 - \alpha)\%$ prediction interval by (4) based on $B = 1,000$ 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}, $r = 1, \ldots, R$, are the R future values generated in first step, $Q^*_M(\cdot)$ are the quantiles of the estimated bootstrap distribution and M $\in$ {S, EnS1, EnS2, ExS}.

In steps 1 and 2 we obtain the "theoretical" and bootstrap interval lengths using $L_T = X^{(\lceil R(1-\alpha/2)\rceil)}_{T+h} - X^{(\lceil R\alpha/2\rceil}_{T+h})$, where $X^{(\cdot)}_{T+h}$ denotes an element in the ordered sample and $L_M = Q^*_M(1 - \alpha/2) - Q^*_M(\alpha/2)$. Finally, steps 1 to 3 are repeated $S = 1,000$ times to obtain $C_{M,i}$, $L_{M,i}$ with $i = 1, \ldots, S$, and we calculate the estimates

$$\begin{aligned} \bar{C}_M &= S^{-1} \sum C_{M,i}, \\ se(\bar{C}_M) &= \Big(S^{-1}(S-1)^{-1} \sum (C_{M,i} - \bar{C_M})^2\Big)^{1/2}, \\ \bar{L}_M &= S^{-1} \sum L_{M,i}, \\ se(\bar{L}_M) &= \Big(S^{-1}(S-1)^{-1} \sum (L_{M,i} - \bar{L}_M)^2\Big)^{1/2}. \end{aligned} \tag{18}$$

The different sieve bootstrap are as follows: **S**, 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; **ExS**, the exogenous sieve bootstrap using the AICC information criterion expressions (11) in step 1a.

In Tables 1 and 2, we present the results of (18) and the estimated above and below coverage for Model 1 and Gaussian and exponential errors, using the least squares estimators for autoregressive parameters, the three sample sizes, nominal coverage 95%, and lead times $h = 1$ and 5. Also, we include the results for standard Gaussian forecast intervals based on the Box-Jenkins (BJ) approach assuming the known order p (q) of Model 1 (2). Notice that, in Table 1, where the error distribution is Gaussian, the results of BJ could be interpreted as benchmarks. In Alonso, Peña and Romo (2003) we present the remaining simulation results.

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%	3.94
1	25	BJ	90.86 (0.19)	4.60 / 4.54	3.67 (0.02)
		S	89.99 (0.24)	4.86 / 5.15	3.81 (0.02)
		EnS1	90.85 (0.21)	4.34 / 4.81	3.94 (0.02)
		EnS2	87.21 (0.28)	5.84 / 6.95	3.79 (0.03)
		ExS	90.42 (0.21)	4.86 / 4.73	3.94 (0.02)
	50	BJ	93.13 (0.10)	3.44 / 3.44	3.83 (0.01)
		S	91.25 (0.16)	4.15 / 4.60	3.77 (0.02)
		EnS1	92.65 (0.13)	3.49 / 3.87	3.92 (0.02)
		EnS2	90.41 (0.17)	4.62 / 4.97	3.73 (0.02)
		ExS	93.00 (0.12)	3.49 / 3.51	3.95 (0.02)
	100	BJ	94.07 (0.07)	2.95 / 2.98	3.86 (0.01)
		S	93.17 (0.10)	3.46 / 3.37	3.86 (0.01)
		EnS1	93.72 (0.09)	3.17 / 3.10	3.91 (0.01)
		EnS2	91.25 (0.12)	4.41 / 4.34	3.65 (0.01)
		ExS	93.97 (0.09)	2.95 / 3.08	3.94 (0.01)
h	n	Theoretical	95%	2.50% / 2.50%	6.46
5	25	BJ	90.43 (0.26)	4.66 / 4.91	6.05 (0.04)
		S	89.90 (0.30)	4.89 / 5.21	6.24 (0.05)
		EnS1	90.87 (0.28)	4.45 / 4.67	6.40 (0.05)
		EnS2	85.39 (0.36)	6.89 / 7.73	5.58 (0.06)
		ExS	89.96 (0.25)	5.07 / 4.97	6.10 (0.05)
	50	BJ	93.22 (0.16)	3.47 / 3.31	6.19 (0.03)
		S	92.26 (0.20)	3.67 / 4.07	6.40 (0.04)
		EnS1	93.07 (0.17)	3.28 / 3.64	6.53 (0.04)
		EnS2	87.94 (0.27)	5.73 / 6.33	5.72 (0.04)
		ExS	92.63 (0.17)	3.71 / 3.66	6.38 (0.04)
	100	BJ	93.25 (0.12)	3.48 / 3.26	6.26 (0.02)
		S	93.08 (0.13)	3.60 / 3.32	6.33 (0.03)
		EnS1	93.83 (0.12)	3.16 / 3.02	6.48 (0.03)
		EnS2	89.00 (0.18)	5.60 / 5.40	5.62 (0.03)
		ExS	93.62 (0.12)	3.21 / 3.18	6.40 (0.03)

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	25	BJ	91.87 (0.19)	0.27 / 7.86	3.56 (0.03)
		S	88.40 (0.39)	6.96 / 4.64	3.74 (0.04)
		EnS1	90.38 (0.33)	5.03 / 4.60	3.95 (0.04)
		EnS2	87.46 (0.42)	6.28 / 6.26	3.88 (0.05)
		ExS	90.75 (0.32)	4.66 / 4.59	4.00 (0.05)
	50	BJ	93.58 (0.12)	0.16 / 6.26	3.78 (0.02)
		S	90.80 (0.30)	5.13 / 4.08	3.65 (0.03)
		EnS1	93.28 (0.22)	2.87 / 3.85	3.83 (0.03)
		EnS2	92.83 (0.24)	2.19 / 4.98	3.85 (0.04)
		ExS	93.83 (0.20)	2.44 / 3.72	3.87 (0.03)
	100	BJ	94.09 (0.07)	0.01 / 5.90	3.82 (0.02)
		S	93.25 (0.22)	3.50 / 3.25	3.72 (0.02)
		EnS1	94.93 (0.15)	1.89 / 3.19	3.82 (0.02)
		EnS2	94.62 (0.16)	0.89 / 4.49	3.72 (0.02)
		ExS	95.09 (0.13)	1.83 / 3.09	3.85 (0.02)
h	n	Theoretical	95%	2.50% / 2.50%	6.69
5	25	BJ	87.92 (0.29)	4.92 / 7.16	5.79 (0.06)
		S	88.59 (0.31)	5.74 / 5.67	6.21 (0.07)
		EnS1	89.32 (0.29)	5.27 / 5.41	6.33 (0.07)
		EnS2	83.82 (0.42)	8.30 / 7.88	5.55 (0.08)
		ExS	88.79 (0.29)	5.45 / 5.76	6.20 (0.07)
	50	BJ	91.15 (0.19)	3.64 / 5.21	6.11 (0.04)
		S	91.46 (0.22)	4.29 / 4.24	6.53 (0.06)
		EnS1	92.06 (0.20)	3.90 / 4.04	6.63 (0.06)
		EnS2	87.65 (0.29)	6.26 / 6.10	5.81 (0.06)
		ExS	91.85 (0.20)	4.05 / 4.10	6.50 (0.06)
	100	BJ	92.54 (0.12)	2.92 / 4.54	6.20 (0.03)
		S	93.31 (0.14)	3.26 / 3.43	6.61 (0.04)
		EnS1	93.74 (0.13)	2.99 / 3.27	6.73 (0.04)
		EnS2	89.44 (0.20)	5.24 / 5.32	5.78 (0.04)
		ExS	93.38 (0.13)	3.33 / 3.29	6.60 (0.04)

For Model 1 with Gaussian errors, methods EnS1 and ExS have a slightly better performance than S in terms of mean coverage for all sample sizes and lead times (up to a 1.5% more of mean coverage for $h = 1$). Notice that EnS1 and ExS are near to BJ results. 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 errors and $h = 5$, similar results are observed; for $h = 1$, EnS1 and ExS have a better performance than S in terms of mean

coverage (more than 1% in mean coverage in all cases, and up to a 3% in sample size $n = 50$). In these cases, BJ cannot be considered as a benchmark, since it assumes an incorrect error distribution (see its asymmetric proportions of above and below coverage in fifth column of Table 2).

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, the method S is outperformed by the EnS1 and ExS approaches that include model variability, in some cases up to a 4% more of mean coverage. A global measure of a procedure M's improvement by introducing model uncertainty can be computed as $(\bar{C}_M - \bar{C}_S)/(0.95 - \bar{C}_S)$, where $\bar{C}_S$ is the mean coverage of the bootstrap method without including model uncertainty. These increases are up to 70%. In almost all cases, both endogenous approaches are improved by the ExS method, which is always better than S. Again, in the Gaussian case, EnS1 and ExS are near to BJ results. In some cases ExS is superior to BJ, revealing the importance of parameter estimation variability. In Alonso, Peña and Romo (2003) we present the simulation results for Model 2.

Acknowledgement

We would like to thank two referees and the coordinating editor for a carefully reading that greatly improved the paper. We gratefully acknowledge the financial support by the Dirección General de Investigación Científica y Técnica of Spain and Cátedra de Calidad BBVA.

Appendix

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

By (11) we have,

$$\frac{\sum_{c=1}^{C} \mathrm{pr}_{S_n}\{p = c\}}{\mathrm{pr}_{S_n}\{p = \widehat{p}\}} = \frac{\sum_{c=1}^{C} \exp\left(-0.5(n+2c)\widehat{\sigma}^2_{c,n}\right)}{\exp(-0.5(n+2\widehat{p})\widehat{\sigma}^2_{\widehat{p},n})}. \tag{19}$$

Analyzing the generic term in (19), $\exp(0.5n(\widehat{\sigma}^2_{\widehat{p},n} - \widehat{\sigma}^2_{c,n}) + \widehat{p}\widehat{\sigma}^2_{\widehat{p},n} - c\widehat{\sigma}^2_{c,n})$, we realize that the term $n(\widehat{\sigma}^2_{\widehat{p},n} - \widehat{\sigma}^2_{c,n}) = O_P(n)$ 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. Then, the generic term goes to 0 as $n \to \infty$.

Proof of Proposition 2. The vector $\widehat{\boldsymbol{\phi}}_p^* = (\widehat{\phi}_{1,n}^*, \ldots, \widehat{\phi}_{p,n}^*)'$ is defined by the bootstrap empirical Yule-Walker equations: $\widehat{\Gamma}_p^* \widehat{\boldsymbol{\phi}}_p^* = -\widehat{\boldsymbol{\gamma}}_p^*$, where $\widehat{\Gamma}_p^* = [\widehat{R}^*(i-j)]_{i,j=1}^p$, $\widehat{\boldsymbol{\gamma}}_p^* = (\widehat{R}^*(1), \ldots, \widehat{R}^*(p))^t$, and $\widehat{R}^*(j) = n^{-1}\sum_{t=1}^{n-|j|}(X_t^* - \bar{X}^*)(X_{t+|j|}^* - \bar{X}^*)$. Then

$$\|\widehat{\boldsymbol{\phi}}_p^* - \boldsymbol{\phi}_p\|_\infty \leq \|\Gamma_p^{-1} - \widehat{\Gamma}_p^{*\,-1}\|_{row}\|\widehat{\boldsymbol{\gamma}}_p^*\|_\infty + \|\Gamma_p^{-1}\|_{row}\|\boldsymbol{\gamma}_p - \widehat{\boldsymbol{\gamma}}_p^*\|_\infty, \tag{20}$$

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 $\|\widehat{\Gamma}_p^* - \Gamma_p\|_{row} \leq |\widehat{\gamma}_0^* - \gamma_0| + 2\|\widehat{\boldsymbol{\gamma}}_p^* - \boldsymbol{\gamma}_p\|_1$, we can concentrate our attention on this last term. To get convergence to zero in (20), it is enough to consider the inequation

$$\begin{aligned}\|\widehat{\boldsymbol{\gamma}}_{p_{\max}}^* - \boldsymbol{\gamma}_{p_{\max}}\|_2^2 &\leq 2\sum_{k=1}^{p_{\max}}(\widehat{R}^*(k) - \mathrm{E}^*[\widehat{R}^*(k)])^2 + 2\sum_{k=1}^{p}(\mathrm{E}^*[\widehat{R}^*(k)] - R(k))^2 \\ &= 2(S_1 + S_2).\end{aligned} \tag{21}$$

We have that $S_2 = O_P((n/\log n)^{\max\{-(2r-3),-(2k-3)\}/(2r+2)})$, since

$$S_2 = \sum_{k=1}^{p_{\max}}\left(\mathrm{E}^*[{\varepsilon_1^*}^2]\sum_{p=0}^{p_{\max}}\sum_{i,j=0}^{+\infty}\widehat{\psi}_{i,n}^p\widehat{\psi}_{j,n}^p\delta_{i+k,j}\Pr^*\{p^*=p\} - \mathrm{E}[\varepsilon_1^2]\sum_{i,j=0}^{+\infty}\psi_i\psi_j\delta_{i+k,j}\right)^2, \tag{22}$$

where $\delta_{i,j} = 1$ if $i = j$, and 0 otherwise, $\widehat{\Psi}_p(z) = \sum_{i=0}^{+\infty}\widehat{\psi}_{i,n}^p z^i = \widehat{\Phi}_p(z)^{-1}$, $\widehat{\Phi}_p(z)$ is a polynomial of degree p estimated in the original sample, and $\Pr^*\{p^*=p\}$ is the probability under the exogenous bootstrap. In the following we omit the index n in $\widehat{\psi}_{j,n}$. We have

$$\begin{aligned}S_2 \leq\ & 2\sum_{k=1}^{p_{\max}}\left(\mathrm{E}^*[{\varepsilon_1^*}^2]\sum_{p=0}^{p_{\max}}\sum_{i,j=0}^{+\infty}(\widehat{\psi}_i^p\widehat{\psi}_j^p - \psi_i\psi_j)\delta_{i+k,j}\Pr^*\{p^*=p\}\right)^2 \\ & +2\sum_{k=1}^{p_{\max}}\left((\mathrm{E}^*[{\varepsilon_1^*}^2] - \mathrm{E}[\varepsilon_1^2])\sum_{i,j=0}^{+\infty}\psi_i\psi_j\delta_{i+k,j}\right)^2 = 2(I_1 + I_2).\end{aligned} \tag{23}$$

Under Assumption C and (14) we can obtain similar results to Theorem 3.1 and 3.2 of Bühlmann (1995) which allows us to establish that

$$I_1 = O_P((n/\log n)^{\max\{-(2r-1),-(2k-3)\}/(2r+2)}). \tag{24}$$

Under assumptions $A1$ and B of this proposition, we can establish that $\mathrm{E}^*[{\varepsilon_t^*}^2] - \mathrm{E}[\varepsilon_t^2] = O_P((n/\log n)^{-r/(2r+2)})$. Therefore, $I_2 = o_P((n/\log(n))^{-(2r-1)/(2r+2)})$. Finally, from (23) and (24), we have that $S_2 = O_P((n/\log n)^{\max\{-(2r-3),-(2k-3)\}/(2r+2)})$.

For the other term in (21) we have $S_1 = O_P((n/\log n)^{-(k-2)/(2r+2)})$, since

$$S_1 \leq \sum_{k=0}^{p_{\max}} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j,h,l=0}^{+\infty} \sum_{p,q=0}^{p_{\max}} \left(\widehat{\psi}_i^{p^*}\widehat{\psi}_j^{p^*}\varepsilon_{t-i}^*\varepsilon_{t+k-j}^* - \widehat{\psi}_i^{p}\widehat{\psi}_j^{p}\mathrm{E}^*[{\varepsilon_1^*}^2]\delta_{i+k,j}\right)$$
$$\left(\widehat{\psi}_h^{p^*}\widehat{\psi}_l^{p^*}\varepsilon_{s-h}^*\varepsilon_{s+k-l}^* - \widehat{\psi}_h^{q}\widehat{\psi}_l^{q}\mathrm{E}^*[{\varepsilon_1^*}^2]\delta_{h+k,l}\right)\Pr^*\{p^*=p\}\Pr^*\{p^*=q\}\,. \quad (25)$$

Then

$$\mathrm{E}^*[S_1] \leq \sum_{k=0}^{p_{\max}} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j,h,l=0}^{+\infty} \sum_{p,r=0}^{p_{\max}} \left(\widehat{\psi}_i^{r}\widehat{\psi}_j^{r}\widehat{\psi}_h^{r}\widehat{\psi}_l^{r}\mathrm{E}^*[\varepsilon_{t-i}^*\varepsilon_{t+k-j}^*\varepsilon_{s-h}^*\varepsilon_{s+k-l}^*]\right.$$
$$\left. -\widehat{\psi}_i^{r}\widehat{\psi}_j^{r}\widehat{\psi}_h^{p}\widehat{\psi}_l^{p}\mathrm{E}^*[{\varepsilon_1^*}^2]^2\delta_{i+k,j}\delta_{h+k,l}\right)\Pr^*\{p^*=p\}\Pr^*\{p^*=r\}\,. \quad (26)$$

Note that we can decompose the sum $\sum_{k=0}^{p_{\max}}\sum_{p,r=0}^{p_{\max}}(\cdot)$ into two: $\sum_{k=0}^{p_{\max}}\sum_{p=0}^{p_{\max}}\sum_{r=p_\varepsilon(n)}^{p_{\max}}(\cdot) = O_P(p^2C_n)$, $\sum_{k=0}^{p_{\max}}\sum_{r=0}^{p_{\max}}\sum_{p=p_\varepsilon(n)}^{p_{\max}}(\cdot) = O_P(p^2C_n)$ (using the Assumption C), and the sum $\sum_{r=0}^{p_{\max}}\sum_{p=p_\varepsilon(n)}^{p_{\max}}\sum_{r=p_\varepsilon(n)}^{p_{\max}}(\cdot) = O_P(n^{-1}(n/\log n)^{1/(2r+2)})$. Therefore, $S_1 = O_P((n/\log n)^{-(k-2)/(2r+2)})$.

Finally, we have

$$p_{\max}^{1/2}\|\widehat{\boldsymbol{\gamma}}_{p_{\max}}^* - \boldsymbol{\gamma}_{p_{\max}}\|_2 = O_P\left((n/\log n)^{\max\{-(k-3),-(2k-4),-(2r-4)\}/(4r+4)}\right),$$

and the Assumption $A2$ with $r > 2$ and Assumption C with $k > 3$ concludes the proof.

Proof of Lemma 3. We have that

$$\widehat{\sigma}_{p,n}^{2\,*} = n^{-1}\sum_{t=p_{\max}}^{n-1}\left(X_{t+1}^* + \phi_p'\mathbf{X}_{t,p}^*\right)^2 + 2n^{-1}\sum_{t=p_{\max}}^{n-1}\left(X_{t+1}^* + \phi_p'\mathbf{X}_{t,p}^*\right)\left(\widehat{\phi}_p^* - \phi_p\right)'\mathbf{X}_{t,p}^*$$
$$+n^{-1}\sum_{t=p_{\max}}^{n-1}\left(\left(\widehat{\phi}_p^* - \phi_p\right)'\mathbf{X}_{t,p}^*\right)^2 = S_1 + S_2 + S_3, \quad (27)$$

where $\mathbf{X}_{t,c}^* = (X_t^*, \ldots, X_{t+1-p}^*)'$. To establish (16), we prove 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 2, we obtain that $S_3 = O_P((n/\log n)^{-(r-2)/(2r+2)})$. To prove that S_1 goes to $\sigma_{p,n}$, we have $\mathrm{E}^*[S_1] - \sigma_{p,n}^2 = O_P((n/\log n)^{-(r-1)/(2r+2)})$ and $\mathrm{Var}^*[S_1] = O_P(n^{-1}(n/\log n)^{1/(2r+2)})$, using similar arguments as in (26). In Alonso, Peña and Romo (2003) we give the details.

Finally, by the Cauchy-Schwarz inequality, we have $S_2 = O_P(n^{-1/2}(n/\log n)^{-(r-3)/(4r+4)})$. Then

$$|\widehat{\sigma}_{p,n}^{2\,*} - \sigma_{p,n}^2| = O_P\left((n/\log n)^{\max\{-(k-3),-(2k-4),-(2r-4)\}/(4r+4)}\right),$$

and the Assumption $A2$ with $r > 2$ and Assumption C with $k > 3$ concludes the proof.

Proof of Proposition 3. Suppose that there exists a positive integer C such that $\lim_{n\to\infty} \mathrm{pr}^*\{\widehat{p}^*(n) \leq C\} > 0$. This is equivalent to

$$\lim_{n\to\infty} \mathrm{pr}^* \left\{\exists p' = p'(n) \leq C : S_n^*(p') \leq S_n^*(p)\right\} > 0. \tag{28}$$

From Lemma 2 we have that $\widehat{p}$ is a divergent sequence, i.e., for any $0 < C < +\infty$ we have that $\mathrm{pr}\{\widehat{p} > C\} \to 1$. Then, (28) and (16) from Lemma 3 implies that for all $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathrm{pr}^* \left\{\exists p' \leq C : -\varepsilon < \frac{n+2\widehat{p}}{n}\sigma^2_{\widehat{p},n} - \frac{n+2p'}{n}\sigma^2_{p',n}\right\} > 0. \tag{29}$$

By assumption $A3$, we have $\sigma_C^2 > \sigma^2$, and note that $\sigma^2_{\widehat{p},n} \xrightarrow{P} \sigma^2$ and $\sigma_C^2 \leq \liminf \sigma^2_{p',n} \leq \limsup \sigma^2_{p',n} \leq \sigma_1^2$. Choose a sufficiently small ε in order to get a contradiction with (29).

Proof of Theorem 1. We can write X_{T+h} and X^*_{T+h} as $X_{T+h} = -\sum_{j=1}^{+\infty} \phi_j X_{T+h-j} + \varepsilon_{T+h}$, and $X^*_{T+h} = -\sum_{j=1}^{+\infty} \widehat{\phi}^*_{j,n} X^*_{T+h-j} + \varepsilon^*_{T+h}$, 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, it only remains to prove that the difference of the first terms in X^*_{T+1} and X_{T+1} goes to 0 in probability. This is

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

By similar arguments as in Proposition 2, we have that

$$S_1 = O_P((n/\log n)^{\max\{-(k-1),-(2k-2),-(2r-2)\}/(4r+4)})$$

and

$$S_2 = o_P((n/\log n)^{\max\{-(k-1),-r\}/(2r+2)}).$$

Then

$$\begin{aligned} &-\sum_{j=1}^{+\infty} \widehat{\phi}_{j,n} X_{T+h-j} \\ &= \sum_{j=1}^{+\infty} \phi_j X_{T+1-j} + O_P\left((n/\log n)^{\max\{-(k-1),-(2k-2),-(2r-2)\}/(4r+4)}\right). \end{aligned}$$

Using Assumption $A1$ with $r > 1$ and Assumption C with $k > 1$, we obtain that $X^*_{T+1} \xrightarrow{d^*} X_{T+1}$ in probability. In Alonso, Peña and Romo (2003) we present the details.

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 linear combinations of the corresponding (and independent) errors $(\varepsilon_{T+1}, \ldots, \varepsilon_{T+h}, \varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+h})$.

References

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

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

Alonso, A. M., Peña, D. and Romo, J. (2002). Forecasting time series with sieve bootstrap. *J. Statist. Plann. Inference* **100**, 1-11.

Alonso, A. M., Peña, D. and Romo, J. (2003). On introducing model uncertainty in time series bootstrap. Technical Report, Universidad Autónoma de Madrid, Madrid. http://www.uam.es/andres.alonso/Papeles/ss2.pdf

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

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

Box, G. E. P., Jenkins, G. M. and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall, San Francisco.

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

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

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. Simulation Comput.* **26**, 961-978.

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

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

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

Hurvich, C. M. and Tsai, C. L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika* **78**, 499-509.

Inoue, A. and Kilian, L. (2002). Bootstrapping smooth functions of slope parameters and innovations variance in VAR(∞) models. *Internat. Econom. Rev.* **43**, 309-331.

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

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

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Amer. Statist. Assoc.* **90**, 928-934.

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

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

Kreiss, J.-P. (1992). Bootstrap procedures for AR(∞) processes. In *Bootstrapping and Related Techniques* (Edited by K.-H. Jöckel, G. Rothe and W. Sendler), 107-113. Springer-Verlag, Berlin.

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

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

Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *J. Multivariate Anal.* **57**, 277-296.

Pascual, L., Romo, J. and Ruiz, E. (2001). Effects of parameter estimation on prediction densities: a bootstrap approach. *Internat. J. Forecasting* **17**, 83-103.

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

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.

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.

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

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

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

Zhang, P. (1993). On the convergence rate of model selection criteria. *Comm. Statist. Theory Methods* **22**, 2765-2775.

Departamento de Matemáticas, Universidad Autonóma de Madrid, Madrid 28049, Spain.

E-mail: andres.alonso@uam.es

Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126, Getafe 28903, Spain.

E-mail: dpena@est-econ.uc3m.es

Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126, Getafe 28903, Spain.

E-mail: romo@est-econ.uc3m.es

(Received December 2002; accepted May 2003)