PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 25, Fasc. 1 (2005), pp. 133–153

ON THE CONSTRUCTION AND PROPERTIES OF BOOTSTRAP-t PREDICTION INTERVALS FOR STATIONARY TIME SERIES

BY

ADAM ZAGDAŃSKI (WROCŁAW AND TORONTO)

Abstract. We consider the construction of unconditional bootstrap-t prediction intervals for stationary time series. Our approach relies on the sieve bootstrap resampling scheme introduced by Bühlmann [8].

Basic theoretical properties concerned with consistency of the bootstrap approximation as well as consistency of constructed intervals are proved.

We generalize results obtained earlier by Stine [26], Masarotto [21] and Grigoletto [16] for autoregressive time series of finite order to the rich class of linear and invertible stationary models.

Finite sample accuracy of proposed bootstrap-t prediction intervals is verified by computer simulations. Empirical results of a comparative study show that our method is a superior alternative to both traditional Box–Jenkins approach and hybrid sieve-bootstrap prediction intervals proposed recently by Różański and Zagdański [24].

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

Key words and phrases: Prediction intervals, sieve bootstrap-t method of sieves.

1. INTRODUCTION

The computation of interval forecasts can be of vital importance in assessing future uncertainty as well as in many practical applications including planning (see e.g. Chatfield [12] for details). Construction of reliable prediction intervals for time series has been an ongoing problem.

The classical approach to constructing prediction intervals is based on Gaussian approximation of the prediction error distribution. This strategy is commonly used and is well known as the Box–Jenkins method. However, we cannot expect Box–Jenkins prediction intervals to perform very well for non-Gaussian series. Moreover, using this approach we do not incorporate the variability coming from model uncertainty.

Many authors consider more general bootstrap-based procedures of constructing prediction intervals. Let us only briefly name main results. Stine [26] considered construction of bootstrap prediction intervals for autoregressive processes of known order ($AR(p)$). Masarotto [21] and Grigoletto [16] generalized this construction for autoregressive processes $AR(p)$ with finite but possibly unknown order p. Thombs and Schucany [27] used the bootstrap method to construct conditional prediction intervals for autoregressive models of known order ($AR(p)$). Cao et al. [11] proposed modification of Thombs–Schucany's procedure which improves computational efficiency. Kim [20] constructed prediction intervals for vector autoregressive models of known order ($VAR(p)$) using the bootstrap-after-bootstrap approach which has built-in bias-correction procedure.

Approaches mentioned above are "model-based", i.e. rely on a finite-dimensional parametric model assumption. In order to generalize this construction to the broader class of time series the nonparametric resampling scheme should be used. Bühlmann [8], [9] has proposed resampling procedure, called *sieve bootstrap*, which has the advantage that no particular finite parametric model for data is assumed. This approach is based on Grenander's [15] method of sieves the main idea of which is approximating an infinite-dimensional nonparametric model by a sequence of finite-dimensional parametric models.

Alonso et al. [2] considered generalization of results obtained by Thombs and Schucany [27] and Cao et al. [11] for a general class of linear processes. Nonparametric conditional prediction intervals are constructed by using the sieve bootstrap approach. Simulation results indicate that the proposed method has better coverage and mean lengths results and is alternative for classical Box–Jenkins Gaussian prediction intervals.

Recently, Różański and Zagdański [24] have proposed construction of consistent unconditional hybrid prediction intervals using Bühlmann's sieve bootstrap scheme. This approach extends results obtained earlier by Stine [26], Masarotto [21] and Grigoletto [16] for autoregressive time series of finite order to the rich class of linear and invertible stationary models.

In this article we propose another promising method of constructing bootstrap prediction intervals for stationary time series. Namely, we use the sieve bootstrap resampling scheme to construct the so-called bootstrap-t (or studentized) prediction intervals. It is worth noting that for the independent set-up the bootstrap-t method was introduced by Efron [14]. However, this method has become particularly popular since some appealing theoretical properties were proved. Hall [17] showed, for instance, that in a quite general situation (so-called smooth function model) bootstrap-t has good second-order properties (i.e. bootstrap-t is second-order correct and second-order accurate). Moreover, empirical finite sample studies have revealed that the bootstrap-t methods perform well in terms of coverage error even for a small sample size, across a wide range of distributions.

The article is organized as follows. Section 2 contains model assumptions and detailed description of the sieve bootstrap algorithm. In Section 3 we discuss the problem of interval forecasts and construct bootstrap-t prediction intervals. Theoretical results concerning consistency of the bootstrap-t are given in Section 4. In Section 5 we investigate the consistency of constructed studentized prediction intervals. The last Section 6 is devoted to numerical results.

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

2. SIEVE BOOTSTRAP

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

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \ t\in Z, \tag{2.1}$$

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

Moreover, we will require the process X_t to be invertible, which 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 $\{X_t\}_{t\in Z}$ as an autoregressive process of order infinity ($AR(\infty)$), i.e.

$$X_t - \sum_{j=1}^{\infty} \phi_j X_{t-j} = \varepsilon_t, \quad t\in Z, \tag{2.2}$$

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

Using the notation $\Phi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j$, $z\in C$, and $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$, $\psi_0 = 1$, $z\in C$, one can represent X_t as:

$$AR(\infty)\colon \ \Phi(B)(X_t) = \varepsilon_t \quad \text{or} \quad MA(\infty)\colon \ X_t = \Psi(B)\varepsilon_t,$$

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

Let us also denote by $\mathcal{F}_t = \sigma(\{\varepsilon_s\colon s \leqslant t\})$ a σ-algebra generated by $\{\varepsilon_s\}_{s=-\infty}^{t}$. In the sequel, some of the following assumptions will be imposed:

(A1) $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ ($t\in Z$), where $\{\varepsilon_t\}_{t\in Z}$ is an i.i.d. sequence and $E[\varepsilon_t] = 0$, $E[\varepsilon_t]^s < \infty$ for some $s \geqslant 4$.

(A1′) $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ ($t\in Z$), where $\{\varepsilon_t\}_{t\in Z}$ is a stationary ergodic sequence and

$$E[\varepsilon_t \mid \mathcal{F}_{t-1}] \equiv 0, \quad E[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] \equiv \sigma^2 < \infty, \quad E|\varepsilon_t|^s < \infty \text{ for some } s \geqslant 4.$$

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

(B) $p = p(n) \to \infty$, $p(n) = o(n)$ $(n \to \infty)$ and $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})'$ satisfies the empirical Yule–Walker equations, i.e.

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

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

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

2.2. Sieve bootstrap algorithm. 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. $p(n) \to \infty$ when $(n \to \infty)$ but $p(n) = o(n)$. In other words, we approximate infinite-dimensional nonparametric model (which can be represented as $AR(\infty)$ process) by a sequence of finite-dimensional parametric models. This strategy is well known as the method of sieves (Grenander [15]). Let us now present the detailed algorithm of the sieve bootstrap.

Step 1. Let $X_1, \ldots, X_n$ be a realization of the process $\{X_t\}_{t \in Z}$. Applying Bühlmann's suggestion (cf. [8]) we choose an 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 practice, 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. $\phi_1, \ldots, \phi_{p(n)}$, using the observation $\{X_t\}_{t=1}^{n}$. Estimates $\hat{\phi}_p = (\hat{\phi}_1, \ldots, \hat{\phi}_{p(n)})'$ are determined by the Yule–Walker method (Brockwell and Davis [6], pp. 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, for instance, the recursive Durbin–Levinson algorithm (e.g. Brockwell and Davis [6]).

Step 3. We compute residuals

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

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

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

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

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

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

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

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

3. PREDICTION INTERVALS

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

$$P_n X_{n+h} = P_{\overline{\text{sp}}\{X_1,\ldots,X_n\}} X_{n+h}. \tag{3.1}$$

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], pp. 167–168).

Besides calculating optimal predictors we have to assess their accuracy. A well-known measure of uncertainty of the corresponding forecasts is prediction mean squared error. 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 prediction interval.

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

$$I(h, X) = [L(X), R(X)], \tag{3.2}$$

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

3.1. Box–Jenkins 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

$$I_G(h) = [P_n X_{n+h} - \Phi_{1-\alpha}\sigma_n(h),\ P_n X_{n+h} + \Phi_{1-\alpha}\sigma_n(h)], \tag{3.3}$$

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.

3.2. Hybrid bootstrap prediction intervals. We consider now the application of the sieve bootstrap procedure to constructing unconditional prediction intervals. The main idea of using sieve bootstrap to construct interval forecast is a generating replication $X_1^*, \ldots, X_n^*$ on the basis of the observed series $X_1, \ldots, X_n$, 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.

$$X_{n+h}^* = \hat{\phi}_1^* X_{n+h-1}^* + \ \ldots\ + \hat{\phi}_{p(n)}^* X_{n+h-p(n)}^* + \varepsilon_{n+h}^*, \tag{3.4}$$

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

Recently, Różański and Zagdański [24] have proposed construction of the hybrid sieve-bootstrap prediction intervals. Hybrid intervals are constructed approximating the unknown distribution of the prediction error

$$\Delta_n(h) = X_{n+h} - \hat{X}_{n+h} \tag{3.5}$$

by the corresponding bootstrap distribution of

$$\Delta_n^*(h) := X_{n+h}^* - \hat{X}_{n+h}^*, \tag{3.6}$$

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

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

$$I_B(h) = [\hat{X}_{n+h} + q_\alpha^*,\ \hat{X}_{n+h} + q_{1-\alpha}^*]. \tag{3.7}$$

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

3.3. Bootstrap-t prediction intervals. We can construct bootstrap 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

$$T_n(h) = \frac{X_{n+h} - \hat{X}_{n+h}}{\hat{\sigma}_n(h)} \tag{3.8}$$

is estimated by the corresponding bootstrap distribution of

$$T_n^*(h) = \frac{X_{n+h}^* - \hat{X}_{n+h}^*}{\hat{\sigma}_n^*(h)}, \tag{3.9}$$

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

Therefore, studentized sieve-bootstrap prediction interval can be expressed in the following form:

$$I_{B-t}(h) = [\hat{X}_{n+h} + t_\alpha^* \hat{\sigma}_n(h),\ \hat{X}_{n+h} + t_{1-\alpha}^* \hat{\sigma}_n(h)], \tag{3.10}$$

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

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

$$\hat{I}_{B-t}(h) = [\hat{X}_{n+h} + \hat{t}_\alpha^* \hat{\sigma}_n(h),\ \hat{X}_{n+h} + \hat{t}_{1-\alpha}^* \hat{\sigma}_n(h)]. \tag{3.11}$$

4. CONSISTENCY OF BOOTSTRAP-t

Let us introduce the necessary notation:

$\tilde{X}_{n+h}$ — the best (in mean squared sense) linear predictor for X_{n+h} constructed by assuming that the whole history $\{X_n, X_{n-1}, \ldots\}$ is known;

$\tilde{X}_{n+h}^n$ — the best linear predictor constructed on the basis of observations $\{X_n, \ldots, X_1\}$.

We formulate now some elementary facts concerning representation of the prediction error and the prediction mean squared error.

It is easily seen that the theoretical predictor $\tilde{X}_{n+h}$ is given by

$$\tilde{X}_{n+h} = \sum_{j=h}^{\infty} \psi_j \varepsilon_{n+h-j}. \tag{4.1}$$

As a straightforward consequence of (4.1) we infer that the prediction error and the prediction mean squared error may be represented in the following forms:

$$X_{n+h} - \tilde{X}_{n+h} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{n+h-j}, \tag{4.2}$$

$$E(X_{n+h} - \tilde{X}_{n+h})^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{4.3}$$

Let us also note that both predictors and prediction mean squared errors (PMSE) are equal when X_t is a finite-order autoregressive process $AR(p)$, i.e. the following lemma holds true.

LEMMA 4.1 (predictor and PMSE for $AR(p)$). *If X_t is an autoregressive process $AR(p)$ and $n > p$, then:*

$$\tilde{X}^n_{n+h} = \tilde{X}_{n+h} \quad \text{and} \quad E(X_{n+h} - \tilde{X}^n_{n+h})^2 = E(X_{n+h} - \tilde{X}_{n+h})^2.$$

In Section 3 we have used the following notation for predictors based on autoregressive approximation:

$\hat{X}_{n+h}$ – predictor for X_{n+h} calculated on the basis of the model $AR(p(n))$ with estimated coefficients $(\hat{\phi}_1, \ldots, \hat{\phi}_{p(n)})$ by using the observations $\{X_1, \ldots, X_n\}$;

$\hat{X}^*_{n+h}$ – predictor for X^*_{n+h} calculated on the basis of the model $AR(p(n))$ with parameters $(\hat{\phi}^*_1, \ldots, \hat{\phi}^*_{p(n)})$ estimated by using the bootstrap replicate $\{X^*_1, \ldots, X^*_n\}$.

For simplicity of the notation, we introduce the following symbols for appropriate prediction errors:

$$\Delta_n(h) = X_{n+h} - \hat{X}_{n+h}, \qquad \Delta^*_n(h) = X^*_{n+h} - \hat{X}^*_{n+h}.$$

In the sequel the following notation for prediction mean squared errors will also be used:

$\sigma^2(h)$ – prediction mean squared error corresponding to theoretical predictor $\tilde{X}_{n+h}$ based on the model $AR(\infty)$;

$\tilde{\sigma}^2_n(h)$ – prediction mean squared error for predictor based on the model $AR(p(n))$ with known coefficients $\boldsymbol{\phi}_{p(n)} = (\phi_{1,n}, \ldots, \phi_{p(n),n})$ satisfying theoretical Yule–Walker equations, i.e.

$$\Gamma_{p(n)} \boldsymbol{\phi}_{p(n)} = \boldsymbol{\gamma}_{p(n)}, \qquad \text{where } \Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}, \ \boldsymbol{\gamma}_{p(n)} = \big(\gamma(1), \ldots, \gamma(p(n))\big)',$$

$\hat{\sigma}^{*2}_n(h)$ – prediction mean squared error for bootstrap predictor $\hat{X}^*_{n+h}$.

Moreover, let $\hat{\sigma}^2_n(h)$ denote the estimate of prediction mean squared error $\tilde{\sigma}^2_n(h)$ obtained by substitution of unknown coefficients of the model $AR(p(n))$ by the corresponding Yule–Walker estimators $(\hat{\phi}_1, \ldots, \hat{\phi}_{p(n)})$.

Formulas (4.1) and (4.3) allow us to write:

$$\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2, \qquad \tilde{\sigma}^2_n(h) = \sigma^2 \sum_{j=0}^{h-1} \psi^2_{j,n},$$

$$\hat{\sigma}^2_n(h) = \hat{\sigma}^2 \sum_{j=0}^{h-1} \hat{\psi}_j^2, \qquad \hat{\sigma}^{*2}_n(h) = \hat{\sigma}^{*2} \sum_{j=0}^{h-1} \hat{\psi}^{*2}_j,$$

where $\{\psi_{j,n}\}_{j=0}^{\infty}$ are coefficients of the $MA(\infty)$ representation corresponding to the $AR(p(n))$ process with parameters $(\phi_{1,n}, \ldots, \phi_{p(n),n})$, $\{\hat{\psi}_j\}_{j=0}^{\infty}$ denote coefficients of the $MA(\infty)$ representation corresponding to the $AR(p(n))$ model

with estimated parameters $(\hat{\phi}_1, \ldots, \hat{\phi}_{p(n)})$, and $\{\hat{\psi}_j^*\}_{j=0}^{\infty}$ stand for coefficients of the $MA(\infty)$ representation for the $AR(p(n))$ process with replicated parameters $(\hat{\phi}_1^*, \ldots, \hat{\phi}_{p(n)}^*)$.

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

LEMMA 4.2 (Alonso et al. [1], Proposition 3.5). *Suppose that assumptions* (A1) *for* $s = 4$, (A2) *for* $r > 1$ $(r \in N)$, *and* (B) *for* $p(n) = o\left((n/\log(n))^{1/(2r+2)}\right)$ *hold true. Furthermore, assume that* $\boldsymbol{\phi}_{p(n)} = (\phi_{1,n}, \ldots, \phi_{p(n),n})$ *satisfies the theoretical Yule–Walker equations, i.e.*

$$\Gamma_{p(n)} \boldsymbol{\phi}_{p(n)} = \gamma_{p(n)}, \quad \text{where } \Gamma_{p(n)} = [\gamma(i-j)]_{i,j=1}^{p(n)}, \ \gamma_{p(n)} = \left(\gamma(1), \ldots, \gamma(p(n))\right)'.$$

Then

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

LEMMA 4.3. *Suppose that assumptions* (A1) *for* $s = 4$, (A2) *for* $r > 2$ $(r \in N)$ *and* (B) *for* $p(n) = o\left((n/\log(n))^{1/(2r+2)}\right)$ *hold true. Then*

$$\max_{1 \leq u \leq p(n)} |\hat{\psi}_u^* - \psi_u| \xrightarrow{P^*} 0 \text{ in probability.}$$

Proof of Lemma 4.3. Adopting the idea used by Bühlmann [7] (in the proof of Theorem 3.2) one may write:

$$\sigma^2 \psi_u = \sum_{j=0}^{\infty} \phi_j \gamma(u+j) \Rightarrow \psi_u = \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)$$

and

$$\hat{\sigma}^{*2} \hat{\psi}_u^* = \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j) \Rightarrow \hat{\psi}_u^* = \hat{\sigma}^{*-2} \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j).$$

Then

$$|\hat{\psi}_u^* - \psi_u| = \left|\hat{\sigma}^{*-2} \sum_{j=0}^{\infty} \hat{\phi}_j^* \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)\right|$$

$$= \left|\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} \hat{\phi}_j^* \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)\right|$$

$$= \left|\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} (\hat{\phi}_j^* - \phi_j) \hat{\gamma}^*(u+j) + \hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j) - \sigma^{-2} \sum_{j=0}^{\infty} \phi_j \gamma(u+j)\right|$$

$$= \left|\hat{\sigma}^{*-2} \sum_{j=0}^{p(n)} (\hat{\phi}_j^* - \phi_j) \hat{\gamma}^*(u+j) + (\hat{\sigma}^{*-2} - \sigma^{-2}) \sum_{j=0}^{p(n)} \phi_j \hat{\gamma}^*(u+j)\right.$$

$$\left. + \sigma^{-2} \sum_{j=0}^{p(n)} \phi_j \left(\hat{\gamma}^*(u+j) - \gamma(u-j)\right) - \sigma^{-2} \sum_{j=p(n)+1}^{\infty} \phi_j \gamma(u+j)\right|$$

$$\leqslant \left|\hat{\sigma}^{*-2}\sum_{j=0}^{p(n)}(\hat{\phi}_j^*-\phi_j)\hat{\gamma}^*(u+j)\right|+\left|(\hat{\sigma}^{*-2}-\sigma^{-2})\sum_{j=0}^{p(n)}\phi_j\hat{\gamma}^*(u+j)\right|$$

$$+\left|\sigma^{-2}\sum_{j=0}^{p(n)}\phi_j(\hat{\gamma}^*(u+j)-\gamma(u-j))\right|+\left|\sigma^{-2}\sum_{j=p(n)+1}^{\infty}\phi_j\gamma(u+j)\right|$$

$$=I_1+I_2+I_3+I_4.$$

Now let us bound the consecutive components. We have

$$I_1 \leqslant \hat{\sigma}^{*-2}\sum_{j=0}^{p(n)}|\hat{\phi}_j^*-\phi_j|\,|\hat{\gamma}^*(u+j)|$$

$$\leqslant \max_{0\leqslant j\leqslant p(n)}|\hat{\phi}_j^*-\phi_j|\sum_{j=0}^{p(n)}|\hat{\varrho}^*(u+j)| \leqslant \max_{1\leqslant j\leqslant p(n)}|\hat{\phi}_j^*-\phi_j|\cdot p(n).$$

Using Lemma 4.2 we get

$$\max_{1\leqslant j\leqslant p(n)}|\hat{\phi}_j^*-\phi_j| = O_{P^*}\left((n/\log n)^{-(r-1)/(2r+2)}\right) \text{ in probability.}$$

Hence we obtain

$$I_1 = O_{P^*}\left((n/\log n)^{-(r-1)/(2r+2)}\cdot p(n)\right) = O_{P^*}\left((n/\log n)^{-(r-2)/(2r+2)}\right) \text{ in probability.}$$

Further we have

$$I_2 \leqslant |\hat{\sigma}^{*-2}-\sigma^{-2}|\cdot\left|\sum_{j=0}^{p(n)}\phi_j\hat{\gamma}^*(u+j)\right|$$

$$=|\hat{\sigma}^{*-2}-\sigma^{-2}|\cdot\hat{\sigma}^{*-2}\cdot\left|\sum_{j=0}^{p(n)}\phi_j\hat{\varrho}^*(u+j)\right|$$

$$\leqslant |\hat{\sigma}^{*-2}-\sigma^{-2}|\cdot\hat{\sigma}^{*-2}\cdot\left|\sum_{j=0}^{p(n)}\phi_j\right| \leqslant o_P(1)\cdot\sum_{j=0}^{\infty}|\phi_j|.$$

In the inequalities above we have used Lemma 5.3 of Bühlmann [8]. Finally, since $\sum_{j=0}^{\infty}|\phi_j|<\infty$, we have $I_2 = o_P(1)$. We obtain also

$$I_3 \leqslant \sigma^{-2}\sum_{j=0}^{p(n)}|\phi_j|\cdot|\hat{\gamma}^*(u+j)-\gamma(u+j)|$$

$$\leqslant \sigma^{-2}\max_{0\leqslant j\leqslant p(n)}|\hat{\gamma}^*(u+j)-\gamma(u+j)|\sum_{j=0}^{p(n)}|\phi_j|$$

$$\leqslant \sigma^{-2}\max_{1\leqslant u\leqslant p(n)}\max_{0\leqslant j\leqslant p(n)}|\hat{\gamma}^*(u+j)-\gamma(u+j)|\sum_{j=0}^{\infty}|\phi_j|$$

$$\leqslant \sigma^{-2}\max_{1\leqslant k\leqslant 2p(n)}|\hat{\gamma}^*(k)-\gamma(k)|\sum_{j=0}^{\infty}|\phi_j|.$$

Let us put $\tilde{p}(n) = 2 \cdot p(n)$. Since $\sum_{j=0}^{\infty} |\phi_j| < \infty$, it is sufficient to show that

$$\max_{1 \leqslant k \leqslant \tilde{p}(n)} |\hat{\gamma}^*(k) - \gamma(k)| \xrightarrow{P^*} 0 \text{ in probability.}$$

It is easily seen that for $p(n) = o\left((n/\log n)^{1/(2r+2)}\right)$ (assumption (B)) also

$$\tilde{p}(n) = 2p(n) = o\left((n/\log n)^{1/(2r+2)}\right).$$

Further, we can use a similar boundedness to that applied in the proof of Proposition 3.5 by Alonso et al. [1]. Let us set the following notation for vector norms:

$$\|x\|_\infty = \max_{1 \leqslant i \leqslant k} |x_i|, \quad \|x\|_1 = \sum_{i=1}^{k} |x_i|, \quad \|x\|_2 = \sqrt{\sum_{i=1}^{k} x_i^2},$$

where $x = (x_1, \ldots, x_k)$.

Thus, we have

$$\|\hat{\gamma}^*_{\tilde{p}(n)} - \gamma_{\tilde{p}(n)}\|_\infty \leqslant \|\hat{\gamma}^*_{\tilde{p}(n)} - \gamma_{\tilde{p}(n)}\|_1 \leqslant \left(\tilde{p}(n)\right)^{1/2} \|\hat{\gamma}^*_{\tilde{p}(n)} - \gamma_{\tilde{p}(n)}\|_2,$$

where $\hat{\gamma}^*_{\tilde{p}(n)} = \left(\hat{\gamma}^*(1), \ldots, \hat{\gamma}^*\left(\tilde{p}(n)\right)\right)$, $\gamma_{\tilde{p}(n)} = \left(\gamma(1), \ldots, \gamma\left(\tilde{p}(n)\right)\right)$, and further

$$\left(\tilde{p}(n)\right)^{1/2} \|\hat{\gamma}^*_{\tilde{p}(n)} - \gamma_{\tilde{p}(n)}\|_2 = O_{P^*}\left((n/\log n)^{-(r-1)/(2r+2)}\right) \text{ in probability.}$$

Finally, we have shown that $I_3 = O_{P^*}\left((n/\log n)^{-(r-1)/(2r+2)}\right)$ in probability. Moreover, we have

$$I_4 \leqslant \sigma^{-2} \sum_{j=p(n)+1}^{\infty} |\phi_j| \, |\gamma(u+j)| = \sum_{j=p(n)+1}^{\infty} |\phi_j| \, |\varrho(u+j)| \leqslant \sum_{j=p(n)+1}^{\infty} |\phi_j|.$$

Assumption (A2) implies that $I_4 = o\left(p(n)^{-r}\right)$ and the proof is complete. ∎

LEMMA 4.4. *Suppose that assumptions* (A2) *for* $r \in N$ $(r \geqslant 1)$ *and* (B) *for* $p(n) = o\left((n/\log n)^{1/(2r+2)}\right)$ *hold true. Then*

$$\hat{\sigma}_n^2(h) \xrightarrow{P} \sigma^2(h).$$

Proof. Prediction mean squared error for the h-step predictor may be represented as $\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$, where σ^2 is the white noise variance and $\{\psi_j\}_{j=0}^{\infty}$ are coefficients of the $MA(\infty)$ expansion for X_t. Similarly one may write $\hat{\sigma}_n^2(h) = \hat{\sigma}^2 \sum_{j=0}^{h-1} \hat{\psi}_j^2$, where $\hat{\sigma}^2$ is the Yule–Walker estimate of white noise variance and $\{\hat{\psi}_j\}$ are coefficients of the $MA(\infty)$ representation for the $AR(p(n))$ model with estimated parameters $(\hat{\phi}_1, \ldots, \hat{\phi}_{p(n)})$. We have used the well-known fact (see, for instance, Brockwell and Davis [6]) that autoregressive model with parameters estimated by the Yule–Walker method is always causal.

From Theorem 3.2 of Bühlmann [7] we have

$$\sup_{j \in N} |\hat{\psi}_j - \psi_j| = O\left((\log(n)/n)^{1/2}\right) + O\left(p(n)^{-r}\right) \text{ a.s.}$$

and

$$|\hat{\sigma}^2-\sigma^2| = O\left((\log(n)/n)^{1/2}\right)+o\left(p(n)^{-r}\right) \text{ a.s.},$$

which completes the proof of Lemma 4.4. ∎

LEMMA 4.5. *If the assumptions of Lemma* 4.3 *hold, then*

$$\hat{\sigma}_n^{*2}(h) \xrightarrow{P^*} \sigma^2(h) \text{ in probability.}$$

Proof. As before one may write

$$\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \quad \text{and} \quad \hat{\sigma}_n^{*2}(h) = \hat{\sigma}^{*2} \sum_{j=0}^{h-1} \hat{\psi}_j^{*2},$$

where $\hat{\sigma}^{*2} = E^* \varepsilon_t^{*2}$ and $\{\hat{\psi}_j^*\}_{j=0}^{\infty}$ are coefficients of the $MA(\infty)$ representation for the $AR(p(n))$ process with parameters $\hat{\phi}_1^*, \ldots, \hat{\phi}_{p(n)}^*$.

From Lemma 5.3 of Bühlmann [8] it follows that $\hat{\sigma}^{*2}-\sigma^2 \xrightarrow{P} 0$. This implies also that $\hat{\sigma}^{*2}-\sigma^2 \xrightarrow{P^*} 0$ in probability.

Applying Lemma 4.3 we conclude the proof of Lemma 4.5. ∎

We formulate now the main result of this section concerned with the consistency of the bootstrap-t approximation. For this purpose the notion of weakly approaching sequences of random distributions introduced by Belyaev and Sjöstedt-de Luna (see [4] and [5]) will be used.

THEOREM 4.1 (*consistency of bootstrap-t*). *Suppose that assumptions* (A1) *for* $s = 4$, (A2) *for* $r > 2$ $(r \in N)$ *and* (B) *for* $p(n) = o\left((n/\log n)^{1/(2r+2)}\right)$ *hold true. Then*

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

Proof. Using the generalized version of Lemma 1 in Belyaev and Sjöstedt-de Luna [5] we have:

(4.4) $\mathcal{L}^*\left(\frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)}\right) \overset{wa(P)}{\longleftrightarrow} \mathcal{L}\left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)}\right)$ if and only if for all subsequences $\{n_k\}$ there exists a subsequence $\{n_{k_l}\}$ and a random variable Y_0 such that

$$\mathcal{L}^*\left(\frac{\Delta_{n_{k_l}}^*(h)}{\hat{\sigma}_{n_{k_l}}^*(h)}\right) \xrightarrow{w} \mathcal{L}(Y_0) \text{ in probability} \quad \text{and} \quad \mathcal{L}\left(\frac{\Delta_{n_{k_l}}(h)}{\hat{\sigma}_{n_{k_l}}(h)}\right) \xrightarrow{w} \mathcal{L}(Y_0).$$

From Theorem 5.1 of Różański and Zagdański [24] we obtain

$$\mathcal{L}^*\left(\Delta_n^*(h)\right) \overset{wa(P)}{\longleftrightarrow} \mathcal{L}\left(\Delta_n(h)\right).$$

Hence for all $\{n_k\}$ there exists $\{n_{k_l}\}$ and Δ_0 such that

$$\mathcal{L}^*\left(\Delta_{n_{k_l}}^*(h)\right) \xrightarrow{w} \mathcal{L}(\Delta_0) \text{ in probability} \quad \text{and} \quad \mathcal{L}\left(\Delta_{n_{k_l}}(h)\right) \xrightarrow{w} \mathcal{L}(\Delta_0)$$

or, equivalently,

$$\Delta^*_{n_{k_l}} \xrightarrow{d^*} \Delta_0 \text{ in probability} \quad \text{and} \quad \Delta_{n_{k_l}} \xrightarrow{d} \Delta_0.$$

Lemmas 4.4 and 4.5 imply

$$\hat{\sigma}^*_n(h) \xrightarrow{d^*} \sigma(h) \text{ in probability} \tag{4.5}$$

and

$$\hat{\sigma}_n(h) \xrightarrow{d} \sigma(h). \tag{4.6}$$

Let $\{n_k\}$ be an arbitrary subsequence. Further, let $\{n_{k_l}\}$ be a subsequence of $\{n_k\}$ for which

$$\Delta^*_{n_{k_l}}(h) \xrightarrow{d^*} \Delta_0 \text{ in probability} \quad \text{and} \quad \Delta_{n_{k_l}}(h) \xrightarrow{d} \Delta_0.$$

According to (4.5) and (4.6) we have

$$\hat{\sigma}^*_{n_{k_l}}(h) \xrightarrow{d^*} \sigma(h) \text{ in probability} \quad \text{and} \quad \hat{\sigma}_{n_{k_l}}(h) \xrightarrow{d} \sigma(h).$$

Since $\sigma(h) = \text{const}$ (is independent of n), we conclude from the Slutsky lemma and its modification that

$$\frac{\Delta^*_{n_{k_l}}(h)}{\hat{\sigma}^*_{n_{k_l}}(h)} \xrightarrow{d^*} \frac{\Delta_0}{\sigma(h)} \text{ in probability} \quad \text{and} \quad \frac{\Delta_{n_{k_l}}(h)}{\hat{\sigma}_{n_{k_l}}(h)} \xrightarrow{d} \frac{\Delta_0}{\sigma(h)},$$

which by virtue of (4.4) yields

$$\mathcal{L}^*\left(\frac{\Delta^*_n(h)}{\hat{\sigma}^*_n(h)}\right) \overset{wa(P)}{\longleftrightarrow} \mathcal{L}\left(\frac{\Delta_n(h)}{\hat{\sigma}_n(h)}\right),$$

and the proof is complete. ∎

5. CONSISTENCY OF BOOTSTRAP-t PREDICTION INTERVALS

In Section 4 the consistency of the bootstrap-t has been investigated. However, in analyzing asymptotic performance of bootstrap confidence intervals the main concern is whether the coverage probability of the confidence intervals converges to the nominal level as $n \to \infty$. Other accuracy measures, such as the length of intervals, are often used to compare different approaches to construction of confidence intervals. We will analyze this topic more carefully in Section 6 devoted to computer simulation. 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 future (unknown) value X_{n+h} $(h \geqslant 1)$, is *consistent* if

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

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

LEMMA 5.1 (Politis et al. [23], Lemma 1.2.1). *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*

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

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*

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

Let us now present the main result concerned with the consistency of prediction intervals.

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

$$I_{B-t}(h) = [\hat{X}_{n+h} + t_\alpha^* \hat{\sigma}_n(h),\ \hat{X}_{n+h} + t_{1-\alpha}^* \hat{\sigma}_n(h)], \tag{5.1}$$

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

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

$$\hat{I}_{B-t}(h) = [\hat{X}_{n+h} + \hat{t}_\alpha^* \hat{\sigma}_n(h),\ \hat{X}_{n+h} + \hat{t}_{1-\alpha}^* \hat{\sigma}_n(h)]. \tag{5.2}$$

THEOREM 5.1 (*consistency of the bootstrap-t prediction intervals*). *Suppose that assumptions* (A1) *with* $s = 4$, (A2) *with* $r > 2$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/(2r+2)}\big)$ *hold true. Denote by* $A_{\varepsilon,h}$ *the random variable*

$$d_{1,h}(\phi_1, \ldots, \phi_{h-1})\varepsilon_1 + \ldots + d_{h-1,h}(\phi_1, \ldots, \phi_{h-1})\varepsilon_{h-1} + \varepsilon_h,$$

where $d_{1,h}, \ldots, d_{h-1,h}$ *are continuous functions given as in the prediction error representation, i.e.*

$$X_{n+h} - \hat{X}_{n+h}$$

$$= d_{1,h}(\phi_1, \ldots, \phi_{h-1})\varepsilon_{n+1} + \ldots + d_{h-1,h}(\phi_1, \ldots, \phi_{h-1})\varepsilon_{n+h-1} + \varepsilon_{n+h} + o_P(1).$$

Furthermore, let μ_α *and* $u_{1-\alpha}$ *be quantiles of the distribution of the random variable* $A_{\varepsilon,h}/\sigma(h)$ *which are continuity points of the distribution function. Then*

$$P\big(X_{n+h} \in \hat{I}_{B-t}(h)\big) \to 1-2\alpha \quad \text{as } n \to \infty.$$

Proof. Let $t_{1-\alpha}$ stand for a $(1-\alpha)$-order quantile of the normalized prediction error $T_n(h) = \Delta_n(h)/\hat{\sigma}_n(h)$. Moreover, let $u_{1-\alpha}$ be a quantile of the distribution of the random variable $A_{\varepsilon,h}/\sigma(h)$, where $\sigma(h)$ is defined as in Section 4.

Applying Lemma 4.4 we see that $\hat{\sigma}_n(h) \xrightarrow{P} \sigma(h)$. In the proof of Theorem 5.1 in Różański and Zagdański [24] it was shown that

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

and

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

Hence, by (5.3) and the Slutsky lemma, we have

$$T_n(h) = \frac{\Delta_n(h)}{\hat{\sigma}_n(h)} \xrightarrow{d} \frac{A_{\varepsilon,h}}{\sigma(h)}. \tag{5.5}$$

Similarly, using Lemma 4.5, the fact (5.4) and the modified Slutsky lemma, we may write:

$$T_n^*(h) = \frac{\Delta_n^*(h)}{\hat{\sigma}_n^*(h)} \xrightarrow{d^*} \frac{A_{\varepsilon,h}}{\sigma(h)} \text{ in probability.} \tag{5.6}$$

Since $u_{1-\alpha}$ is a continuity point of the distribution function of the random variable $A_{\varepsilon,h}/\sigma(h)$, from Lemmas 5.1 and 5.2 and consistency of empirical quantiles we obtain

$$t_{1-\alpha}^* - t_{1-\alpha} = o_P(1), \quad t_{1-\alpha} - u_{1-\alpha} = o(1) \quad \hat{t}_{1-\alpha}^* - t_{1-\alpha}^* = o_P(1). \tag{5.7}$$

Further, (5.7) and the Slutsky lemma imply that

$$\begin{aligned} &P\left(\frac{X_{n+h}-\hat{X}_{n+h}}{\hat{\sigma}_n(h)} \leqslant \hat{t}_{1-\alpha}^*\right) \\ &\quad = P\left(T_n(h) \leqslant (\hat{t}_{1-\alpha}^* - t_{1-\alpha}^*) + (t_{1-\alpha}^* - t_{1-\alpha}) + (t_{1-\alpha} - u_{1-\alpha}) + u_{1-\alpha}\right) \\ &\quad = P\left(T_n(h) + o_P(1) \leqslant u_{1-\alpha}\right) \to P\left(\frac{A_{\varepsilon,h}}{\sigma(h)} \leqslant u_{1-\alpha}\right) = 1-\alpha \end{aligned} \tag{5.8}$$

if $u_{1-\alpha}$ is a continuity point of the distribution function of $A_{\varepsilon,h}/\sigma(h)$.

In the same manner we can see that

$$P\left(\frac{X_{n+h}-\hat{X}_{n+h}}{\hat{\sigma}_n(h)} \leqslant \hat{t}_{\alpha}^*\right) \to \alpha. \tag{5.9}$$

Finally, using (5.8) and (5.9) we conclude the proof of Theorem 5.1. ∎

Remark 5.1. Let us note that consistency of the bootstrap-t prediction intervals was proved for confidence level $(1-2\alpha)$ such that u_α and $u_{1-\alpha}$ — quantiles of the distribution of the random variable $A_{\varepsilon,h}/\sigma(h)$ — are continuity

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

6. SIMULATION RESULTS

6.1. Introduction. We have compared finite sample accuracy of constructed prediction intervals using computer simulations.

The following models were considered:

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

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

(M3) $ARFIMA(0, d, 0)$, $(1-B)^d X_t = \varepsilon_t$ for $d = 0.3$, where the fractional difference operator $\nabla^d = (1-B)^d$ is defined as in Hosking [19]. Let us note that for $d \in (-1/2, 1/2)$ $ARFIMA(0, d, 0)$ is a stationary process, which can be represented as $AR(\infty)$ process (e.g. Hosking [19]).

For all models we use four different noise distributions:

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

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

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

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

Prediction intervals have been constructed using the classical Box–Jenkins approach based on 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 benchmark. On the other hand, to determine bootstrap prediction intervals (i.e. predictors and the corresponding prediction mean squared errors) we use autoregressive approximation by $AR(p(n))$.

The following parameters have been used in simulations:

- sample size: $n = 25, 50, 100, 500$,
- 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 $\hat{p}_{\mathrm{AIC}}(n)$ for models M1–M3 are given in Table 1.

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

Model	Distribution	n	Estimation $p(n)$ using AIC			
			$E(\hat{p}_{\mathrm{AIC}})$	$\mathrm{std}(\hat{p}_{\mathrm{AIC}})$	$\min(\hat{p}_{\mathrm{AIC}})$	$\max(\hat{p}_{\mathrm{AIC}})$
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
	log N	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
	log N	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
M3	N	100	2.26	2.13	0	20
		500	5.40	3.48	1	26
	t	100	2.23	2.24	0	18
		500	5.47	3.63	1	26
	log N	100	2.06	1.79	0	16
		500	5.57	4.11	1	26
	M	100	2.25	2.26	0	17
		500	5.40	3.60	1	25

6.2. Results. Simulation results for models M1 and M2 were given in Różański and Zagdański [24]. In this article, we restrict ourselves to present some new results for model M3, which exhibits long-memory behavior.

Figures 1 and 2 present selected prediction intervals constructed for model M3 with nominal confidence level 95%.

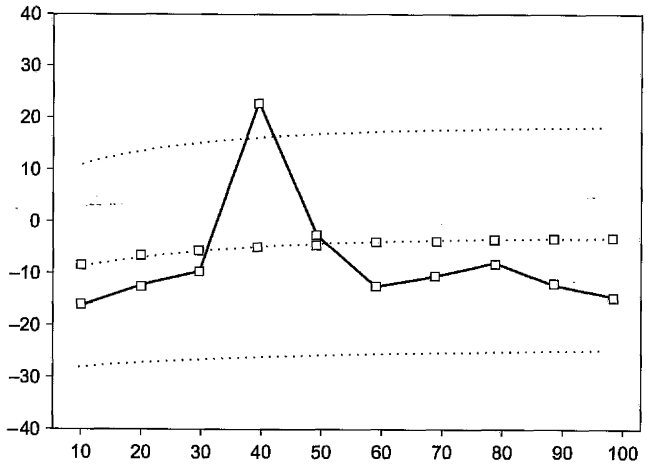

Fig. 1. Gaussian prediction intervals for model M3 and M noise: true future values (solid line with squares), predictors (dotted line with squares), Gaussian intervals (dotted line)

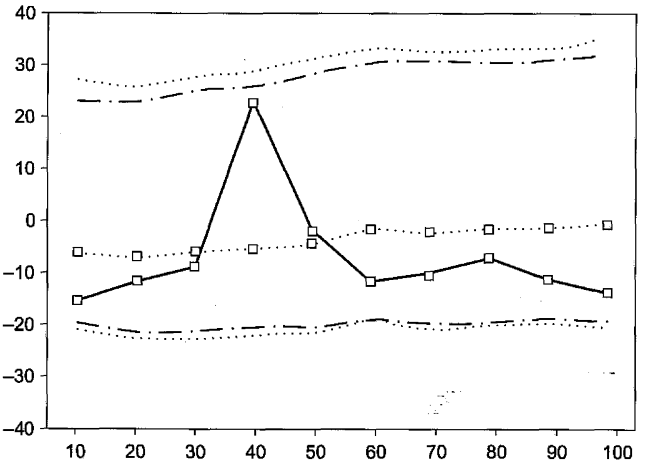

Fig. 2. Bootstrap prediction intervals for M3 and M noise: true future values (solid line with squares), predictors (dotted line with squares), hybrid sieve bootstrap interval (dash-dotted line), sieve bootstrap-t (dotted line)

The accuracy of constructed prediction intervals has been investigated with the aid of empirical coverage probabilities and mean of interval length

(*E*(*length*)) based on 1000 Monte Carlo repetitions. The comparisons were carried out for nominal confidence levels equal to 80% and 95% and for forecast horizons $h = 1, \ldots, 5$. Besides empirical coverage probabilities we calculated (given in parentheses) their standard errors.

One can see that in all analyzed cases for studentized intervals we obtain better empirical coverage than for hybrid intervals. However, for small sample size the coverage results for both bootstrap intervals are not always satisfactory, which is due to bias present in estimation of the model parameters. Moreover, studentized prediction intervals yield much better coverage results than Gaussian Box–Jenkins intervals in the case of bimodal (M) noise distribution. For nominal confidence level 80% we observe (e.g. Table 2) that for non-Gaussian series Box–Jenkins intervals are too conservative and their mean length is the largest one.

Table 2. Empirical coverage and mean interval length for model M3, $n = 100$, and nominal confidence level 80%

Model	Distribution	h	Box–Jenkins		hybrid bootstrap		bootstrap-t	
			coverage	*E*(*length*)	coverage	*E*(*length*)	coverage	*E*(*length*)
M3	N	1	80.7% (1.248)	2.564	77.9% (1.312)	2.489	79.2% (1.284)	2.546
		2	81.2% (1.236)	2.678	78.3%(1.304)	2.596	80.2% (1.260)	2.661
		3	82.3% (1.207)	2.725	79.9% (1.267)	2.646	81.2% (1.236)	2.714
		4	79.2% (1.284)	2.752	76.4% (1.343)	2.663	77.7% (1.316)	2.734
		5	80.8% (1.246)	2.770	78.4% (1.301)	2.671	80.1% (1.263)	2.741
	t	1	88.0% (1.028)	2.564	76.4% (1.343)	1.907	78.1% (1.308)	2.012
		2	88.8% (0.997)	2.678	78.2% (1.306)	2.042	79.8% (1.270)	2.162
		3	88.5% (1.009)	2.725	78.2% (1.306)	2.087	80.6% (1.251)	2.213
		4	87.8% (1.035)	2.752	78.9% (1.290)	2.115	80.2% (1.260)	2.241
		5	87.9% (1.031)	2.770	77.2% (1.327)	2.127	78.9% (1.290)	2.257
	log N	1	93.2% (0.796)	2.564	76.6% (1.339)	1.601	81.6% (1.225)	1.741
		2	92.4% (0.838)	2.678	74.1% (1.385)	1.752	79.4% (1.279)	1.913
		3	92.0% (0.858)	2.725	75.1% (1.368)	1.807	78.3% (1.304)	-1.970
		4	93.1% (0.801)	2.752	74.2% (1.384)	1.841	79.0% (1.288)	2.010
		5	93.4% (0.785)	2.770	77.0% (1.331)	1.851	80.8% (1.246)	2.022
	M	1	90.7% (0.918)	2.564	79.7% (1.272)	2.099	81.2% (1.236)	2.134
		2	89.4% (0.973)	2.678	77.5% (1.321)	2.309	79.6% (1.274)	2.358
		3	90.4% (0.932)	2.725	77.2% (1.327)	2.372	78.4% (1.301)	2.427
		4	89.7% (0.961)	2.752	76.9% (1.333)	2.403	78.9% (1.290)	2.459
		5	91.9% (0.863)	2.770	78.6% (1.297)	2.404	80.0% (1.265)	2.464

Table 3. Empirical coverage and mean interval length for model M3, $n = 100$, and nominal confidence level 95%

Model	Distribution	h	Box–Jenkins		hybrid bootstrap		bootstrap-t	
			coverage	E(*length*)	coverage	E(*length*)	coverage	E(*length*)
M3	N	1	95.3% (0.669)	3.922	93.4% (0.785)	3.808	93.6% (0.774)	3.912
		2	95.3% (0.669)	4.095	93.1%(0.801)	3.976	94.0% (0.751)	4.091
		3	95.0% (0.689)	4.167	94.5% (0.721)	4.056	94.8% (0.702)	4.173
		4	95.2% (0.676)	4.209	93.2% (0.796)	4.071	94.1% (0.745)	4.192
		5	95.4% (0.662)	4.237	93.8% (0.763)	4.088	94.3% (0.733)	4.213
	t	1	95.7% (0.641)	3.922	92.6% (0.828)	3.672	94.1% (0.745)	3.937
		2	96.0% (0.620)	4.095	93.3% (0.791)	3.841	94.2% (0.739)	4.158
		3	95.6% (0.649)	4.167	93.9% (0.757)	3.921	95.1% (0.683)	4.232
		4	95.5% (0.656)	4.209	93.6% (0.774)	3.961	94.3% (0.733)	4.280
		5	96.3% (0.597)	4.237	93.7% (0.768)	3.968	94.6% (0.715)	4.293
	log N	1	95.8% (0.634)	3.922	92.3% (0.843)	3.349	95.8% (0.634)	3.799
		2	96.5% (0.581)	4.095	92.5% (0.833)	3.573	96.2% (0.605)	4.034
		3	95.3% (0.669)	4.167	91.5% (0.882)	3.647	95.0% (0.689)	4.117
		4	96.5% (0.581)	4.209	92.7% (0.823)	3.700	95.6% (0.649)	4.178
		5	96.2% (0.605)	4.237	91.3% (0.891)	3.682	95.0% (0.689)	4.184
	M	1	90.7% (0.918)	3.922	94.1% (0.745)	3.949	95.6% (0.649)	4.173
		2	90.1% (0.944)	4.095	93.4% (0.785)	4.065	95.1% (0.683)	4.317
		3	90.9% (0.910)	4.167	93.5% (0.780)	4.122	96.1% (0.612)	4.384
		4	90.7% (0.918)	4.209	93.6% (0.774)	4.157	95.5% (0.656)	4.418
		5	93.2% (0.796)	4.237	93.2% (0.796)	4.169	94.8% (0.702)	4.436

REFERENCES

[1] A. M. Alonso, D. Peña and J. Romo, *Introducing model uncertainty in time series bootstrap*, Working Paper 01-14, Statistics and Econometric Series 9, Universidad Carlos III de Madrid (2001).

[2] A. M. Alonso, D. Peña and J. Romo, *Forecasting time series with sieve bootstrap*, J. Statist. Plann. Inference 100 (2002), pp. 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 Stoch. Process. 3 (19) (1997), pp. 100–109.

[5] Y. K. Belyaev and S. Sjöstedt-de Luna, *Weakly approaching sequences of random distributions*, J. Appl. Probab. 37 (2000), pp. 807–822.

[6] P. Brockwell and R. Davis, *Time Series Theory and Method*, Springer, New York 1987.

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

[8] P. Bühlmann, *Sieve bootstrap for time series*, Bernoulli 3 (2) (1997), pp. 123–148.

[9] P. Bühlmann, *Sieve bootstrap for smoothing in nonstationary time series*, Ann. Statist. 26 (1) (1998), pp. 48–83.

[10] P. Bühlmann, *Bootstrap for Time Series*, Statist. Sci. 17 (1) (2002), pp. 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), pp. 961–978.
[12] Ch. Chatfield, *Calculating Interval Forecasts*, J. Bus. Econom. Statist. 11 (2) (1993), pp. 121–135.
[13] M. Deistler and E. J. Hannan, *The Statistical Theory of Linear Systems*, Wiley, New York 1988.
[14] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM CMNS-Natl. Sci. Found. Monogr 38. SIAM, Philadelphia 1982, PA.
[15] U. Grenander, *Abstract Inference*, Wiley, New York 1981.
[16] M. Grigoletto, *Bootstrap prediction intervals for autoregressions: some alternatives*, Internat. J. Forecast. 14 (1998), pp. 447–456.
[17] P. Hall, *Theoretical comparison of bootstrap confidence intervals*, Ann. Statist. 16 (1988), pp. 927–985.
[18] E. J. Hannan and L. Kavalieris, *Regressions, autoregression models*, J. Time Ser. Anal. 7 (1986), pp. 27–49.
[19] J. R. M. Hosking, *Fractional differencing*, Biometrika 68 (1) (1981), pp. 165–176.
[20] J. H. Kim, *Bootstrap-after-bootstrap prediction intervals for autoregressive models*, J. Bus. Econom. Statist. 19 (1) 2001, pp 117–128.
[21] G. Masarotto, *Bootstrap prediction intervals for autoregressions*, Internat. J. Forecast. 6 (1990), pp. 229–239.
[22] A. M. Polansky, *Stabilizing bootstrap-t confidence intervals for small samples*, Canad. J. Statist. 28 (3) (2000), pp. 501–516.
[23] D. N. Politis, D. N. Romano and M. Wolf, *Subsampling*, Springer, New York 1999.
[24] R. Różański and A. Zagdański, *On the Consistency of Sieve Bootstrap Prediction Intervals for Stationary Time Series*, Discuss. Math. Probab. Statist. 24 (2004), pp. 5–40.
[25] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer, New York 1995.
[26] R. A. Stine, *Estimating properties of autoregressive forecast*, J. Amer. Statist. Assoc. 82 (400) (1987), pp. 1073–1078.
[27] L. A. Thombs and R. Schucany, *Bootstrap prediction intervals for autoregression*, J. Amer. Statist. Assoc. 85 (410) (1990), pp. 486–492.

Institute of Mathematics
Wrocław University of Technology
and
Department of Public Health Sciences, Faculty of Medicine
University of Toronto
E-mail: zagdan@im.pwr.wroc.pl

Received on 16.3.2005;
revised version on 21.6.2005