On sieve bootstrap prediction intervals

Andrés M. Alonso[a,b,*], Daniel Peña[b], Juan Romo[b]

[a] *Department of Mathematics, Universidad Autónoma de Madrid, Madrid 28049, Spain*
[b] *Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Madrid, Spain*

Received December 2001; received in revised form May 2003

Abstract

In this paper we consider a sieve bootstrap method for constructing nonparametric prediction intervals for a general class of linear processes. We show that the sieve bootstrap provides consistent estimators of the conditional distribution of future values given the observed data.

MSC: 62M10; 62M20; 62G09

Keywords: Sieve bootstrap; Prediction intervals; Linear processes

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 addition, some bootstrap approaches have been proposed in order to avoid the use of a specified errors distribution, see e.g. Stine (1987) and Thombs and Schucany (1990) for AR(p) models, and Pascual et al. (1998) for ARMA(p,q) models.

In Alonso et al. (2002) we propose an AR(∞)-sieve bootstrap procedure to construct prediction intervals for a general class of linear models that includes stationary and invertible ARMA processes. We illustrate with an extensive Monte Carlo study showing that sieve bootstrap prediction intervals provide better coverage results than some previous methods in general cases (see also Alonso et al.

* Corresponding author. Department of Mathematics, Universidad Autónoma de Madrid, Madrid 28049, Spain.
E-mail address: andres.alonso@uam.es (A.M. Alonso).

(2000) for additional simulations and real data examples). Our approach uses the sieve bootstrap of Kreiss (1992) and Bühlmann (1997) based on residual resampling from a sequence of approximating autoregressions for $\{X_t\}_{t\in Z}$ with order $p = p(n)$ that increases as a function of the sample size n. The proposed bootstrap prediction intervals could be applied to a more general class of linear models without specifying a finite dimensional model as have done previous bootstrap proposals.

The paper is organized as follows. Section 2 presents the sieve bootstrap for estimating forecast intervals. Section 3 establishes that the sieve bootstrap provides consistent estimators of the conditional distribution of future values X_{T+h}.

2. Sieve bootstrap forecast intervals

Let $\{X_t\}_{t\in Z}$ be a real valued stationary process with expectation $E[X_t]=\mu_X$, that admits an AR(∞) representation. The method proceeds as follows:

1. Given a sample $\{X_1,\ldots,X_n\}$, select the order $p = p(n)$ of the autoregressive approximation by AICC criterion (cf. Section 9.3 of Brockwell and Davis, 1991).
2. Obtain the Yule–Walker estimators of the autoregressive coefficients: $(\hat{\phi}_1,\hat{\phi}_2,\ldots,\hat{\phi}_p)^{\mathrm{t}}$.
3. Compute the residuals: $\hat{\varepsilon}_t = \sum_{j=0}^{p} \hat{\phi}_j(X_{t-j} - \bar{X})$, where $\hat{\phi}_0 = 1$, and $t \in (p+1,\ldots,n)$.
4. Compute the empirical distribution function of the centered residuals: $\hat{F}_{\tilde{\varepsilon}}(x) = (n-p)^{-1} \sum_{t=p+1}^{n} 1_{\{\tilde{\varepsilon}_t \leqslant x\}}$, where $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\varepsilon}^{(\cdot)}$ and $\hat{\varepsilon}^{(\cdot)} = (n-p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_t$.
5. Draw a resample ε_t^* of i.i.d. observations from $\hat{F}_{\tilde{\varepsilon}}$.
6. Define X_t^* by the recursion:

$$\sum_{j=0}^{p} \hat{\phi}_j(X_{t-j}^* - \bar{X}) = \varepsilon_t^*, \tag{2.1}$$

where the starting p observations are equals 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. However, it is not effective for bootstrap prediction, because it does not replicate the conditional distribution of X_{T+h} given the observed data. But, if we proceed by fixing the last p observations as done in Cao et al. (1997) we can obtain resamples of the future values X_{T+h}^* given $X_{T-p+1}^* = X_{T-p+1},\ldots,X_T^* = X_T$.

7. Compute the estimation of $(\hat{\phi}_1^*,\ldots,\hat{\phi}_p^*)^{\mathrm{t}}$, as in Step 2.
8. Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{p} \hat{\phi}_j^*(X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*, \tag{2.2}$$

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

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

$$[Q^*(\alpha/2), Q^*(1-\alpha/2)], \tag{2.3}$$

where $Q^*(\cdot)$ are the quantiles of the estimated bootstrap distribution.

3. Asymptotic results

The asymptotic validity of the proposed intervals (2.3) depends on the limiting behavior of the distribution $F^*_{X^*_{T+h}}$, and it is sufficient to establish convergence in the conditional distribution of the bootstrap version X^*_{T+h} to X_{T+h}. Notice that the proposed bootstrap procedure has two main parts: (i) obtaining the estimates $\hat{\phi}^*_p$ in order to have information about the distribution of $\hat{\phi}_p$, and (ii) computing the future values X^*_{T+h}. First, in Proposition 1, we prove the convergence in probability of $\hat{\phi}^*_p$ to $\hat{\phi}_p$, and in Theorem 1 we prove the large-sample validity of a conditional sieve approach.

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

Assumption A1. $X_t - \mu_X = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ $(t \in \mathbb{Z})$ with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ i.i.d. and $E[\varepsilon_t|\mathscr{F}_{t-1}] \equiv 0$, $E[\varepsilon_t^2|\mathscr{F}_{t-1}] \equiv \sigma^2 < \infty$, $E[|\varepsilon_t|^s] < \infty$ for some $s \geqslant 4$, and $\mathscr{F}_{t-1}$ is the σ-field generated by $\{\varepsilon_s\}_{s=-\infty}^{t-1}$.

Assumption A2. $\Psi(z)$ is bounded away from zero for $|z| \leqslant 1$, and $\sum_{j=0}^{+\infty} j^r|\psi_j| < \infty$ for some $r \in \mathbb{N}$.

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

Assumption B. $p = p(n) \to \infty$, $p(n) = \mathrm{o}(n)$ $(n \to \infty)$, and the $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots \hat{\phi}_{p,n})^{\mathrm{t}}$ satisfy the empirical Yule–Walker equations

$$\hat{\Gamma}_p \hat{\phi}_p = -\hat{\gamma}_p, \tag{3.1}$$

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

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

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

$$\max_{1 \leqslant j \leqslant p(n)} |\hat{\phi}^*_j - \hat{\phi}_j| \xrightarrow{\mathrm{P}^*} 0, \quad \textit{in probability.} \tag{3.2}$$

Proof. The vector $\hat{\phi}^*_p$ is defined by the bootstrap empirical Yule–Walker equations

$$\hat{\Gamma}^*_p \hat{\phi}^*_p = -\hat{\gamma}^*_p, \tag{3.3}$$

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

$$\begin{aligned} \|\hat{\phi}^*_p - \hat{\phi}_p\|_\infty &= \|(\hat{\Gamma}_p^{-1} - \hat{\Gamma}_p^{*-1})\hat{\gamma}^*_p + \hat{\Gamma}_p^{-1}(\hat{\gamma}_p - \hat{\gamma}^*_p)\|_\infty \\ &\leqslant \|\hat{\Gamma}_p^{*-1} - \hat{\Gamma}_p^{-1}\|_{\mathrm{row}} \|\hat{\gamma}^*_p\|_\infty + \|\hat{\Gamma}_p^{-1}\|_{\mathrm{row}} \|\hat{\gamma}^*_p - \hat{\gamma}_p\|_\infty, \end{aligned} \tag{3.4}$$

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

From the proof of Theorem 2.1 of Hannan and Kavalieris (1986), we have that $\|\hat{\Gamma}_p\|_{\text{row}}$ and $\|\hat{\Gamma}_p^{-1}\|_{\text{row}}$ are uniformly bounded in p. Since $\hat{\Gamma}_p^{-1} - \hat{\Gamma}_p^{*-1} = \hat{\Gamma}_p^{-1}(\hat{\Gamma}_p^* - \hat{\Gamma}_p)\hat{\Gamma}_p^{*-1}$, and $\|\hat{\Gamma}_p^* - \hat{\Gamma}_p\|_{\text{row}} \leqslant |\hat{\gamma}_0^* - \hat{\gamma}_0| + 2\|\hat{\gamma}_p^* - \hat{\gamma}_p\|_1$, we can concentrate our attention on this last term

$$\|\hat{\gamma}_p^* - \hat{\gamma}_p\|_1 \leqslant \|\hat{\gamma}_p^* - \gamma_p\|_1 + \|\gamma_p - \hat{\gamma}_p\|_1. \tag{3.5}$$

From Theorem 3 of An et al. (1982), the second term is $\mathrm{O}_{\text{a.s.}}((n/\log(n))^{-r/(2r+2)})$.

Since

$$\|\hat{\gamma}_p^* - \gamma_p\|_\infty \leqslant \|\hat{\gamma}_p^* - \gamma_p\|_1 \leqslant p^{1/2}\|\hat{\gamma}_p^* - \gamma_p\|_2, \tag{3.6}$$

to get convergence to zero in (3.4), it is enough to consider the last term in (3.6).

$$\begin{aligned}\|\hat{\gamma}_p^* - \gamma_p\|_2^2 &= \sum_{k=1}^{p} (\hat{R}^*(k) - R(k))^2 \\ &\leqslant 2\sum_{k=1}^{p} (\hat{R}^*(k) - E^*[\hat{R}^*(k)])^2 + 2\sum_{k=1}^{p} (E^*[\hat{R}^*(k)] - R(k))^2 \\ &= 2(S_1 + S_2).\end{aligned} \tag{3.7}$$

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

$$S_2 = \sum_{k=1}^{p} \left(E^*[\varepsilon_1^{*2}] \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n}\delta_{i+k,j} - E[\varepsilon_1^2] \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \psi_i\psi_j\delta_{i+k,j} \right)^2, \tag{3.8}$$

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

$$\begin{aligned} S_2 = \sum_{k=1}^{p} \Bigg(& E^*[\varepsilon_1^{*2}] \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} (\hat{\psi}_{i,n}\hat{\psi}_{j,n} - \psi_i\psi_j)\delta_{i+k,j} \\ & + (E^*[\varepsilon_1^{*2}] - E[\varepsilon_1^2]) \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \psi_i\psi_j\delta_{i+k,j} \Bigg)^2 \\ \leqslant 2\sum_{k=1}^{p} & \left(E^*[\varepsilon_1^{*2}] \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} (\hat{\psi}_{i,n}\hat{\psi}_{j,n} - \psi_i\psi_j)\delta_{i+k,j} \right)^2 \\ & + 2\sum_{k=1}^{p} \left((E^*[\varepsilon_1^{*2}] - E[\varepsilon_1^2]) \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \psi_i\psi_j\delta_{i+k,j} \right)^2 = I_1 + I_2. \end{aligned} \tag{3.9}$$

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

$$\sup_{i\in\mathbb{N}} |\hat{\psi}_{i,n} - \psi_i| = \mathrm{O}_{\text{a.s.}}((\log(n)/n)^{1/2}) + \mathrm{O}_{\text{a.s.}}(p^{-r}) \tag{3.10}$$

and

$$\sup_{n \geqslant n_1} \sum_{i=0}^{+\infty} i^r |\hat{\psi}_{i,n}| = \mathrm{O}_{\mathrm{a.s.}}(1), \tag{3.11}$$

where n_1 is a random variable.

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

$$\begin{aligned} I_1 &\leqslant 2E^*[\varepsilon_1^{*2}]^2 p \left(\sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} |\hat{\psi}_{i,n}\hat{\psi}_{j,n} - \psi_i\psi_j| \right)^2 \\ &\leqslant 2E^*[\varepsilon_1^{*2}]^2 p \left(\sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} |\hat{\psi}_{i,n}\psi_j - \psi_i\psi_j| + \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} |\hat{\psi}_{i,n}\hat{\psi}_{j,n} - \hat{\psi}_{i,n}\psi_j| \right)^2 \\ &= \mathrm{O}_{\mathrm{P}}(p)(\mathrm{O}_{\mathrm{a.s.}}((\log(n)/n)^{1/2} p) + \mathrm{O}_{\mathrm{a.s.}}(p^{-r+1}) + \mathrm{o}_{\mathrm{a.s.}}(p^{-r}))^2 \\ &= \mathrm{O}_{\mathrm{P}}((n/\log(n))^{-(2r-3)/(2r+2)}). \end{aligned} \tag{3.12}$$

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

$$E^*[\varepsilon_t^{*2}] - E[\varepsilon_t^2] = \mathrm{o}_{\mathrm{P}}((\log(n)/n)^{1/2} p). \tag{3.13}$$

Therefore,

$$I_2 = \mathrm{o}_{\mathrm{P}}((\log(n)/n) p^3) = \mathrm{o}_{\mathrm{P}}((n/\log(n))^{-(2r-1)/(2r+2)}). \tag{3.14}$$

For the other term in (3.7), we have $S_1 = \mathrm{O}_{\mathrm{P}}(n^{-1}(n/\log(n))^{1/(2r+2)})$, since

$$\begin{aligned} S_1 &= \sum_{k=1}^{p} \left(n^{-1} \sum_{t=1}^{n-k} \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n}\varepsilon_{t-i}^*\varepsilon_{t+k-j}^* - \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n} E^*[\varepsilon_1^{*2}]\delta_{i+k,j} \right)^2 \\ &= \sum_{k=1}^{p} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j=0}^{+\infty} \sum_{h,l=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n}\hat{\psi}_{h,n}\hat{\psi}_{l,n} (\varepsilon_{t-i}^*\varepsilon_{t+k-j}^* - E^*[\varepsilon_1^{*2}]\delta_{i+k,j}) \\ &\quad \times (\varepsilon_{s-h}^*\varepsilon_{s+k-l}^* - E^*[\varepsilon_1^{*2}]\delta_{h+k,l}). \end{aligned} \tag{3.15}$$

Taking E^* in the above expression, we have

$$\begin{aligned} E^*[S_1] = \sum_{k=1}^{p} n^{-2} \sum_{t,s=1}^{n-k} \sum_{i,j=0}^{+\infty} \sum_{h,l=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n}\hat{\psi}_{h,n}\hat{\psi}_{l,n} (E^*[\varepsilon_{t-i}^*\varepsilon_{t+k-j}^*\varepsilon_{s-h}^*\varepsilon_{s+k-l}^*] \\ - E^*[\varepsilon_1^{*2}]^2 \delta_{i+k,j}\delta_{h+k,l}). \end{aligned} \tag{3.16}$$

Notice that

$$E^*[\varepsilon^*_{t-i}\varepsilon^*_{t+k-j}\varepsilon^*_{s-h}\varepsilon^*_{s+k-l}] = \begin{cases} E^*[\varepsilon_1^{*4}] & \text{if } t-i=t+k-j=s-h=s+k-l, \\ E^*[\varepsilon_1^{*2}]^2 & \text{if two pairs different indices,} \\ 0 & \text{otherwise} \end{cases} . \tag{3.17}$$

and

$$\begin{aligned} &E^*[\varepsilon^*_{t-i}\varepsilon^*_{t+k-j}\varepsilon^*_{s-h}\varepsilon^*_{s+k-l}] - E^*[\varepsilon_1^{*2}]^2\delta_{i+k,j}\delta_{h+k,l} \\ &= \begin{cases} E^*[\varepsilon_1^{*4}] - E^*[\varepsilon_1^{*2}]^2 & \text{if } t-i=t+k-j=s-h=s+k-l, \\ 0 & \text{if } t-i=t+k-j\neq s-h=s+k-l, \\ E^*[\varepsilon_1^{*2}]^2 & \text{if } t-i=s-h\neq t+k-j=s+k-l, \\ & \text{or } t-i=s+k-l\neq s-h=t+k-j, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.18}$$

Because of Theorem 3.2 of Bühlmann (1995), we have for some random variable n_1 that $\sup_{n\geqslant n_1} \sum_{i,j,h,l=0}^{+\infty} \hat{\psi}_{i,n}\hat{\psi}_{j,n}\hat{\psi}_{h,n}\hat{\psi}_{l,n} = \mathrm{O}_{\mathrm{a.s.}}(1)$. On the other hand, in (3.16) 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, $E^*[S_1] = \mathrm{O}_{\mathrm{P}}(pn^{-1})$.

Finally, we have

$$p^{1/2}\|\hat{\gamma}^*_p - \gamma_p\|_2 = \mathrm{O}_{\mathrm{P}}((n/\log(n))^{-(r-2)/(2r+2)}) \tag{3.19}$$

and the Assumption A2 with $r > 2$ concludes the proof. □

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

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

$$X^*_{T+h} \xrightarrow{\mathrm{d}} X_{T+h}, \quad \textit{in probability.} \tag{3.20}$$

Proof. We can write X_{T+h} and X^*_{T+h} as

$$X_{T+h} = -\sum_{j=1}^{+\infty}\phi_j X_{T+h-j} + \varepsilon_{T+h}, \tag{3.21}$$

$$X^*_{T+h} = -\sum_{j=1}^{+\infty}\hat{\phi}_{j,n} X^*_{T+h-j} + \varepsilon^*_{T+h}, \tag{3.22}$$

where $\hat{\phi}_{j,n}$ denote the estimates of ϕ_j with a sample of size n: $(X_{T-n+1},\ldots,X_T)$, $\hat{\phi}_{j,n}=0$ for $j > p(n)$, and $X^*_t = X_t$ for $t \leqslant T$. For simplicity of notation we prove the theorem for $h=1$:

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

$$-\sum_{j=1}^{+\infty}(\hat{\phi}_{j,n}-\phi_j)X_{T+1-j} = -\sum_{j=1}^{p(n)}(\hat{\phi}_{j,n}-\phi_j)X_{T+1-j} + \sum_{j=p(n)+1}^{+\infty}\phi_j X_{T+1-j}$$
$$= S_{1,1} + S_{2,1}. \tag{3.23}$$

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

$$E[|S_{2,1}|] \leqslant E[|X_t|] \sum_{j=p(n)+1}^{+\infty}|\phi_j| = \mathrm{o}(\mathrm{p}^{-r}) \tag{3.24}$$

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

$$|S_{1,1}| \leqslant \left|\sum_{j=1}^{p(n)}(\hat{\phi}_{j,n}-\phi_{j,n})X_{T+1-j}\right| + \left|\sum_{j=1}^{p(n)}(\phi_{j,n}-\phi_j)X_{T+1-j}\right| = I_1 + I_2, \tag{3.25}$$

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

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

$$\max_{1\leqslant j\leqslant p}|\hat{\phi}_{j,n}-\phi_{j,n}| = \mathrm{O_{a.s.}}((\log(n)/n)^{1/2}). \tag{3.26}$$

Therefore,

$$I_1 \leqslant \left(\sum_{j=1}^{p(n)}(\hat{\phi}_{j,n}-\phi_{j,n})^2\right)^{1/2}\left(\sum_{j=1}^{p(n)}X^2_{T+1-j}\right)^{1/2}$$
$$\leqslant p(n)^{1/2}\max_{1\leqslant j\leqslant p}|\hat{\phi}_{j,n}-\phi_{j,n}|\mathrm{O_P}(p(n)^{1/2}) = \mathrm{O_P}(p(n)(\log(n)/n)^{1/2})$$
$$= \mathrm{O_P}((n/\log(n))^{-r/(2r+2)})). \tag{3.27}$$

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

$$\sum_{j=0}^{p(n)}|\phi_{j,n}-\phi_j| \leqslant c\sum_{j=p(n)+1}^{+\infty}|\phi_j|, \tag{3.28}$$

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

$$E[I_2] \leqslant E[|X_t|]\sum_{j=1}^{p(n)}|\phi_{j,n}-\phi_j| = \mathrm{o}(p^{-r}). \tag{3.29}$$

Finally,

$$-\sum_{j=1}^{+\infty}\hat{\phi}_{j,n}X_{T+h-j}=\sum_{j=1}^{+\infty}\phi_j X_{T+1-j}+\mathrm{O_P}((n/\log(n))^{-r/(2r+2)})). \tag{3.30}$$

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

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

Acknowledgements

We would like to thank Mike Wiper for his careful reading which greatly improved the paper. This research was partially supported by the CYCIT project BEC 2000-0167 and by the Cátedra de Calidad BBVA.

References

Alonso, A.M., Peña, D., Romo, J., 2000. Forecasting time series with sieve bootstrap. W.P. 00-07, Universidad Carlos III de Madrid, Madrid.

Alonso, A.M., Peña, D., Romo, J., 2002. Forecasting time series with sieve bootstrap. J. Statist. Planning Inference 100, 1–11.

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

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

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer, New York.

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

Bühlmann, P., 1997. Sieve bootstrap for time series. Bernoulli 3, 123–148.

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

Hannan, E.J., Deistler, M., 1988. The Statistical Theory of Linear Systems. Wiley, New York.

Hannan, E.J., Kavalieris, L., 1986. Regression, autoregression models. J. Time Ser. Anal. 7, 27–49.

Kreiss, J.-P., 1992. Bootstrap procedures for AR(∞)-processes. In: Jöckel, K.H., Rothe, G., Sendler, W. (Eds.), Bootstrapping and Related Techniques. Springer, Heidelberg, pp. 107–113.

Pascual, L., Romo, J., Ruiz, E., 1998. Bootstrap predictive inference for ARIMA processes. W.P. 98-86, Universidad Carlos III de Madrid, Madrid.

Stine, R.A., 1987. Estimating properties of autoregressive forecasts. J. Amer. Statist. Assoc. 82, 1072–1078.

Thombs, L.A., Schucany, W.R., 1990. Bootstrap prediction intervals for autoregression. J. Amer. Statist. Assoc. 85, 486–492.