

ON THE CONSISTENCY OF SIEVE BOOTSTRAP PREDICTION INTERVALS FOR STATIONARY TIME SERIES

ROMAN RÓŻAŃSKI AND ADAM ZAGDAŃSKI

Institute of Mathematics
Wrocław University of Technology

e-mail: rozanski@im.pwr.wroc.pl

e-mail: zagdan@im.pwr.wroc.pl

Abstract

In the article, we consider construction of prediction intervals for stationary time series using Bühlmann's [8], [9] sieve bootstrap approach. Basic theoretical properties concerning consistency are proved. We extend the results obtained earlier by Stine [21], Masarotto and Grigoletto [13] for an autoregressive time series of finite order to the rich class of linear and invertible stationary models. Finite sample performance of the constructed intervals is investigated by computer simulations.

Keywords: prediction intervals, sieve bootstrap, method of sieves.

2000 Mathematics Subject Classification: 62G09, 62G15, 62M20.

1. INTRODUCTION

Computing interval forecasts are an important part of the forecasting process. Namely, we construct prediction intervals sometimes called prediction bounds or forecast limits in order to indicate the likely uncertainty in point forecasts.

The most common approach to constructing prediction intervals known as the Box-Jenkins method is based on the Gaussian approximation of the prediction error distribution. However, we cannot expect Gaussian prediction intervals to perform very well for non-Gaussian series.

Thus, many authors consider more general bootstrap-based procedures of construction prediction intervals. Let us now briefly review the state of the art.

Stine [21] considered the construction of bootstrap prediction intervals for autoregressive processes of known order ($AR(p)$). Assuming that the error distribution is symmetric and with finite moments the consistency of constructed prediction intervals was proved.

Thombs and Schucany [22] used the bootstrap method to construct conditional prediction intervals for autoregressive models of known order ($AR(p)$). The main idea was approximation of the unknown conditional distribution of future value X_{n+h} given observations. To generate bootstrap replicates Thombs and Schucany used both a forward and backward representation of an autoregressive model.

Cao [11] proposed a modification of Thombs-Schucany's procedure which improves computational efficiency. In this case, the resampling mechanism is not used to draw bootstrap replicates of the original series but only replicates of some future values of the series. Moreover, Cao [11] obtained promising simulation results on the application of the smoothed bootstrap, i.e., replacing empirical distribution function by a kernel estimator for drawing resampled bootstrap errors.

Masarotto [17] used a bootstrap method to construct prediction intervals for autoregressive processes $AR(p)$ of finite but unknown order p using an optimal linear predictor.

Grigoletto [13] considered a modification of Masarotto's algorithm, which allows a substantial variance reduction in estimation of the percentiles of prediction error distribution.

Kim [16] constructed prediction intervals for vector autoregressive models of known order ($VAR(p)$) using the bootstrap-after-bootstrap approach which has a built-in bias-correction procedure.

Approaches discussed above to constructing bootstrap prediction intervals are „model-based”, i.e., rely on the finite dimensional parametric model assumption. In order to generalize this construction to a broader class of time series a nonparametric resampling scheme should be used.

Bühlmann [8], [9] has proposed a resampling procedure called a sieve bootstrap which has the advantage that no particular finite parametric model for data is assumed. This approach is based on Grenander's [14]

method of sieves whose main idea is approximating an infinite-dimensional, nonparametric model by a sequence of finite-dimensional parametric models. The sieve bootstrap scheme is valid for the rich subclass of linear stationary processes which can be represented as an autoregressive process of order infinity ($AR(\infty)$). Moreover, Bühlmann [8], [10] showed that for many linear processes the stationary sieve bootstrap for $AR(\infty)$ models has generally a better performance than other nonparametric resampling techniques called the blockwise bootstrap.

Recently, Alonso, Peña and Romo [2] considered a generalization of results obtained by Thombs and Schucany [22] and Cao [11] for a general class of linear processes. Nonparametric conditional prediction intervals are constructed using the sieve bootstrap approach. Simulation results indicate that the method has better coverage and mean lengths results and is an alternative to the classical Box-Jenkins Gaussian prediction intervals.

In this article, we use the sieve bootstrap resampling scheme to generalize the construction of unconditional bootstrap prediction intervals (proposed e.g. by Stine [21], Masarotto [17] and Grigoletto [13] for $AR(p)$ processes) to a wider class of linear models represented as $AR(\infty)$ processes.

The article is organized as follows. Section 2 contains model formulation and a detailed description of the sieve bootstrap procedure. In Section 3, we discuss the problem of interval forecasts and construct hybrid sieve bootstrap prediction intervals. Theoretical relating to the concerning consistency of the hybrid bootstrap are given in Section 4. In Section 5, we justify also the consistency of constructed prediction intervals. In Section 6, we briefly present the construction of studentized prediction intervals and announce (without proofs) theoretical results which may be obtained. The last section is devoted to numerical results.

Simulations have been carried out using computers of the Wrocław Centre of Networking and Supercomputing.

2. SIEVE BOOTSTRAP

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a stationary, real valued process with zero expectation. If $\{X_t\}_{t \in \mathbb{Z}}$ is a purely stochastic process, then by Wold's theorem (see e.g. Anderson [3]) $\{X_t\}_{t \in \mathbb{Z}}$ may be represented as a moving average process of order infinity ($MA(\infty)$), i.e.,

$$(2.1) \quad X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1, \quad t \in \mathbb{Z},$$

where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of uncorrelated random variables with $E[\epsilon_t] = 0$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

Moreover, we will require the process X_t to be invertible, and the narrows a bit the class of stationary processes.

Appropriate conditions guaranteeing the invertibility given e.g. in Anderson ([3], Theorem 7.6.9) allow us to represent $\{X_t\}_{t \in \mathbb{Z}}$ as an autoregressive process of order infinity ($AR(\infty)$), i.e.,

$$(2.2) \quad X_t - \sum_{j=1}^{\infty} \phi_j X_{t-j} = \epsilon_t, \quad t \in \mathbb{Z},$$

where $\sum_{j=1}^{\infty} \phi_j^2 < \infty$.

Using the following notations:

$$\Phi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j, \quad z \in C,$$

and

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad \psi_0 = 1, \quad z \in C,$$

one can represent X_t as:

$$AR(\infty) : \quad \Phi(B)(X_t) = \epsilon_t$$

or

$$MA(\infty) : \quad X_t = \Psi(B)\epsilon_t,$$

where B stands for a backward shift operator, i.e., $B X_t = X_{t-1}$.

Let us also denote by $\mathcal{F}_t = \sigma(\{\epsilon_s : s \leq t\})$ σ -algebra generated by $\{\epsilon_s\}_{s=-\infty}^t$.

In the sequel, some of the following assumptions will be imposed:

- (A1') $X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = 1$ ($t \in \mathbb{Z}$) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a stationary and ergodic sequence and $E[\epsilon_t | \mathcal{F}_{t-1}] = 0$, $E[\epsilon_t^2 | \mathcal{F}_{t-1}] \equiv \sigma^2 < \infty$, $E|\epsilon_t|^s < \infty$ for some $s \geq 4$.
- (A1) $X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = 1$ ($t \in \mathbb{Z}$) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence and $E[\epsilon_t] = 0$, $E|\epsilon_t|^s < \infty$ for some $s \geq 4$.

- (A2) $\Psi(z)$ is bounded away from zero for $|z| \leq 1$, $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$ for some natural r .
- (B) $p = p(n) \rightarrow \infty$, $p(n) = o(n)$ ($n \rightarrow \infty$) and $\hat{\phi}_p = (\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n})'$ satisfies empirical Yule-Walker equations, i.e.,:

$$\hat{\Gamma}_p \hat{\phi}_p = \hat{\gamma}_p,$$

where $\hat{\Gamma}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$, $\hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))'$, and $\hat{\gamma}$ is the sample autocovariance function given by:

$$\hat{\gamma}(j) = \frac{1}{n} \sum_{t=1}^{n-|j|} (X_t - \bar{X}) (X_{t+|j|} - \bar{X}), \quad |j| \leq n-1,$$

where $\bar{X} = \sum_{t=1}^n X_t$.

Roughly speaking, the main idea of the sieve bootstrap is approximation of the process X_t by a sequence of autoregressive processes of order $p = p(n)$ growing „sufficiently slow” with sample size n , i.e., $p(n) \rightarrow \infty$ when ($n \rightarrow \infty$) but $p(n) = o(n)$. In other words, we approximate the infinite dimensional non-parametric model (which can be represented as $AR(\infty)$ process) by a sequence of finite dimensional parametric models. This procedure is well known as the method of sieves (Grenander [14]).

Let us now present the detailed algorithm of the sieve bootstrap.

Step 1. Let X_1, \dots, X_n be a realization of the process $\{X_t\}_{t \in \mathbb{Z}}$. Applying Bühlmann’s suggestion [8] we choose the approximating order $p = p(n)$ using the Akaike information criterion (AIC) in a range $[0, p_{max}(n)]$, where $p_{max}(n)$ is increasing with n . In practise, we use $p_{max}(n) = 10 \log_{10}(n)$ (The standard value used in statistical packages).

Step 2. We estimate coefficients of the $AR(p(n))$ model, i.e. $\phi_1, \dots, \phi_{p(n)}$, using observation $\{X_t\}_{t=1}^n$. Estimates $\hat{\phi}_p = (\hat{\phi}_1, \dots, \hat{\phi}_{p(n)})'$ are determined using the Yule-Walker method (Brockwell and Davis, [6], p. 232–233):

$$\hat{\Gamma}_p \hat{\phi}_p = \hat{\gamma}_p,$$

where $\hat{\Gamma}_p$ and $\hat{\gamma}_p$ are given as in assumption B.

Taking into account the computational efficiency, we can determine Yule-Walker estimates more effectively using the recursive Durbin-Levinson algorithm:

$$\begin{aligned}\hat{\phi}_{11} &= \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}, \quad \hat{\nu}_1 = \hat{\gamma}(0) \left[1 - \hat{\phi}_{11}^2 \right], \\ \hat{\phi}_{mm} &= \left(\hat{\gamma}(m) - \sum_{j=1}^{m-1} \hat{\phi}_{m-1,j} \hat{\gamma}(m-j) \right) / \hat{\nu}_{m-1}, \\ \begin{bmatrix} \hat{\phi}_{m,1} \\ \vdots \\ \hat{\phi}_{m,m-1} \end{bmatrix} &= \begin{bmatrix} \hat{\phi}_{m-1,1} \\ \vdots \\ \hat{\phi}_{m-1,m-1} \end{bmatrix} - \hat{\phi}_{mm} \begin{bmatrix} \hat{\phi}_{m-1,m-1} \\ \vdots \\ \hat{\phi}_{m-1,1} \end{bmatrix}, \\ \hat{\nu}_m &= \hat{\nu}_{m-1} \left(1 - \hat{\phi}_{mm}^2 \right).\end{aligned}$$

Step 3. We compute residuals

$$\hat{\varepsilon}_{t,n} = X_t - \sum_{j=1}^{p(n)} \hat{\phi}_{j,n} X_{t-j}, \quad t = p+1, \dots, n.$$

Step 4. We can construct the replication of observations. For this purpose, the residuals are centered:

$$\tilde{\varepsilon}_{t,n} = \hat{\varepsilon}_{t,n} - \frac{1}{n-p} \sum_{t=p+1}^n \hat{\varepsilon}_{t,n}, \quad t = p+1, \dots, n,$$

and next we draw residuals ε_t^* from the empirical cumulative distribution based on $\{\tilde{\varepsilon}_{t,n}\}_{t=p+1}^n$, i.e. ε_t^* i.i.d. $\sim \hat{F}_{\varepsilon,n}$, where

$$\hat{F}_{\varepsilon,n}(u) = \frac{1}{n-p} \sum_{t=p+1}^n 1_{[\tilde{\varepsilon}_{t,n} \leq u]}.$$

Step 5. Finally, we define a bootstrap replication $\{X_1^*, \dots, X_n^*\}$ by the recursive equation:

$$(2.3) \quad X_t^* = \sum_{j=1}^{p(n)} \hat{\phi}_{j,n} X_{t-j}^* + \varepsilon_t^*.$$

In practice, we can generate replication $\{X_t^*\}$ starting the recursion from some initial values, e.g. equal resampled innovations ε_t^* .

Remark 2.1. In the sieve bootstrap, the crucial assumption is that observations X_1, \dots, X_n are finite realizations of process $AR(\infty)$ as in (2.2).

It is worth pointing out that this representation contains an important class of $ARMA(p, q)$ models:

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \sum_{k=1}^q \psi_k \epsilon_{t-k} + \epsilon_t, \quad t \in Z,$$

with an invertible moving average polynomial, i.e., $\Psi(z) = \sum_{k=0}^q \psi_k z^k$, $z \in C$, has all roots outside the unit disk $\{z \in C, |z| \leq 1\}$.

Remark 2.2. Assumption A2 is satisfied for models with polynomial decaying coefficients $\{\psi_j\}_{j=0}^\infty$ (or equivalently $\{\phi_j\}_{j=0}^\infty$). For example, $ARMA(p, q)$ models satisfy this assumption with the exponentially decaying coefficient $\{\psi_j\}$. Additionally, assumption A2 implies that $\Phi(z)$ is bounded away from zero for $|z| \leq 1$ and $\sum_{j=0}^\infty j^r |\phi_j| < \infty$.

3. PREDICTION INTERVALS

It is well known (see e.g. Brockwell and Davis [6], p. 159–162) that for a stationary process with mean 0, the best (in mean square sense) linear combination of X_1, \dots, X_n for predicting X_{n+h} ($h \geq 1$), is the projection of X_{n+h} onto the closed linear subspace $\overline{\text{span}}\{X_1, \dots, X_n\}$. Thus, we can represent the optimal h -step predictor as:

$$P_n X_{n+h} = P_{\overline{\text{span}}\{X_1, \dots, X_n\}} X_{n+h}.$$

The above predictor may be derived from appropriate projection equations or alternatively using some recursive algorithm, for instance the innovations algorithm (Brockwell and Davis [6], p. 167–168).

Besides calculating optimal predictors we have to assess their accuracy. A well-known measure of uncertainty of the corresponding forecasts is the prediction mean squared error (PMSE). The other possibility is to construct interval forecasts. Therefore, computing prediction intervals is an important part of the forecasting process intended to indicate the likely uncertainty in point forecasts.

Let us now present the formal definition of a prediction interval.

Definition 3.1 (Prediction interval). A prediction interval $I(h, \underline{X})$ with nominal confidence level $1 - 2\alpha$ is a random interval based on past observations $\underline{X} = (X_1, X_2, \dots, X_n)$ and constructed for future (unknown) observation X_{n+h} , $h \geq 1$:

$$I(h, \underline{X}) = [L(\underline{X}), R(\underline{X})],$$

so that $P(L(\underline{X}) \leq X_{n+h} \leq R(\underline{X})) = 1 - 2\alpha$.

3.1. Gaussian prediction intervals

For a stationary Gaussian process one may construct prediction intervals using the fact (Brockwell and Davis [6], p. 175) that the prediction error $\tilde{\Delta}_n(h) := X_{n+h} - P_n X_{n+h}$ is normally distributed with zero mean and variance $\sigma_n^2(h)$.

The $(1 - 2\alpha)$ Gaussian prediction interval is given by:

$$(3.1) \quad I_G(h) = [P_n X_{n+h} - \Phi_{1-\alpha} \sigma_n(h), P_n X_{n+h} + \Phi_{1-\alpha} \sigma_n(h)],$$

where $\Phi_{1-\alpha}$ denotes the $(1 - \alpha)$ -quantile of the standard normal distribution.

Therefore, Gaussian prediction intervals, known as Box-Jenkins prediction intervals, are constructed assuming that the noise distribution is normal and possible departures from normality may badly influence their accuracy. Moreover, applying the Box-Jenkins approach we do not take into account the variability, which accompanies the estimation of model's parameters. All these reasons may result in unsatisfactory coverage results of constructed Gaussian prediction intervals.

For non-Gaussian time series one may construct prediction intervals with the aid of the bootstrap method.

3.2. Sieve bootstrap prediction intervals

We consider now the application of the sieve bootstrap procedure to constructing unconditional prediction intervals.

The main idea of using the sieve bootstrap to construct interval forecasts is generating replication X_1^*, \dots, X_n^* on the basis of the observed series X_1, \dots, X_n , and then extending this replication to the future time $n + h$. Let us note that replications of future observations X_{n+h}^* can be easily determined applying the autoregressive approximation ($AR(p(n))$), i.e.:

$$(3.2) \quad X_{n+h}^* = \hat{\phi}_1^* X_{n+h-1}^* + \dots + \hat{\phi}_{p(n)}^* X_{n+h-p(n)}^* + \epsilon_{n+h}^*,$$

where $\hat{\phi}_1^*, \dots, \hat{\phi}_{p(n)}^*$ – bootstrap replicates of estimators calculated on the basis of X_1^*, \dots, X_n^* .

Remark 3.1. It is worth pointing out that alternatively one could construct bootstrap replicates of future observations X_{n+h}^* as follows:

$$(3.3) \quad X_{n+h}^* = \hat{\phi}_1 X_{n+h-1}^* + \dots + \hat{\phi}_{p(n)} X_{n+h-p(n)}^* + \epsilon_{n+h}^*$$

However, using this approach (e.g. Cao [11]) we do not incorporate the variability coming from parameter estimation. Moreover, the resampling scheme given by (3.2) seems to be more appropriate when replications of the conditional distribution are constructed (as in Thombs and Schucany [22] or Alonso, Peña and Romo [2]) and hence this procedure may be regarded as more general.

The prediction interval will be constructed approximating the unknown distribution of the prediction error

$$\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$$

by the corresponding bootstrap distribution of

$$\Delta_n^*(h) := X_{n+h}^* - \hat{X}_{n+h}^*,$$

where:

- \hat{X}_{n+h} – the estimate of the optimal linear predictor for X_{n+h} constructed on the basis of observations X_1, \dots, X_n using the autoregressive approximation $AR(p(n))$,
- \hat{X}_{n+h}^* – the optimal linear predictor for X_{n+h}^* based on X_1^*, \dots, X_n^* .

This strategy is known as the hybrid bootstrap (Shao and Tu [20]).

In Section 6, we will also analyze another approach to constructing prediction intervals called the bootstrap-t, which is based on the approximation of distribution of the standardized prediction error, i.e., $T_n(h) = \Delta_n(h)/\hat{\sigma}_n(h)$.

Denoting now by q_α^* i $q_{1-\alpha}^*$ corresponding quantiles of distribution $\Delta_n^*(h)$ we may express hybrid bootstrap prediction intervals in the following form:

$$(3.4) \quad I_B(h) = [\hat{X}_{n+h} + q_\alpha^*, \hat{X}_{n+h} + q_{1-\alpha}^*].$$

In practice, the quantiles q_α^* and $q_{1-\alpha}^*$ are replaced by their Monte Carlo approximation based on B replicates of the series (where B is sufficiently large).

Remark 3.2. Let us note that Masarotto [17] and Grigoletto [13] also considered the construction of prediction intervals based on the bootstrap approximation of prediction errors or alternatively standardized prediction errors. However, validity of their approach was established only for autoregressive time series models of finite and possibly unknown order. Moreover, they did not examine the robustness and the effectiveness of the bootstrap approach when the assumption concerning the finite autoregressive structure for time series is not satisfied.

4. CONSISTENCY OF THE HYBRID BOOTSTRAP

In this section, the consistency of the hybrid bootstrap will be established.

Let us first formulate some auxiliary results which will be needed in the proof of the main theorem.

The basic issue concerning the resampled innovations ϵ_t^* was proved by Bühlmann [8].

Lemma 4.1 (Bühlmann [8], Lemma 5.4). *Let assumptions A1' with $s = 4$, A2 with $r = 1$ and B with $p(n) = o((n/\log(n))^{1/2})$ hold. Then for each $t \in Z$:*

$$\epsilon_t^* \xrightarrow{d^*} \epsilon_t \quad \text{in probability.}$$

Lemma 4.2 and 4.3 will be useful in investigating the asymptotic of the prediction error and the bootstrap prediction error.

Lemma 4.2. *Suppose that assumptions A2 with $r \geq 1$ ($r \in N$) and B with $p(n) = o((n/\log(n))^{1/(2r+2)})$ hold. Then*

$$\sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} = o_P(1).$$

Proof of Lemma 4.2. Let us denote

$$\sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} = S_1 + S_2.$$

It is easily seen that

$$E|S_1| \leq \sqrt{EX_t^2} \sum_{j=p(n)+1}^{\infty} |\phi_j|.$$

By assumption A2 it follows that

$$\sum_{j=p(n)+1}^{\infty} j^r |\phi_j| > \sum_{j=p(n)+1}^{\infty} (p(n))^r |\phi_j|.$$

Hence $E|S_1| = o(p(n)^{-r})$.

Applying Markov's inequality we obtain that $S_1 = o_P(p(n)^{-r})$.

Let us consider now S_2 term. One can write

$$\begin{aligned} |S_2| &\leq \left| \sum_{j=1}^{p(n)} (\phi_j - \phi_{j,n}) X_{n+1-j} \right| + \left| \sum_{j=1}^{p(n)} (\phi_{j,n} - \hat{\phi}_j) X_{n+1-j} \right| \\ &= I_1 + I_2, \end{aligned}$$

where $\underline{\phi}_{p(n)} = (\phi_{1,n}, \dots, \phi_{p(n),n})'$ satisfies theoretical Yule-Walker equations, i.e.,

$$\Gamma_{p(n)} \underline{\phi}_{p(n)} = \underline{\gamma}_{p(n)},$$

where $\Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}$, $\underline{\gamma}_{p(n)} = (\gamma(1), \dots, \gamma(p(n)))'$.

In order to bound I_1 we use extended Baxter's inequality (Deistler and Hannan [12], Theorem 6.6.12, p. 271)

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

where c is constant depending on the theoretical model.

Therefore, we may write:

$$\begin{aligned} E I_1 &= E \left| \sum_{j=1}^{p(n)} (\phi_j - \phi_{j,n}) X_{n+1-j} \right| \\ &\leq \sum_{j=1}^{p(n)} E |X_{n+1-j}| \cdot |\phi_{j,n} - \phi_j| \\ &\leq \sqrt{E X_t^2} \cdot c \cdot \sum_{j=p(n)+1}^{\infty} |\phi_j| = o(p(n)^{-r}). \end{aligned}$$

By virtue of Cauchy inequality and Theorem 2.1 (Hannan and Kavalieris [15]), i.e.

$$\max_{1 \leq j \leq p} |\hat{\phi}_j - \phi_{j,n}| = O_{a.s.} \left(\sqrt{\log n/n} \right),$$

we obtain

$$I_2 \leq \sqrt{p(n)} \max_{1 \leq j \leq p(n)} |\hat{\phi}_j - \phi_{j,n}| \sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2}.$$

It is clear that

$$\sqrt{\sum_{j=1}^{p(n)} X_{n+1-j}^2} = O_P(\sqrt{p(n)}).$$

Finally, we have:

$$I_2 = O_P\left(p(n)(\log(n)/n)^{1/2}\right) = O_P\left((n/\log(n))^{\frac{-r}{2(r+1)}}\right),$$

which completes the proof of Lemma 4.2. ■

Lemma 4.3. *Let us assume that:*

$$(i) \quad \left| P^*(X_n^* \leq u) - P(X \leq u) \right| \xrightarrow{P} 0,$$

for arbitrary u being a continuity point of the distribution function F_X ,

$$(ii) \quad Y_n^* \xrightarrow{P^*} 0 \quad \text{in probability,}$$

$$(iii) \quad V_n \xrightarrow{P} 0.$$

Then

$$\left| P^*(X_n^* + Y_n^* \leq u) - P(X + V_n \leq u) \right| \xrightarrow{P} 0.$$

The proof of Lemma 4.3 is straightforward and therefore is omitted.

Lemmas 4.4 and 4.5 show the useful representation of the prediction error and the bootstrap prediction error respectively.

Lemma 4.4. *For a arbitrary time horizon h ($h \in N$) we may represent the prediction error as:*

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= d_{1,h}(\underline{\phi}_{h-1}) \epsilon_{n+1} + \dots \\ &\quad + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1), \end{aligned}$$

where $d_{1,h}, \dots, d_{h-1,h}$ are some continuous functions and $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Proof of Lemma 4.4. The proof is by induction on h .

($h = 1$)

In this case we have:

$$\begin{aligned} X_{n+1} - \hat{X}_{n+1} &= \epsilon_{n+1} + \left(\sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} \right) \\ &= \epsilon_{n+1} + o_P(1), \end{aligned}$$

where the latter equality follows directly from Lemma 4.2.

($h > 1$)

Let us assume that for an arbitrary time horizon k , such that $1 \leq k \leq h-1$, we can represent the prediction error $X_{n+k} - \hat{X}_{n+k}$ in the following form:

$$\begin{aligned} (4.1) \quad X_{n+k} - \hat{X}_{n+k} &= \tilde{d}_{1,k}(\phi_1, \dots, \phi_{k-1}) \epsilon_{n+1} + \dots \\ &\quad + \tilde{d}_{k-1,k}(\phi_1, \dots, \phi_{k-1}) \epsilon_{n+k-1} + \epsilon_{n+k} + o_P(1). \end{aligned}$$

Further, the following recursive formula for the prediction error will be useful:

$$\begin{aligned} (4.2) \quad X_{n+h} - \hat{X}_{n+h} &= \sum_{j=1}^{h-1} \hat{\phi}_j (X_{n+h-j} - \hat{X}_{n+h-j}) \\ &\quad + \sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+h-j} \\ &\quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h}. \end{aligned}$$

Finally, using (4.2) and Lemma 4.2 we have:

$$\begin{aligned}
X_{n+h} - \hat{X}_{n+h} &= \sum_{j=1}^{h-1} \hat{\phi}_j \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) + \sum_{j=1}^{p(n)} \left(\phi_j - \hat{\phi}_j \right) X_{n+h-j} \\
&\quad + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+h-j} + \epsilon_{n+h} \\
&= \sum_{j=1}^{h-1} \hat{\phi}_j \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) + \epsilon_{n+h} + o_P(1) \\
&= \sum_{j=1}^{h-1} \phi_j \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) \\
&\quad + \sum_{j=1}^{h-1} \left(\hat{\phi}_j - \phi_j \right) \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) + \epsilon_{n+h} + o_P(1) \\
&= \sum_{j=1}^{h-1} \phi_j \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) + \epsilon_{n+h} + o_P(1),
\end{aligned}$$

the last equality being a consequence of the stochastic boundness of the prediction error (i.e., $X_{n+h} - \hat{X}_{n+h} = O_P(1)$) and fact that $\max_{1 \leq j \leq p(n)} |\phi_j - \hat{\phi}_j| = o_P(1)$ (Theorem 2.1 Hannan and Kavalieris [15]).

Applying now assumption (4.1) for all components $X_{n+h-j} - \hat{X}_{n+h-j}$, ($j \leq h-1$) we conclude the proof of Lemma 4.4. \blacksquare

Using similar arguments as in the proof of Lemma 4.4 we obtain the following:

Lemma 4.5. *For an arbitrary time horizon h ($h \in N$) we can represent the bootstrap prediction error as:*

$$\begin{aligned}
X_{n+h}^* - \hat{X}_{n+h}^* &= d_{1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+1}^* + \dots \\
&\quad + d_{h-1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1),
\end{aligned}$$

where $d_{1,h}, \dots, d_{h-1,h}$ are continuous functions as in the representation of the prediction error (Lemma 4.4) and $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Let us also mention (without proof) the version of continuity theorem, which will be used to establish the connection between the pointwise convergence of distribution functions and the convergence of corresponding characteristic functions.

Theorem 4.1 (Continuity theorem). *Suppose that:*

(i) *the sequence Y_n is tight,*

(ii) $P^*(X_n^* \leq u) - P(Y_n \leq u) \xrightarrow{P} 0$ for $u \in E$,

where E is a dense set (in R) and the Lebesgue measure of E^c equals 0

($P^(\cdot)$ stands for conditional distribution, given observations).*

Then

$$\phi_{X_n^*}(t) - \phi_{Y_n}(t) \xrightarrow{P} 0 \text{ for arbitrary } t \in R.$$

We present now the main theorem of this section.

Theorem 4.2 (Consistency of the hybrid bootstrap). *Let A1 with $s = 4$, A2 with $r \geq 2$ and B with $p(n) = o((\frac{n}{\log n})^{\frac{1}{2r+2}})$ hold. Then*

$$(4.3) \quad \left| P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq u) - P(X_{n+h} - \hat{X}_{n+h} \leq u) \right| = o_P(1)$$

for each u being a continuity point of distribution function of random variable $d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h$, where $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Proof of Theorem 4.2. We have divided the proof into two steps to make it clear.

Step I:

$h = 1$

According to $AR(\infty)$ representation for X_t we have

$$X_{n+1} = \sum_{j=1}^{\infty} \phi_j X_{n+1-j} + \epsilon_{n+1}.$$

The estimator of the best one-step linear predictor for X_{n+1} based on the autoregressive approximation $AR(p(n))$ is given by

$$\hat{X}_{n+1} = \hat{\phi}_1 X_n + \dots + \hat{\phi}_{p(n)} X_{n-p(n)+1}.$$

Similarly, one can express the best linear predictor for X_{n+1}^* , constructed using observation X_1^*, \dots, X_n^* , where $X_{n+1}^* = \sum_{j=1}^{p(n)} \hat{\phi}_j^* X_{n+1-j}^* + \epsilon_{n+1}^*$:

$$\hat{X}_{n+1}^* = \hat{\phi}_1^* X_n^* + \dots + \hat{\phi}_{p(n)}^* X_{n+1-p(n)}^*.$$

By Lemma 4.2 and i.i.d. and conditional i.i.d. properties of the sequences ϵ_t and ϵ_t^* accordingly, we get:

$$\begin{aligned} & \left| P^* \left(X_{n+1}^* - \hat{X}_{n+1}^* \leq u \right) - P \left(X_{n+1} - \hat{X}_{n+1} \leq u \right) \right| \\ &= \left| P^* \left(\epsilon_{n+1}^* \leq u \right) - P \left(\epsilon_{n+1} \leq u - \left(\sum_{j=1}^{p(n)} (\phi_j - \hat{\phi}_j) X_{n+1-j} \right. \right. \right. \\ & \quad \left. \left. \left. + \sum_{j=p(n)+1}^{\infty} \phi_j X_{n+1-j} \right) \right) \right| \\ &= \left| P^* \left(\epsilon_{n+1}^* \leq u \right) - P(\epsilon_{n+1} \leq u - o_P(1)) \right| \\ &= \left| P^* \left(\epsilon_1^* \leq u \right) - P(\epsilon_1 \leq u - o_P(1)) \right| \\ &\leq |P^* \left(\epsilon_1^* \leq u \right) - P(\epsilon_1 \leq u)| + |P(\epsilon_1 \leq u) - P(\epsilon_1 + o_P(1) \leq u)|. \end{aligned}$$

Applying Lemma 4.1 and Slutsky's Lemma we obtain for an arbitrary u - the continuity point of distribution ϵ_t

$$\left| P^* \left(X_{n+1}^* - \hat{X}_{n+1}^* \leq u \right) - P \left(X_{n+1} - \hat{X}_{n+1} \leq u \right) \right| = o_P(1),$$

which completes the Proof of Step I.

Step II:

$h > 1$

Let us recall that by Lemmas 4.4 and 4.5 the prediction error and the bootstrap prediction error may be expressed as:

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+1} + \dots \\ &\quad + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1), \end{aligned}$$

and

$$\begin{aligned} X_{n+h}^* - \hat{X}_{n+h}^* &= d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+1}^* + \dots \\ &\quad + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1), \end{aligned}$$

where $d_{1,h}, \dots, d_{h-1,h}$ are continuous functions and $\underline{\phi}_{h-1} = (\phi_1, \dots, \phi_{h-1})$.

Thus, we have:

$$\begin{aligned} (4.4) \quad &P^* \left(X_{n+h}^* - \hat{X}_{n+h}^* \leq u \right) - P \left(X_{n+h} - \hat{X}_{n+h} \leq u \right) = \\ &= P^* \left(d_{1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+1}^* + \dots + d_{h-1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+h-1}^* \right. \\ &\quad \left. + \epsilon_{n+h}^* + o_{P^*}(1) \leq u \right) + \\ &\quad - P \left(d_{1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+1} + \dots + d_{h-1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{n+h-1} \right. \\ &\quad \left. + \epsilon_{n+h} + o_P(1) \leq u \right) \\ &= P^* \left(d_{1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_1^* + \dots + d_{h-1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{h-1}^* + \epsilon_h^* + o_{P^*}(1) \leq u \right) + \\ &\quad - P \left(d_{1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_1 + \dots + d_{h-1,h} \left(\underline{\phi}_{h-1} \right) \epsilon_{h-1} + \epsilon_h + o_P(1) \leq u \right). \end{aligned}$$

In the latter equality we used the fact that ϵ_t is i.i.d. sequence and ϵ_t^* is conditionally i.i.d. sequence.

Lemma 4.1 and i.i.d. properties of ϵ_t^* i ϵ_t imply:

$$(4.5) \quad \begin{aligned} & P^*(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1^* + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1}^* + \epsilon_h^* \leq u) + \\ & -P(d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h \leq u) \xrightarrow{P} 0 \end{aligned}$$

for an arbitrary u being a continuity point of distribution function of random variable

$$d_{1,h}(\underline{\phi}_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\underline{\phi}_{h-1}) \epsilon_{h-1} + \epsilon_h.$$

Applying (4.4) and Lemma 4.3 we get:

$$P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq u) - P(X_{n+h} - \hat{X}_{n+h} \leq u) \xrightarrow{P} 0,$$

and the proof is complete. ■

Remark 4.1 (Uniform convergence). Under the assumptions of Theorem 4.2 with an additional assumption of the continuity of the distribution of the sequence $\{\epsilon_t\}$ one may obtain the consistency in uniform metric, i.e.,:

$$\sup_{x \in R} \left| P^*(X_{n+h}^* - \hat{X}_{n+h}^* \leq x) - P(X_{n+h} - \hat{X}_{n+h} \leq x) \right| \xrightarrow{P} 0.$$

Remark 4.2. Using essentially the same arguments as in the proof of Theorem 4.2 one may prove the consistency of the hybrid bootstrap when replicates of future bootstrap observations X_{n+h}^* are constructed using formula (3.3) (as in Remark 3.1) instead of formula (3.2). However, in this case we have to impose stronger condition that assumption A2 holds for $r > 2$.

Let us mention two important consequences of Theorem 4.2.

Corollary 4.1 (Convergence of characteristic functions). *Since Theorem 4.2 holds for a dense (in R) set and the sequence $\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$ is tight, we obtain from the continuity theorem (Theorem 4.1) the closeness of the corresponding characteristic functions:*

$$\phi_{\Delta_n^*(h)}^*(t) - \phi_{\Delta_n(h)}(t) \xrightarrow{P} 0$$

for arbitrary $t \in R$, where $\phi_{\Delta_n^*(h)}^*$ and $\phi_{\Delta_n(h)}$ denote characteristic functions of $\Delta_n^*(h)$ and $\Delta_n(h)$, accordingly.

Corollary 4.2 (Consistency in the Lévy metric). *Applying a modified version of Lemma 10 (Belyaev [5]) and the tightness of $\Delta_n(h)$, we obtain that the closeness of characteristic functions (formulated in Corollary 4.1) is equivalent to the closeness of distribution functions in the Lévy metric, i.e.,:*

$$\rho_L \left(F_{\Delta_n^*(h)}^*, F_{\Delta_n(h)} \right) \xrightarrow{P} 0 \text{ when } n \rightarrow \infty,$$

where ρ_L denotes the Lévy metric and $F_{\Delta_n^*(h)}^*$ and $F_{\Delta_n(h)}$ stand for the distribution function of $\Delta_n^*(h)$ and $\Delta_n(h)$ accordingly.

5. CONSISTENCY OF PREDICTION INTERVALS

In Section 4, the consistency of the hybrid bootstrap has been investigated.

However, when studying the asymptotic performance of bootstrap confidence intervals the main concern is whether the coverage probability of the confidence intervals converges to the nominal level when $n \rightarrow \infty$. In this way the consistency and the accuracy of bootstrap confidence intervals are defined. (see Definition 4.1 and 4.2, Shao and Tu [20]).

Other accuracy measures, such as the length of intervals are often used to compare different approaches to the construction of confidence intervals. We will analyze this topic more carefully in Section 7 devoted to computer simulation.

It is worth pointing out that for the i.i.d. case the consistency of bootstrap distribution estimator usually implies the consistency of bootstrap confidence intervals (see Theorem 4.1, Shao and Tu [20]). However, the reverse implication may not be true (Example 3.8, Shao and Tu [20]).

Let us now introduce the formal definition of the consistent prediction interval.

Definition 5.1 (Consistent prediction interval). A prediction interval $\hat{I}(h)$ with nominal confidence level $(1 - 2\alpha)$, constructed for a future (unknown) value X_{n+h} , ($h \geq 1$) is consistent if

$$P \left(X_{n+h} \in \hat{I}(h) \right) \rightarrow 1 - 2\alpha, \text{ when } n \rightarrow \infty.$$

To prove the consistency of prediction intervals we will use auxiliary results on convergence of quantiles for a weakly convergent sequence of distribution functions (Lemma 5.1 and its modification for the conditional case Lemma 5.2).

Lemma 5.1 (Lemma 1.2.1, Politis and Romano [19]). *If $\{G_n\}$ is a sequence of distribution functions, weakly convergent to the distribution function G (i.e., $G_n \Rightarrow G$) and if G is continuous and strictly increasing in $y = G^{-1}(\alpha)$, then*

$$G_n^{-1}(\alpha) \rightarrow G^{-1}(\alpha) \quad \text{when } n \rightarrow \infty.$$

The straightforward modification of Lemma 5.1 for the conditional distributions is the following:

Lemma 5.2. *If $\{F_n^*\}$ is a sequence of distribution functions, weakly convergent to the distribution function F in probability (i.e., $F_n^* \Rightarrow F$ in probability) and if F is continuous and strictly increasing in $y = F^{-1}(\alpha)$, then*

$$F_n^{*-1}(\alpha) \rightarrow F^{-1}(\alpha) \quad \text{in probability.}$$

Let us now present the main result related to the consistency of prediction intervals.

In Section 3, we defined hybrid bootstrap prediction intervals with nominal confidence level $(1 - 2\alpha)$:

$$(5.1) \quad I_B(h) = \left[\hat{X}_{n+h} + q_\alpha^*, \hat{X}_{n+h} + q_{1-\alpha}^* \right],$$

where q_α^* and $q_{1-\alpha}^*$ are quantiles of the distribution of the bootstrap prediction error $\Delta_n^*(h) = X_{n+h}^* - \hat{X}_{n+h}^*$.

Replacing unknown quantiles $q_\alpha^*, q_{1-\alpha}^*$ by their Monte Carlo approximations based on B bootstrap replicates of the statistics $\Delta_n^*(h)$, we obtain:

$$(5.2) \quad \hat{I}_B(h) = \left[\hat{X}_{n+h} + \hat{q}_\alpha^*, \hat{X}_{n+h} + \hat{q}_{1-\alpha}^* \right].$$

Theorem 5.1 (Consistency of hybrid prediction intervals). *Let assumptions A1 with $s = 4$, A2 with $r > 1$ and B with $p(n) = o((n/\log(n))^{1/(2r+2)})$ hold. Furthermore, let $c_\alpha, c_{1-\alpha}$ – quantiles of the distribution of a random variable $A_{\epsilon, h}$ (given as in the proof) be continuity points of that distribution function. Then*

$$P\left(X_{n+h} \in \hat{I}_B(h)\right) \longrightarrow 1 - 2\alpha, \quad \text{when } n \longrightarrow \infty.$$

Proof of Theorem 5.1. Let us first note that:

$$\begin{aligned} P\left(X_{n+h} \in \hat{I}_B(h)\right) &= P\left(\hat{X}_{n+h} + \hat{q}_\alpha^* \leq X_{n+h} \leq \hat{X}_{n+h} + \hat{q}_{1-\alpha}^*\right) \\ &= P\left(\hat{q}_\alpha^* \leq X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*\right) \\ &= P\left(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*\right) - P\left(X_{n+h} - \hat{X}_{n+h} < \hat{q}_\alpha^*\right). \end{aligned}$$

It remains to prove that, with n tending to ∞ :

$$P\left(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*\right) \longrightarrow 1 - \alpha$$

and

$$P\left(X_{n+h} - \hat{X}_{n+h} < \hat{q}_\alpha^*\right) \longrightarrow \alpha.$$

Let us recall that the prediction error and the bootstrap prediction error may be written as:

$$\begin{aligned} X_{n+h} - \hat{X}_{n+h} &= d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+1} + \dots \\ &\quad + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1), \end{aligned}$$

and

$$\begin{aligned} X_{n+h}^* - \hat{X}_{n+h}^* &= d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+1}^* + \dots \\ &\quad + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1). \end{aligned}$$

Let us denote

$$A_{\epsilon,h} := d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{h-1} + \epsilon_h.$$

Hence

$$(5.3) \quad \Delta_n(h) = X_{n+h} - \hat{X}_{n+h} \xrightarrow{d} A_{\epsilon,h}.$$

Moreover, by Lemma 5.4 (Bühlmann [8]) we obtain:

$$(5.4) \quad \Delta_n^*(h) = X_{n+h}^* - \hat{X}_{n+h}^* \xrightarrow{d^*} A_{\epsilon,h} \quad \text{in probability.}$$

Let $q_{1-\alpha}$ denote the quantile of order $(1-\alpha)$ of the prediction error $\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}$. Additionally, let $c_{1-\alpha}$ stand for the quantile of order $(1-\alpha)$ of distribution of a random variable $A_{\epsilon,h}$. From Lemma 5.1, Lemma 5.2 and the consistency of empirical quantiles we have:

$$q_{1-\alpha}^* - q_{1-\alpha} = o_P(1),$$

$$q_{1-\alpha} - c_{1-\alpha} = o(1),$$

$$\hat{q}_{1-\alpha}^* - q_{1-\alpha}^* = o_P(1).$$

Therefore, we may write:

$$\begin{aligned} & P\left(X_{n+h} - \hat{X}_{n+h} \leq \hat{q}_{1-\alpha}^*\right) \\ &= P\left(X_{n+h} - \hat{X}_{n+h} \leq (\hat{q}_{1-\alpha}^* - q_{1-\alpha}^*) + (q_{1-\alpha}^* - q_{1-\alpha}) \right. \\ &\quad \left. + (q_{1-\alpha} - c_{1-\alpha}) + c_{1-\alpha}\right) \\ &= P\left(X_{n+h} - \hat{X}_{n+h} + o_P(1) \leq c_{1-\alpha}\right). \end{aligned}$$

By virtue of Slutsky's Lemma

$$X_{n+h} - \hat{X}_{n+h} + o_P(1) \xrightarrow{d} A_{\epsilon,h}.$$

We thus get

$$P\left(X_{n+h} - \hat{X}_{n+h} + o_P(1) \leq c_{1-\alpha}\right) \longrightarrow P(A_{\epsilon,h} \leq c_{1-\alpha}) = 1 - \alpha,$$

if $c_{1-\alpha}$ is a continuity point of the distribution function of a random variable $A_{\epsilon,h}$.

In the same manner we can see that:

$$P\left(X_{n+h} - \hat{X}_{n+h} < \hat{q}_\alpha^*\right) \longrightarrow \alpha,$$

if c_α is a continuity point of the distribution function of a random variable $A_{\epsilon,h}$, which completes the proof of Theorem 5.1. ■

Remark 5.1. Let us note that the consistency of the hybrid bootstrap prediction intervals was proved for confidence level $(1 - 2\alpha)$, such that $c_\alpha, c_{1-\alpha}$ – quantiles of the distribution of random variable $A_{\epsilon,h} = d_{1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_1 + \dots + d_{h-1,h}(\phi_1, \dots, \phi_{h-1}) \epsilon_{h-1} + \epsilon_h$ are continuity points of distribution function.

Assuming additionally that the distribution function of ϵ_t is continuous we obtain the consistency of bootstrap prediction intervals with arbitrary confidence level $(1 - 2\alpha)$.

6. STUDENTIZED PREDICTION INTERVALS

In Sections 3–5, we have analyzed the construction and theoretical properties of the hybrid bootstrap prediction intervals.

It is worth pointing out that we can construct prediction intervals also adopting the idea of studentization, which yields so called bootstrap-t or studentized prediction intervals. More precisely, the unknown distribution of studentized statistics

$$T_n(h) = \frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)},$$

is estimated by the corresponding bootstrap distribution of

$$T_n^*(h) = \frac{X_{n+h}^* - \hat{X}_{n+h}^*}{\hat{\sigma}_n^*(h)},$$

where $\hat{\sigma}_n^2(h)$ and $\hat{\sigma}_n^{*2}(h)$ are the corresponding prediction mean squared errors (PMSE).

The studentized prediction interval can be expressed in the following form:

$$(6.1) \quad I_{B-t}(h) = \left[\hat{X}_{n+h} + t_{\alpha}^* \hat{\sigma}_n(h), \hat{X}_{n+h} + t_{1-\alpha}^* \hat{\sigma}_n(h) \right],$$

where t_{α}^* and $t_{1-\alpha}^*$ are quantiles of $T_n^*(h)$.

Replacing unknown quantiles $t_{\alpha}^*, t_{1-\alpha}^*$ by their Monte Carlo estimates based on B bootstrap samples of $T_n^*(h)$, we obtain:

$$(6.2) \quad \hat{I}_{B-t}(h) = \left[\hat{X}_{n+h} + \hat{t}_{\alpha}^* \hat{\sigma}_n(h), \hat{X}_{n+h} + \hat{t}_{1-\alpha}^* \hat{\sigma}_n(h) \right].$$

A detailed discussion of the theoretical properties of studentized prediction intervals is beyond the scope of this paper and will be presented in another one. Below, we only announce some results concerning consistency.

Theorem 6.1 contains a result on the consistency of the bootstrap-t, which uses an idea of weakly approaching sequences of random distributions (Belyaev [4], [5]).

Theorem 6.1 (Consistency of bootstrap-t). *Let assumptions A1 with $s = 4$, A2 with $r > 2$ ($r \in \mathbb{N}$) and B with $p(n) = o((n/\log(n))^{\frac{1}{2r+2}})$ hold. Then*

$$\mathcal{L}^* \left(\frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)} \right) \xrightarrow{wa(P)} \mathcal{L} \left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)} \right).$$

Theorem 6.2 is devoted to the consistency of bootstrap-t prediction intervals.

Theorem 6.2 (Consistency of the studentized prediction intervals). *Let assumptions A1 with $s = 4$, A2 with $r > 2$ and B with $p(n) = o((n/\log(n))^{1/(2r+2)})$ hold. Furthermore, let $u_{\alpha}, u_{1-\alpha}$ – quantiles of the distribution of a random variable $A_{\epsilon,h}/\sigma(h)$ (given as in the proof of Theorem 5.1) be continuity points of the distribution function. Then*

$$P \left(X_{n+h} \in \hat{I}_{B-t}(h) \right) \longrightarrow 1 - 2\alpha, \quad \text{when } n \longrightarrow \infty.$$

7. SIMULATION RESULTS

In order to investigate the finite sample performance of constructed prediction intervals we have conducted some computer simulations. In our numerical studies the following models were considered:

$$(M1) \quad ARMA(1, 1), \quad X_t = 0.8X_{t-1} - 0.6\varepsilon_{t-1} + \varepsilon_t,$$

$$(M2) \quad AR(48), \quad X_t = \sum_{j=1}^{48} \phi_j X_{t-j} + \varepsilon_t, \quad \phi_j = (-1)^{j+1} 7.5/(j+1)^3$$

$$(j = 1, \dots, 48).$$

For both models we use four different noise distributions:

(N) standard normal: $N(0, 1)$,

(t) t-Student: $t(3)/\sqrt{3}$,

(logN) log-normal: $(\log N(0, 1) - \sqrt{e})/\sqrt{e(e-1)}$,

(M) mixture of normal distributions: $0.9 N(-1, 1) + 0.1 N(9, 1)$.

Prediction intervals have been constructed using the classical Box-Jenkins approach based on the Gaussian approximation and two sieve bootstrap methods, i.e., hybrid prediction intervals and studentized (bootstrap-t) prediction intervals.

Gaussian intervals have been constructed assuming that the true underlying model for data is known. Thus they may be treated as a benchmark. On the other hand, to determine bootstrap prediction intervals (i.e., predictors and corresponding prediction mean squared errors) we use autoregressive approximation by $AR(p(n))$.

The following parameters have been used in simulations:

- sample size: $n = 25, 50, 100$,
- number of bootstrap replications: $B = 1000$,
- number of Monte Carlo repetitions: 1000.

The order $p(n)$ of autoregressive approximations is chosen by minimizing AIC (Akaike Information Criterion) in a range $p(n) \in [0, 10 \log_{10}(n)]$ (which is default for instance in S-PLUS). Results of data-driven choice of $\hat{p}_{AIC}(n)$ for models M1–M2 are given in Table 1.

Table 1. Estimation of order $p(n)$ using AIC.

Model	Distribution	n	Estimation $p(n)$ by AIC			
			$E(\hat{p}_{AIC})$	$std(\hat{p}_{AIC})$	$min(\hat{p}_{AIC})$	$max(\hat{p}_{AIC})$
M1	N	25	0.88	1.24	0	8
		50	1.52	1.89	0	16
		100	2.35	2.32	0	20
	t	25	0.70	1.06	0	7
		50	1.45	1.67	0	14
		100	2.22	2.23	0	20
	logN	25	0.79	1.21	0	8
		50	1.45	1.78	0	13
		100	2.27	2.16	0	18
	M	25	0.84	1.30	0	10
		50	1.48	1.85	0	12
		100	2.36	2.28	0	18
M2	N	25	1.56	1.03	0	8
		50	2.05	1.54	1	15
		100	2.77	2.12	1	19
	t	25	1.48	0.89	0	7
		50	1.96	1.36	1	12
		100	2.52	1.66	1	16
	logN	25	1.46	0.85	0	8
		50	1.93	1.50	1	14
		100	2.52	1.73	1	14
	M	25	1.54	1.03	0	10
		50	2.09	1.63	1	14
		100	2.65	1.99	1	19

Remark 7.1. Additionally, we have used computer simulations to compare two strategies for constructing bootstrap replicates of future observations X_{n+h}^* , given by (3.2) and (3.3) accordingly. Some comments on this issue have already been given in Remark 3.1 and Remark 4.2. However, simulation results did not show significant differences in both approaches and in order to save the space we decided to present results only for the case given by formula (3.2), which is treated as a default one. Appropriate simulation results for the second approach are available from the authors upon request.

Figures 1–4 present selected prediction intervals constructed for models (M1-M2) under consideration with nominal confidence level 95%.

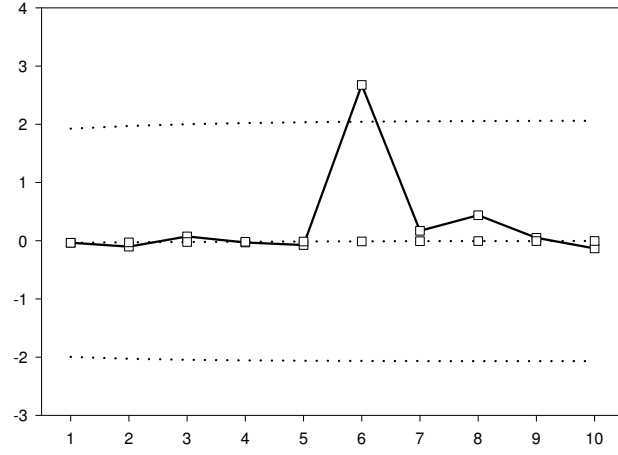


Figure 1. Gaussian prediction intervals for model M1 and M noise: true future values (solid line with squares), predictors (dotted line with squares), hybrid sieve bootstrap interval (dashdot line), sieve bootstrap-t (dotted line).

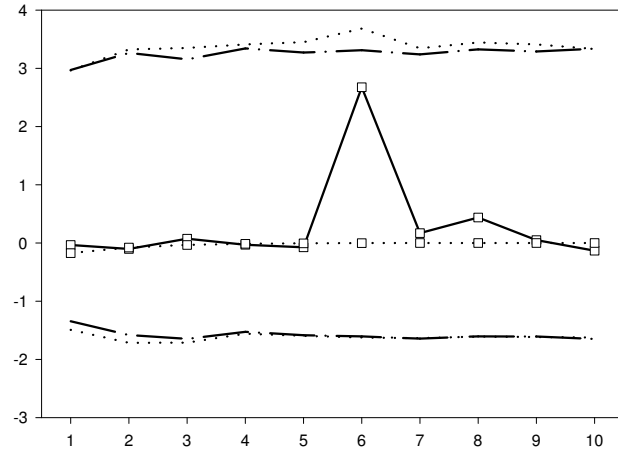


Figure 2. Bootstrap prediction intervals for M1 and M noise: true future values (solid line with squares), predictors (dotted line with squares), hybrid sieve bootstrap interval (dashdot line), sieve bootstrap-t (dotted line).

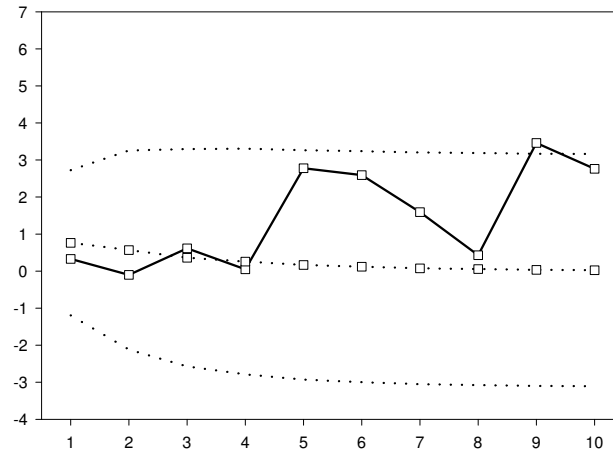


Figure 3. Gaussian prediction intervals for model M2 and logN noise: true future values (solid line with squares), predictors (dotted line with squares), hybrid sieve bootstrap interval (dashdot line), sieve bootstrap-t (dotted line).

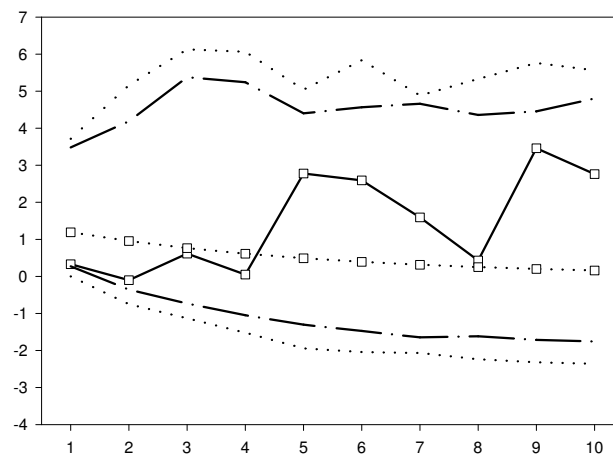


Figure 4. Bootstrap prediction intervals for M2 and logN noise: true future values (solid line with squares), predictors (dotted line with squares), hybrid sieve bootstrap interval (dashdot line), sieve bootstrap-t (dotted line).

Table 2. Empirical coverage probabilities for $n = 25$, nominal confidence level = 95%.

Model	Distri bution	h	Box-Jenkins	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	94.8%(0.702)	89.1%(0.985)	92.2%(0.848)	91.4%(0.887)	92.6%(0.828)
		2	94.3%(0.733)	90.2%(0.940)	92.3%(0.843)	92.3%(0.843)	92.1%(0.853)
		3	95.6%(0.649)	90.2%(0.940)	92.7%(0.823)	92.2%(0.848)	92.8%(0.817)
		4	94.6%(0.715)	88.5%(1.009)	92.3%(0.843)	92.6%(0.828)	92.9%(0.812)
		5	94.3%(0.733)	89.5%(0.969)	92.5%(0.833)	92.4%(0.838)	92.9%(0.812)
	t	1	95.9%(0.627)	93.1%(0.801)	94.4%(0.727)	94.9%(0.696)	95.0%(0.689)
		2	95.3%(0.669)	92.6%(0.828)	92.8%(0.817)	94.0%(0.751)	93.9%(0.757)
		3	95.0%(0.689)	91.5%(0.882)	91.5%(0.882)	91.9%(0.863)	92.1%(0.853)
		4	96.3%(0.597)	91.8%(0.868)	92.9%(0.812)	93.2%(0.796)	93.3%(0.791)
		5	96.2%(0.605)	92.1%(0.853)	92.9%(0.812)	92.9%(0.812)	93.0%(0.807)
	logN	1	96.6%(0.573)	85.3%(1.120)	89.0%(0.989)	92.5%(0.833)	93.4%(0.785)
		2	95.8%(0.634)	84.3%(1.150)	88.2%(1.020)	91.5%(0.882)	92.3%(0.843)
		3	97.5%(0.494)	85.6%(1.110)	89.1%(0.985)	91.3%(0.891)	92.6%(0.828)
		4	96.2%(0.605)	85.0%(1.129)	89.3%(0.977)	91.9%(0.863)	93.0%(0.807)
		5	95.4%(0.662)	84.6%(1.141)	88.0%(1.028)	90.4%(0.932)	92.2%(0.848)
	M	1	89.7%(0.961)	84.6%(1.141)	87.6%(1.042)	93.5%(0.780)	93.2%(0.796)
		2	90.7%(0.918)	87.1%(1.060)	89.5%(0.969)	95.3%(0.669)	94.9%(0.696)
		3	90.0%(0.949)	86.9%(1.067)	89.4%(0.973)	95.5%(0.656)	94.5%(0.721)
		4	91.6%(0.877)	85.6%(1.110)	88.2%(1.020)	95.1%(0.683)	95.3%(0.669)
		5	89.3%(0.978)	85.0%(1.129)	87.1%(1.060)	94.9%(0.696)	94.9%(0.696)
M2	N	1	94.5%(0.721)	89.3%(0.978)	92.0%(0.858)	92.4%(0.838)	92.9%(0.812)
		2	95.5%(0.656)	85.2%(1.123)	88.3%(1.016)	90.6%(0.923)	90.3%(0.936)
		3	95.0%(0.689)	81.3%(1.233)	85.9%(1.101)	90.2%(0.940)	90.1%(0.944)
		4	93.8%(0.763)	81.7%(1.223)	85.3%(1.120)	89.3%(0.978)	89.4%(0.973)
		5	95.8%(0.634)	82.6%(1.199)	87.3%(1.053)	90.4%(0.932)	90.2%(0.940)
	t	1	95.7%(0.641)	91.6%(0.877)	92.4%(0.838)	94.1%(0.745)	93.7%(0.768)
		2	95.8%(0.634)	85.8%(1.104)	87.0%(1.064)	89.7%(0.961)	90.0%(0.949)
		3	95.5%(0.656)	84.8%(1.135)	86.9%(1.067)	90.8%(0.914)	90.8%(0.914)
		4	94.8%(0.702)	82.7%(1.196)	85.8%(1.104)	90.4%(0.932)	90.7%(0.918)
		5	95.0%(0.689)	84.5%(1.144)	86.1%(1.094)	90.8%(0.914)	90.2%(0.940)
	logN	1	95.9%(0.627)	83.9%(1.162)	88.8%(0.997)	91.4%(0.887)	92.6%(0.828)
		2	95.8%(0.634)	77.0%(1.331)	80.8%(1.246)	87.7%(1.039)	88.4%(1.013)
		3	96.0%(0.620)	76.3%(1.345)	80.0%(1.265)	87.2%(1.057)	87.5%(1.046)
		4	96.2%(0.605)	75.8%(1.354)	78.7%(1.295)	87.1%(1.060)	87.8%(1.035)
		5	97.0%(0.539)	77.3%(1.325)	79.5%(1.277)	87.0%(1.064)	87.4%(1.049)
	M	1	91.7%(0.872)	87.1%(1.060)	89.3%(0.978)	94.7%(0.708)	94.6%(0.715)
		2	93.5%(0.780)	79.1%(1.286)	82.4%(1.204)	93.5%(0.780)	93.4%(0.785)
		3	96.0%(0.620)	76.3%(1.345)	79.1%(1.286)	92.3%(0.843)	91.8%(0.868)
		4	96.3%(0.597)	78.1%(1.308)	80.6%(1.251)	92.3%(0.843)	92.5%(0.833)
		5	96.7%(0.565)	79.1%(1.286)	81.8%(1.220)	92.9%(0.812)	93.0%(0.807)

Table 3. Empirical coverage probabilities for $n = 50$, nominal confidence level = 95% .

Model	Distri bution	h	Box-Jenkins	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	94.7%(0.708)	90.8%(0.914)	92.5%(0.833)	93.0%(0.807)	92.4%(0.838)
		2	95.7%(0.641)	92.8%(0.817)	94.0%(0.751)	94.2%(0.739)	93.7%(0.768)
		3	94.4%(0.727)	92.6%(0.828)	93.7%(0.768)	93.9%(0.757)	93.8%(0.763)
		4	94.3%(0.733)	92.5%(0.833)	93.7%(0.768)	93.3%(0.791)	94.0%(0.751)
		5	95.3%(0.669)	92.9%(0.812)	93.5%(0.780)	93.8%(0.763)	93.4%(0.785)
	t	1	96.0%(0.620)	93.0%(0.807)	93.6%(0.774)	95.3%(0.669)	95.0%(0.689)
		2	95.7%(0.641)	93.3%(0.791)	93.6%(0.774)	94.8%(0.702)	94.2%(0.739)
		3	95.8%(0.634)	92.6%(0.828)	93.0%(0.807)	93.8%(0.763)	93.7%(0.768)
		4	95.7%(0.641)	93.7%(0.768)	93.8%(0.763)	95.7%(0.641)	94.7%(0.708)
		5	95.6%(0.649)	93.2%(0.796)	93.8%(0.763)	94.5%(0.721)	94.5%(0.721)
	logN	1	96.7%(0.565)	87.1%(1.060)	89.5%(0.969)	93.4%(0.785)	94.9%(0.696)
		2	95.7%(0.641)	86.7%(1.074)	89.8%(0.957)	93.2%(0.796)	93.6%(0.774)
		3	96.2%(0.605)	88.6%(1.005)	90.5%(0.927)	93.8%(0.763)	94.7%(0.708)
		4	95.4%(0.662)	87.6%(1.042)	89.7%(0.961)	92.8%(0.817)	94.0%(0.751)
		5	97.1%(0.531)	88.5%(1.009)	90.7%(0.918)	93.7%(0.768)	94.8%(0.702)
	M	1	91.3%(0.891)	90.4%(0.932)	92.3%(0.843)	94.8%(0.702)	95.0%(0.689)
		2	90.4%(0.932)	90.3%(0.936)	91.6%(0.877)	95.3%(0.669)	95.3%(0.669)
		3	89.6%(0.965)	89.8%(0.957)	90.8%(0.914)	94.1%(0.745)	94.4%(0.727)
		4	89.4%(0.973)	90.0%(0.949)	91.5%(0.882)	95.6%(0.649)	95.7%(0.641)
		5	89.4%(0.973)	89.4%(0.973)	90.8%(0.914)	94.8%(0.702)	95.3%(0.669)
M2	N	1	94.9%(0.696)	89.8%(0.957)	91.8%(0.868)	92.6%(0.828)	92.8%(0.817)
		2	95.3%(0.669)	88.2%(1.020)	90.3%(0.936)	91.7%(0.872)	91.6%(0.877)
		3	95.0%(0.689)	89.6%(0.965)	90.8%(0.914)	92.5%(0.833)	93.1%(0.801)
		4	94.9%(0.696)	88.2%(1.020)	89.7%(0.961)	92.0%(0.858)	92.0%(0.858)
		5	95.0%(0.689)	88.2%(1.020)	90.5%(0.927)	91.9%(0.863)	92.1%(0.853)
	t	1	96.7%(0.565)	92.3%(0.843)	93.6%(0.774)	94.8%(0.702)	94.9%(0.696)
		2	96.4%(0.589)	91.7%(0.872)	92.2%(0.848)	94.4%(0.727)	94.0%(0.751)
		3	96.1%(0.612)	88.7%(1.001)	90.3%(0.936)	92.5%(0.833)	91.8%(0.868)
		4	95.6%(0.649)	88.1%(1.024)	89.0%(0.989)	91.2%(0.896)	90.8%(0.914)
		5	95.4%(0.662)	89.0%(0.989)	89.5%(0.969)	92.1%(0.853)	92.1%(0.853)
	logN	1	96.2%(0.605)	88.4%(1.013)	91.4%(0.887)	94.6%(0.715)	95.0%(0.689)
		2	95.1%(0.683)	86.2%(1.091)	88.4%(1.013)	93.0%(0.807)	92.7%(0.823)
		3	95.3%(0.669)	85.4%(1.117)	86.0%(1.097)	91.7%(0.872)	92.3%(0.843)
		4	95.2%(0.676)	85.8%(1.104)	86.4%(1.084)	90.6%(0.923)	91.0%(0.905)
		5	96.0%(0.620)	85.2%(1.123)	86.1%(1.094)	90.8%(0.914)	90.5%(0.927)
	M	1	90.0%(0.949)	90.4%(0.932)	92.1%(0.853)	95.3%(0.669)	95.5%(0.656)
		2	93.8%(0.763)	88.3%(1.016)	89.7%(0.961)	95.8%(0.634)	95.6%(0.649)
		3	95.2%(0.676)	84.9%(1.132)	86.5%(1.081)	93.1%(0.801)	93.6%(0.774)
		4	95.8%(0.634)	85.1%(1.126)	86.5%(1.081)	92.3%(0.843)	92.6%(0.828)
		5	94.8%(0.702)	84.1%(1.156)	84.9%(1.132)	92.1%(0.853)	91.7%(0.872)

Table 4. Mean of interval length for standard and smoothed bootstrap for $n = 25$ and confidence level 95%.

Model	Distri bution	h	Box-Jenkins	hybrid bootstrap		bootstrap-t	
				standard	smoothed	standard	smoothed
M1	N	1	3.920	3.693	3.910	3.907	3.925
		2	3.998	3.733	3.984	3.993	4.015
		3	4.047	3.768	4.029	4.044	4.067
		4	4.077	3.784	4.043	4.071	4.081
		5	4.097	3.793	4.054	4.083	4.097
	t	1	3.920	4.129	4.154	4.390	4.328
		2	3.998	4.137	4.192	4.466	4.388
		3	4.047	4.129	4.210	4.482	4.415
		4	4.077	4.155	4.220	4.486	4.417
		5	4.097	4.161	4.224	4.498	4.428
	logN	1	3.920	3.743	3.820	4.288	4.274
		2	3.998	3.767	3.859	4.380	4.358
		3	4.047	3.775	3.883	4.419	4.394
		4	4.077	3.790	3.884	4.427	4.412
		5	4.097	3.797	3.903	4.422	4.424
	M	1	3.920	3.728	3.806	4.443	4.411
		2	3.998	3.766	3.853	4.550	4.505
		3	4.047	3.793	3.873	4.595	4.551
		4	4.077	3.804	3.896	4.622	4.570
		5	4.097	3.809	3.907	4.626	4.602
M2	N	1	3.920	3.724	3.969	4.096	4.118
		2	5.373	4.233	4.621	5.051	5.069
		3	5.867	4.420	4.828	5.387	5.399
		4	6.094	4.504	4.924	5.527	5.542
		5	6.190	4.547	4.979	5.597	5.610
	t	1	3.920	4.109	4.164	4.535	4.471
		2	5.373	4.375	4.564	5.467	5.388
		3	5.867	4.520	4.725	5.842	5.728
		4	6.094	4.591	4.810	5.976	5.899
		5	6.190	4.629	4.852	6.071	5.955
	logN	1	3.920	3.494	3.604	4.164	4.152
		2	5.373	3.777	3.938	5.204	5.158
		3	5.867	3.933	4.102	5.637	5.511
		4	6.094	4.020	4.181	5.789	5.642
		5	6.190	4.073	4.231	5.851	5.735
	M	1	3.920	3.700	3.802	4.491	4.490
		2	5.373	4.085	4.273	5.565	5.541
		3	5.867	4.308	4.477	5.960	5.913
		4	6.094	4.419	4.581	6.134	6.087
		5	6.190	4.478	4.648	6.235	6.179

Table 5. Standard deviation of interval length for standard and smoothed bootstrap for $n = 25$ and confidence level 95%.

Model	Distribution	h	hybrid bootstrap		bootstrap-t	
			standard	smoothed	standard	smoothed
M1	N	1	0.755	0.722	0.748	0.697
		2	0.730	0.722	0.748	0.703
		3	0.713	0.715	0.736	0.706
		4	0.710	0.713	0.754	0.708
		5	0.706	0.715	0.754	0.706
	t	1	2.196	2.150	2.481	2.458
		2	2.188	2.133	2.505	2.457
		3	2.090	2.126	2.440	2.425
		4	2.175	2.131	2.495	2.376
		5	2.175	2.124	2.511	2.423
	logN	1	2.704	2.721	3.188	3.125
		2	2.703	2.748	3.332	3.336
		3	2.698	2.756	3.351	3.369
		4	2.698	2.744	3.386	3.369
		5	2.694	2.738	3.340	3.328
	M	1	0.844	0.833	1.013	0.989
		2	0.847	0.850	1.043	1.023
		3	0.855	0.876	1.070	1.058
		4	0.860	0.876	1.095	1.067
		5	0.868	0.876	1.087	1.079
M2	N	1	0.732	0.701	0.792	0.752
		2	0.778	0.848	0.985	0.976
		3	0.842	0.919	1.137	1.127
		4	0.868	0.959	1.203	1.195
		5	0.893	0.972	1.248	1.224
	t	1	2.269	2.212	2.561	2.523
		2	2.221	2.205	3.200	3.053
		3	2.232	2.226	3.529	3.291
		4	2.256	2.270	3.589	3.379
		5	2.266	2.268	3.728	3.490
	logN	1	2.204	2.226	2.874	2.705
		2	2.197	2.246	4.190	3.755
		3	2.235	2.300	4.778	4.126
		4	2.281	2.319	4.917	4.164
		5	2.319	2.356	4.833	4.277
	M	1	0.858	0.872	1.014	1.030
		2	1.084	1.133	1.389	1.397
		3	1.212	1.256	1.582	1.572
		4	1.268	1.306	1.678	1.665
		5	1.303	1.342	1.737	1.719

The accuracy of the constructed prediction intervals have been investigated with the aid of empirical coverage probabilities and mean of interval length ($E(length)$) based on 1000 Monte Carlo repetitions. A comparison was carried out for nominal confidence levels equal 80% and 95% and for forecast horizons $h = 1, \dots, 5$. Besides empirical coverage probabilities we calculated (given in parentheses) their standard errors. Below, only selected results are presented.

One can see that in all analyzed cases for studentized intervals we obtain a better empirical coverage than for hybrid intervals.

However, for a small sample size ($n = 25$) the coverage results for bootstrap intervals are not always satisfactory, which is due to a bias present in the estimation of model parameters.

Moreover, studentized prediction intervals yield much better coverage results than Gaussian Box-Jenkins intervals in the case of bimodal (M) noise distribution, especially for model $M1$.

For the nominal confidence level 80% we observe that non-Gaussian series Box-Jenkins intervals are too conservative and their mean length is the largest, too.

Additionally, in a numerical comparison the smoothed bootstrap has been taken into account.

The main idea of the smoothed bootstrap is generating resampled residuals ϵ_t^* using the smoothed estimator of the distribution function $\tilde{F}_{\epsilon,n}$ instead of the empirical distribution function as in the standard bootstrap approach.

We have used in simulation the kernel estimator of the distribution function:

$$\tilde{F}_{\epsilon}(x) = \frac{1}{n - p(n)} \sum_{t=1}^{n-p(n)} \tilde{K}\left(\frac{x - \tilde{\epsilon}_t}{b_n}\right),$$

where $\tilde{K}(z) = \int_{-\infty}^z K(u) du$, the kernel K is a symmetric probability density function and b_n is a smoothing parameter (bandwidth).

A detailed discussion including theoretical issues concerned with a smoothed bootstrap modification to construct prediction intervals will be given in a separate paper. We restrict ourselves to the presentation of some numerical results.

We study the application of the smoothed bootstrap to stabilization of studentized prediction intervals for a small sample size ($n = 25$). The analyses were based on the comparison of the mean interval length (Table 4) and standard deviation of interval length (Table 5) (similarly as Polansky [18]). We observe that the smoothed bootstrap yields a reduction of standard deviations of the interval length and for non-Gaussian error distribution also a reduction of the mean interval length. Additionally, empirical coverage probabilities remain on the same accuracy level as for the standard bootstrap (Table 2).

One can see that the application of the smoothed bootstrap improves also coverage results for hybrid bootstrap intervals (Table 2 and 3). However, this improvement is accompanied by an increase of the mean interval length (Table 4).

REFERENCES

- [1] A.M. Alonso, D. Peña and J. Romo, *Introducing model uncertainty in time series bootstrap*, Working Paper, Statistics and Econometric Series, Universidad Carlos III de Madrid **9** (2001), 1–14.
- [2] A.M. Alonso, D. Peña and J. Romo, *Forecasting time series with sieve bootstrap*, Journal of Statistical Planning and Inference, 100 (2002) 1–11.
- [3] T. Anderson, *The Statistical Analysis of Time Series*, Wiley, New York 1971.
- [4] Y.K. Belyaev, *The continuity theorem and its application to resampling from sums of random variables*, Theory of Stochastic Processes **3** (19) (1997), 100–109.
- [5] Y.K. Belyaev and S. Sjöstedt - de Luna, *Weakly approaching sequences of random distributions*, Journal of Applied Probability **37** (2000) 807–822.
- [6] P. Brockwell and R. Davis, *Time Series Theory and Method*, Springer-Verlag 1987.
- [7] P. Bühlmann, *Moving-average representation of autoregressive approximations*, Stochastic Processes and their Applications **60** (1995), 331–342.
- [8] P. Bühlmann, *Sieve bootstrap for time series*, Bernoulli **3** (2) (1997), 123–148.
- [9] P. Bühlmann, *Sieve bootstrap for smoothing in nonstationary time series*, The Annals of Statistics **26** (1) (1998), 48–83.
- [10] P. Bühlmann, *Bootstraps for Time Series*, Statistical Science **17** (1) (2002), 52–72.

- [11] R. Cao, et al., *Saving Computer time in constructing consistent bootstrap prediction intervals for autoregressive processes*, Comm. Statist. Simulation Comput. **26** (3) (1997), 961–978.
- [12] M. Deistler and E.J. Hannan, *The Statistical Theory of Linear Systems*, Wiley, New York 1988.
- [13] M. Grigoletto, *Bootstrap prediction intervals for autoregressions: some alternatives*, International Journal of Forecasting **14** (1998), 447–456.
- [14] U. Grenander, *Abstract Inference*, Wiley, New York 1981.
- [15] E.J. Hannan and L. Kavalieris, *Regressions, autoregression models*, Journal Time Series Analysis **7** (1986), 27–49.
- [16] J.H. Kim, *Bootstrap-after-bootstrap prediction intervals for autoregressive models*, Journal of Business & Economic Statistics **19** (1) (2001), 117–128.
- [17] G. Masarotto, *Bootstrap prediction intervals for autoregressions*, International Journal of Forecasting **6** (1990), 229–239.
- [18] A.M. Polansky, *Stabilizing bootstrap-t confidence intervals for small samples*, The Canadian Journal of Statistics **28** (3) (2000), 501–516.
- [19] D.N. Politis, J.P. Romano and M. Wolf, *Subsampling*, Springer-Verlag, New York 1999.
- [20] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer-Verlag, New York 1995.
- [21] R.A. Stine, *Estimating properties of autoregressive forecast*, Journal of the American Statistical Association **82** (400) (1987), 1073–1078.
- [22] L.A. Thombs and R. Schucany, *Bootstrap prediction intervals for autoregression*, Journal of the American Statistical Association **85** (410) (1990), 486–492.

Received 14 July 2003

Revised 6 January 2004