PREDICTION INTERVALS AND REGIONS FOR MULTIVARIATE TIME SERIES MODELS WITH SIEVE BOOTSTRAP

BY

ROMAN RÓŻAŃSKI (WROCŁAW), GRZEGORZ CHŁAPIŃSKI (WROCŁAW),

MARCIN HŁAWKA (WROCŁAW),

KRZYSZTOF JAMRÓZ (WROCŁAW),

MACIEJ KAWECKI (WROCŁAW), AND ADAM ZAGDAŃSKI (WROCŁAW)

Abstract. In the paper, the construction of unconditional bootstrap prediction intervals and regions for some class of second order stationary multivariate linear time series models is considered. Our approach uses the sieve bootstrap procedure introduced by Kreiss (1992) and Bühlmann (1997). Basic theoretical results concerning consistency of the bootstrap replications and the bootstrap prediction regions are proved. We present a simulation study comparing the proposed bootstrap methods with the Box–Jenkins approach.

2010 AMS Mathematics Subject Classification: Primary: 62M10, 62M20; Secondary: 62G09, 62G15, 62G05.

Key words and phrases: Multivariate time series models, vector of time series, sieve bootstrap, prediction regions, simultaneous prediction intervals.

1. INTRODUCTION

Determination of forecasts of time series future values based on previous observations is an extremely important – from a practical point of view – part of statistical data analysis. Since the high-speed personal computers have appeared, we can even deal with the prediction for a large number of dimensions of data. Methods of determining the prediction for the future and unknown observations are now frequently used in the world around us. They have a wide range of applications, both to predict the behavior of stock prices, stock indices, interest rates, and similar financial market and economic data ([27], [28]) as well in predicting the data of the general nature and geographic scope. For instance, the vector autoregressive models VAR were used by Di Battista et al. [11] in modeling the diversity of the population of some species in their natural environment, and Mirmirani and Li [22] used VAR models to predict the oil prices.

In this paper we present the construction of bootstrap prediction regions for some class of second order stationary multivariate linear time series models. We consider both hybrid bootstrap and bootstrap-t methods. Using the Bonferroni inequality, we can construct a multivariate bootstrap prediction cube, i.e. we construct a prediction interval for each coordinate. We consider also bootstrap simultaneous prediction intervals based on extreme statistics. They are an alternative to the bootstrap prediction intervals based on the Bonferroni correction. The main advantages of the bootstrap methods are nonparametricity (no specific assumptions about the form of the model) and easiness to implement. Thus, the bootstrap methods are a natural alternative to the methods under general asymptotic statistical considerations (e.g. with the popular assumption of the normality of noise distribution).

We consider the most popular and nonparametric method for constructing the replication of time series data, namely the sieve bootstrap. The algorithm was proposed by Kreiss [18] and Bühlmann [9]. Their idea uses a Grenander sieve [14] involving the approximation of infinite-dimensional model by a sequence of finite-dimensional models whose size increases with the number of observations n. For the class of stationary and invertible time series models ($VAR(\infty)$ models), Bühlmann proposed approximation as a sequence of vector autoregressive ($VAR(p)$) models, where $p = p(n)$ increases to infinity at an appropriate rate. In [21] the consistency of sieve bootstrap for general statistics being estimators of parameters in vector autoregressive time series models is considered under assumptions which essentially imply the assumptions imposed on the time series models and the sieve method investigated in our paper (see the assumptions (LA) in Remark 2.1). However, the characterization of asymptotic behavior of parametric estimators obtained by the authors does not cover the problem of asymptotics of VAR sieve bootstrap for predictors and bootstrap prediction regions considered in this article.

In the case of univariate causal linear time series models admitting the $AR(\infty)$ representation, Alonso et al. [1] constructed the sieve bootstrap estimator $X^\star_{T+h}$ of the future value X_{T+h}. They proved that $X^\star_{T+h} \to^{d^\star} X_{T+h}$ in probability, which implies that the bootstrap distribution $F^\star_{X^\star_{T+h}}$ approximates in probability the unknown distribution of $F_{X_{T+h}}$. Further, using the quantiles $Q^\star()$ of the distribution $F^\star_{X^\star_{T+h}}$ as bootstrap estimators of quantiles $Q()$ of the distribution $F_{X_{T+h}}$, the authors construct a prediction interval for the future value X_{T+h}. In general, the distribution $F^\star_{X^\star_{T+h}}$ and the quantiles $Q^\star()$ are not known. So, in simulations, the authors use the Monte Carlo method to approximate $F^\star_{X^\star_{T+h}}$ and $Q^\star()$. Unfortunately, they did not argue that the Monte Carlo approximation of $Q^\star()$ is a consistent estimator of the quantile $Q()$. At least, one should mention the results of Shi et al. [26].

Since the sieve approximation $p(n)$ is charged with serious error as an estimator of possibly finite but unknown order p of the considered univariate time series

model, Alonso et al. [2] joined the sieve bootstrap and the moving block bootstrap to select the order $p^\star$ and to introduce model uncertainty in procedures of resampling. Unfortunately, no proofs of consistency are given and the simulation results are restricted to the Gaussian errors.

In this article, we construct sieve bootstrap prediction regions for causal linear, invertible ($VAR(\infty)$) multivariate time series, approximating the prediction error by bootstrap replications of the prediction error. There are two ways (described in Section 4.2) of the bootstrap replicating of the prediction error. One can see that, in the univariate case, prediction hybrid bootstrap intervals constructed by the first method are identical with bootstrap prediction intervals constructed in [1] but it does not happen when we construct bootstrap-t (studentized) prediction intervals.

In this work, we generalize the results obtained in [13] from $VAR(p)$ models with finite but unknown order p to some class of second order stationary multivariate linear, $VAR(\infty)$ time series models.

It is worth noting that we have constructed consistent unconditional bootstrap prediction regions and the results allow us to use these prediction regions as effective and useful tools for testing and selection of models. Moreover, it follows from the theorems proved in the present article that the constructed bootstrap prediction regions are good approximations of prediction regions constructed on the base of unknown optimal linear predictors.

2. MODEL AND ASSUMPTIONS

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a second order stationary k-dimensional vector process with mean $EX_t = 0$ and the autocovariance function $\gamma(j) = EX_{t+j}X_t^T$, where $\mathbb{Z}$ is the set of all integers. We assume also that the process $\{X_t\}_{t\in\mathbb{Z}}$ is purely stochastic. Thus, using Wold's Decomposition Theorem (see [3] or [20]), we can represent $\{X_t\}_{t\in\mathbb{Z}}$ as an infinite vector moving average process $VMA(\infty)$,

$$(2.1) \qquad\qquad X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = I_k,$$

where $\sum_{j=0}^{\infty} \|\psi_j\|^2 < \infty$ and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a vector white noise process with the covariance matrix $E\epsilon_t\epsilon_t^T = \Sigma$ (Σ is invertible). Additionally, we assume that the process $\{X_t\}_{t\in\mathbb{Z}}$ is invertible. Thus, it can be represented as an infinite vector autoregressive process $VAR(\infty)$:

$$(2.2) \qquad\qquad X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \epsilon_t,$$

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

Further, we will use the following assumptions:

(A1) $X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = I_k$, $t \in \mathbb{Z}$, where

$$\{\epsilon_t\}_{t \in \mathbb{Z}} = \{(\epsilon_{t,1}, \ldots, \epsilon_{t,i}, \ldots, \epsilon_{t,k})^T\}_{t \in \mathbb{Z}}$$

is an i.i.d. sequence such that $E\epsilon_t = 0$, $E\epsilon_t\epsilon_t^T = \Sigma$ and $E|\epsilon_{t,i}\epsilon_{t,j}\epsilon_{t,l}\epsilon_{t,s}| < \infty$ for $i, j, l, s = 1, \ldots, k$ and all t.

(A2) $\big|\det\big(\Psi(z)\big)\big| = \big|\det\big(\sum_{j=0}^{\infty} \psi_j z^j\big)\big| > 0$ for $|z| \leqslant 1$ and $\sum_{j=0}^{\infty} j^r \|\psi_j\| < \infty$ for some $r \in \mathbb{N}$.

(B) $p = p(n) \to \infty$ as $n \to \infty$ and $\widehat{\Phi}_p = [\widehat{\phi}_{1,n}, \ldots, \widehat{\phi}_{p,n}]^T$ satisfies the Yule–Walker equations, i.e.

(2.3) $$\widehat{\Gamma}_p \widehat{\Phi}_p = \widehat{\gamma}_p,$$

where $\widehat{\Gamma}_p = [\widehat{\gamma}(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p = [\widehat{\gamma}(1), \ldots, \widehat{\gamma}(p)]^T$, and $\widehat{\gamma}(\cdot)$ is the sample auto-covariance function,

$$\widehat{\gamma}(j) = \frac{1}{n} \sum_{t=1}^{n-j} (X_{t+j} - \overline{X}_n)(X_t - \overline{X}_n)^T, \quad \widehat{\gamma}(-j) = \widehat{\gamma}^T(j),$$

where $\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t$.

REMARK 2.1. *In the sequel, the following list of assumptions, called* (LA), *will be also imposed*:
- *assumption* (A1),
- *assumption* (A2) *with* $r \geqslant 1$,
- *assumption* (B) *with* $p(n) = o\big((n/\log n)^{1/(2r+2)}\big)$ *and* $r \geqslant 1$.

Under the assumptions (LA) we prove the main theorems on bootstrap consistency.

3. SIEVE BOOTSTRAP ALGORITHM

The invertibility of the process $\{X_t\}_{t \in \mathbb{Z}}$ implies the $VAR(\infty)$ representation of the process which is crucial in the idea of the sieve bootstrap (see e.g. [9]). The sieve bootstrap algorithm uses the idea of the Grenander method of sieve (see [14]). Namely, we approximate $VAR(\infty)$ given by (2.2) by the sequence of $VAR(p)$ models, where $p = p(n)$ is a sequence growing to infinity sufficiently slow with the sample size n (assumption (B)).

REMARK 3.1. *It is also possible to approximate the process* $\{X_t\}_{t \in \mathbb{Z}}$ *given by* (2.1) *and construct sieve as a sequence of finite* $VMA(q)$, *where* $q = q(n) \to \infty$ (see [8]).

Let $X_1, \ldots, X_n$ be the observations of the process $\{X_t\}_{t \in \mathbb{Z}}$. We describe the sieve bootstrap algorithm in the following steps.

S t e p 1. Choose the approximation order $p = p(n)$ using FPE, the final prediction error (see [20]),

$$FPE(p) = \left(\frac{n + pk + 1}{n - pk - 1} \right)^k \det(\widehat{\Sigma}_p),$$

where $\widehat{\Sigma}_p$ is an estimator of the covariance matrix Σ of the white noise $\{\epsilon_t\}_{t \in \mathbb{Z}}$ in the model $VAR(p)$. Further, we choose $p = p(n)$, which minimizes $FPE(p)$ and $p \in \{p_{\min}(n), \ldots, p_{\max}(n)\}$, where $p_{\min} = \log_{10} n$, $p_{\max} = 10 \log_{10} n$.

S t e p 2. Estimate the coefficients $\phi_{1,n}, \ldots, \phi_{p,n}$ of the model $VAR(p)$ using the Yule–Walker method (assumption (B)) and obtain estimators $\widehat{\phi}_{1,n}, \ldots, \widehat{\phi}_{p,n}$.

S t e p 3. Compute the residuals

$$\widehat{\epsilon}_t = X_t - \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j}, \quad t = p+1, \ldots, n.$$

S t e p 4. Center the residuals

$$\widetilde{\epsilon}_t = \widehat{\epsilon}_t - \frac{1}{n-p} \sum_{j=p+1}^{n} \widehat{\epsilon}_t, \quad t = p+1, \ldots, n,$$

and draw bootstrap residuals ϵ_t^* from the empirical cumulative distribution function $\widehat{F}_{\epsilon,n}$, where

(3.1) $$\widehat{F}_{\epsilon,n}(x) = \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{1}\{\widetilde{\epsilon}_t \leqslant x\},$$

and $\leqslant$ denotes the relation of product order (partial order) in $\mathbb{R}^k$.

S t e p 5. Define bootstrap replications $X_1^*, \ldots, X_n^*$ of $X_1, \ldots, X_n$ by

$$X_t^* = \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j}^* + \epsilon_t^*, \quad t = 1, \ldots, n.$$

In practice, we can generate replications $X_1^*, \ldots, X_n^*$ starting the recursion from some initial values, e.g. $X_t^* = \epsilon_t^*, t = 0, -1, \ldots, -p+1$.

S t e p 6. Generate bootstrap replications $\widehat{\phi}_{j,n}^*$ of the Yule–Walker estimators $\widehat{\phi}_{j,n}$, where $j = 1, 2, \ldots, p(n)$.

The bootstrap construction induces a conditional probability measure P^* given the sample $X_1, \ldots, X_n$. In the sequel, all quantities with respect to P^* will be endowed with asterisk *.

REMARK 3.2. *The FPE criterion, used in Step* 1, *is asymptotically equivalent to the AIC criterion* (see [20]) *used in* [9].

REMARK 3.3. *In Step* 2, *the Yule–Walker estimators can be computed by using Whittle's algorithm, which is a multivariate version of the Durbin–Levinson algorithm* (see [7]).

In the sequel, we will use the following notation for the coefficients of autoregressive models and their related moving average models. For the stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ given by (2.1) or (2.2) the autoregressive coefficients ϕ_j and the moving average coefficients ψ_j are related by

$$\Phi(z) = I_k - \sum_{j=1}^{\infty} \phi_j z^j, \quad \Phi^{-1}(z) = \Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j,$$

$$\psi_0 = I_k, \quad \text{and} \quad \psi_j = \sum_{i=1}^{j} \phi_i \psi_{j-i}, \quad j = 1, 2, \ldots$$

In the model $VAR(p)$, approximating the model (2.2), we denote by $\phi_{j,n}$ the autoregressive coefficients which fulfill the theoretical Yule–Walker equations (2.3) (assumption (B), where the sample autocovariance function is replaced by $\gamma(j) = EX_{t+j}X_t^T$), and by $\psi_{j,n}$ the moving average coefficients:

$$(3.2) \qquad \Phi_n(z) = I_k - \sum_{j=1}^{p} \phi_{j,n} z^j, \quad \Phi_n^{-1}(z) = \Psi_n(z) = \sum_{j=0}^{\infty} \psi_{j,n} z^j,$$

$$\psi_{0,n} = I_k, \quad \text{and} \quad \psi_{j,n} = \sum_{i=1}^{j} \phi_{i,n} \psi_{j-i,n}, \quad j = 1, 2, \ldots$$

We write the Yule–Walker estimators given in assumption (B):

$$(3.3) \qquad \widehat{\Phi}_n(z) = I_k - \sum_{j=1}^{p} \widehat{\phi}_{j,n} z^j, \quad \widehat{\Phi}_n^{-1}(z) = \widehat{\Psi}_n(z) = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} z^j,$$

$$\widehat{\psi}_{0,n} = I_k, \quad \text{and} \quad \widehat{\psi}_{j,n} = \sum_{i=1}^{j} \widehat{\phi}_{i,n} \widehat{\psi}_{j-i,n}, \quad j = 1, 2, \ldots$$

Let $\widehat{\phi}^*_{j,n}$ and $\widehat{\psi}^*_{j,n}$ be the bootstrap replications of Yule–Walker estimators $\widehat{\phi}_{j,n}$, $\widehat{\psi}_{j,n}$,

$$(3.4) \qquad \widehat{\Phi}^*_n(z) = I_k - \sum_{j=1}^{p} \widehat{\phi}^*_{j,n} z^j, \quad \widehat{\Phi}^{*-1}_n(z) = \widehat{\Psi}^*_n(z) = \sum_{j=0}^{\infty} \widehat{\psi}^*_{j,n} z^j,$$

where $\widehat{\psi}^*_{0,n} = I_k$.

REMARK 3.4. *The assumptions* (A1) *and* (A2) *provide us with the correctness of defined functions* $\Psi(z)$ *and* $\Phi(z)$, *which means that the series* $\sum_{j=0}^{\infty} \psi_j z^j$ *and* $\sum_{j=1}^{\infty} \phi_j z^j$ *are convergent for* $|z| \leqslant 1$ *and* $\Phi^{-1}(z) = \Psi(z)$. *The properties of the Yule–Walker equations ensure the correctness of defined functions* $\Psi_n(z)$, $\widehat{\Psi}_n(z)$ *and* $\widehat{\Psi}_n^*(z)$ (*solutions of the Yule–Walker equations give us a causal model, see* [7]).

4. PREDICTION REGIONS

4.1. Linear predictors and Gaussian prediction regions. Forecasting the future values X_{n+h}, $h = 1, 2, \ldots$, is a very common task in the statistical analysis of time series. For the second order stationary process we can construct the best linear predictor, in the mean square sense, as an orthogonal projection of X_{n+h} onto a closed subspace $\overline{\mathrm{sp}}\{X_1, \ldots, X_n\}$ of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ (see [20]). We can represent the h-step predictor as

$$\mathrm{Proj}_n X_{n+h} = \mathrm{Proj}_{\overline{\mathrm{sp}}\{X_1, \ldots, X_n\}} X_{n+h}$$

and the mean squared prediction error as

(4.1) $$\Sigma_X(h) = E\left(X_{n+h} - \mathrm{Proj}_n X_{n+h}\right)\left(X_{n+h} - \mathrm{Proj}_n X_{n+h}\right)^T.$$

Assuming that (A1) holds, we have also

(4.2) $$\Sigma_X(h) = \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T.$$

The mean squared prediction error can be obtained by using the multivariate version of the innovations algorithm (see [7]).

REMARK 4.1. *For the autoregressive model* $VAR\big(p = p(n)\big)$,

$$X_t = \sum_{j=1}^{p} \phi_{j,n} X_{t-j} + \epsilon_{t,n},$$

approximating the model $VAR(\infty)$, *the best linear predictor* (*in the mean square sense*) *and the mean squared error matrix have the forms*

$$\mathrm{Proj}_n X_{n+h, VAR(p)} = \sum_{j=1}^{p} \phi_{j,n} \mathrm{Proj}_n X_{n+h-j, VAR(p)},$$

$$\Sigma_{n,X}(h) = \sum_{j=0}^{h-1} \psi_{j,n} \Sigma \psi_{j,n}^T,$$

where $\mathrm{Proj}_n X_{n+j, VAR(p)} = X_{n+j}$ *for* $j \leqslant 0$, $\Sigma = E\epsilon_t \epsilon_t^T$ *and coefficients* $\psi_{j,n}$ *are given by the recursive equations*

$$\psi_{0,n} = I_k \quad \text{and} \quad \psi_{j,n} = \sum_{i=1}^{j} \phi_{i,n} \psi_{(j-i),n}, \quad j = 1, 2, \ldots$$

From the continuity of the projection operator in L^2 it follows that

$$\mathrm{Proj}_n X_{n+h} - \mathrm{Proj}_n X_{n+h,VAR(p)} \xrightarrow{L^2} 0 \quad and \quad \Sigma_{n,X}(h) \to \Sigma_X(h) \ as \ n \to \infty.$$

Since the form of the best linear predictor (in the mean square sense) and the mean squared error matrix depend on unknown parameters of a model considered, we will use an estimator $\widehat{\mathrm{Proj}}_n X_{n+h} = \widehat{X}_{n+h}$ of the best linear predictor $\mathrm{Proj}_n X_{n+h}$ with appropriately chosen estimators of parameters of the model.

The most common method of constructing the prediction regions for X_{n+h}, $h = 1, 2, \ldots$, is the Box–Jenkins method. This method assumes that the prediction error has at least asymptotically normal distribution

$$X_{n+h} - \widehat{X}_{n+h} \sim \mathcal{N}\big(0, \Sigma_X(h)\big).$$

Thus, the quadratic form below has at least the asymptotically χ-square distribution with k degrees of freedom:

$$(X_{n+h} - \widehat{X}_{n+h})^T \Sigma_X^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \sim \chi^2(k).$$

So, the prediction region for X_{n+h}, $h = 1, 2, \ldots$, with nominal confidence level $1 - \alpha$ has a shape of k-dimensional ellipse

$$(4.3) \quad \mathcal{E}_{B-J}(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T \Sigma_X^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant \chi^2_{1-\alpha}(k)\},$$

where $\chi^2_\alpha(k)$ is an α quantile of χ-square distribution with k degrees of freedom. We can also use the Bonferroni inequality and construct a k-dimensional prediction cube

$$(4.4) \quad I_{B-J}(h) = \{X_{n+h,j} \in [\widehat{X}_{n+h,j} + \sigma_{X,j}(h)z_{\alpha/(2k)},$$
$$\widehat{X}_{n+h,j} + \sigma_{X,j}(h)z_{1-\alpha/(2k)}], \ j = 1, \ldots, k\},$$

where z_α is an α quantile of the normal distribution $\mathcal{N}(0,1)$ and $\sigma_{X,j}(h)$ is a square root of the jth diagonal element of the mean squared error matrix $\Sigma_X(h)$.

4.2. Bootstrap prediction regions. The bootstrap methods are very common in the problem of constructing the confidence intervals (see e.g. [12]). The same idea can be used in the construction of prediction regions for X_{n+h}, $h = 1, 2, \ldots$ Namely, we may apply the sieve bootstrap method and generate bootstrap replications $X_1^*, \ldots, X_n^*$ using observations $X_1, \ldots, X_n$. Then, we construct the bootstrap replications of the future observations on the base of the $VAR(p)$ approximation

$$(4.5) \quad \text{for } h = 1\colon \quad X_{n+1}^* = \sum_{j=1}^p \widehat{\phi}_{j,n}^* X_{n+1-j} + \epsilon_{n+1}^*,$$

$$\text{for } h > 1\colon \quad X_{n+h}^* = \sum_{j=1}^{h-1} \widehat{\phi}_{j,n}^* X_{n+h-j}^* + \sum_{j=h}^p \widehat{\phi}_{j,n}^* X_{n+h-j} + \epsilon_{n+h}^*,$$

where $\widehat{\phi}^*_{j,n}$ is a bootstrap replication of the estimator $\widehat{\phi}_{j,n}$. It is worth noting that we applied a modification of a standard procedure, proposed by Bühlmann [9] for the one-dimensional case, to generate the future bootstrap observations in the equation (4.5). Namely, we generated future observations starting recursion from $X_1, \ldots, X_n$, in contrast to the standard method in which the recursion is started from the bootstrap replications $X_1^*, \ldots, X_n^*$. The results of the simulations showed that this type of modification improved the empirical probability of coverage of the bootstrap prediction regions.

Using the $VAR(p)$ approximation, we construct the estimator of the best linear predictor

$$(4.6) \qquad \widehat{X}_{n+h} = \sum_{j=1}^{p(n)} \widehat{\phi}_{j,n} \widehat{X}_{n+h-j},$$

where $\widehat{X}_{n+j} = X_{n+j}, j \leqslant 0$.

We construct the bootstrap prediction regions using two methods: hybrid bootstrap and bootstrap-t.

In the hybrid bootstrap, the unknown distribution of the prediction error

$$(4.7) \qquad H_n(h) = X_{n+h} - \widehat{X}_{n+h}$$

can be approximated by two bootstrap variants:

$$(4.8) \qquad H_n^*(h) = X_{n+h}^* - \widehat{X}_{n+h},$$

$$(4.9) \qquad \widetilde{H}_n^*(h) = X_{n+h}^* - \widehat{X}_{n+h}^*,$$

where $\widehat{X}_{n+h}^*$ is the bootstrap replication of $\widehat{X}_{n+h}$ given by (4.6). It can be proved that both bootstrap variants of the prediction error (4.7) are consistent. However, we will focus on (4.8) because of its good simulation results. Since the proof of consistency for the bootstrap variant (4.9) goes along the same lines as the proof of consistency for the bootstrap variant (4.8), we will omit it. Thus, the bootstrap prediction cube, constructed by using the Bonferroni inequality, has the form

$$(4.10) \quad I_B(h) = \{ X_{n+h,j} \in [\widehat{X}_{n+h,j} + q_{\alpha/(2k),j}^*,$$
$$\widehat{X}_{n+h,j} + q_{1-\alpha/(2k),j}^*], \; j = 1, \ldots, k \},$$

where $q_{\alpha,j}^*$ is an α quantile of the distribution $X_{n+h,j}^* - \widehat{X}_{n+h,j}, j = 1, \ldots, k$.

Using the hybrid bootstrap, we can also create a bootstrap prediction region in the shape of k-dimensional ellipse:

$$(4.11) \qquad \mathcal{E}_B(h) = \{ (X_{n+h} - \widehat{X}_{n+h})^T (X_{n+h} - \widehat{X}_{n+h}) \leqslant q_{1-\alpha}^* \},$$

where q_α^* is an α quantile of the distribution $\| H_n^*(h) \|^2$.

In the bootstrap-t method we approximate the studentized unknown distribution of the prediction error

$$(4.12) \qquad T_n(h) = \widehat{\Sigma}_{n,X}^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h})$$

by its bootstrap replication

$$(4.13) \qquad T_n^*(h) = \widehat{\Sigma}_{n,X}^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}),$$

where $\widehat{\Sigma}_{n,X}(h)$ is an estimator of the mean squared error matrix $\Sigma_X(h)$, and $\widehat{\Sigma}_X^*(h)$ is a bootstrap replication of the estimator. By Remark 4.1, we define the estimator $\widehat{\Sigma}_{n,X}(h)$ as

$$(4.14) \qquad \widehat{\Sigma}_{n,X}(h) = \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T,$$

where $\widehat{\psi}_{j,n} = \sum_{i=1}^{j} \widehat{\phi}_{i,n} \widehat{\psi}_{(j-i),n}$. Thus, the bootstrap prediction cube, constructed by using the Bonferroni inequality, has the form

$$(4.15) \quad I_{B-t}(h) = \{ X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h) t_{\alpha/(2k),j}^*,$$
$$\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h) t_{1-\alpha/(2k),j}^*], \ j = 1, \ldots, k \},$$

where $t_{\alpha,j}^*$ is an α quantile of the distribution $\widehat{\sigma}_{X,j}^{*-1}(h)(X_{n+h,j}^* - \widehat{X}_{n+h})$ for $j = 1, \ldots, k$.

Using the bootstrap-t, we can also create a bootstrap prediction region in the shape of k-dimensional ellipse:

$$(4.16) \qquad \mathcal{E}_{B-t}(h) = \{ (X_{n+h} - \widehat{X}_{n+h})^T \widehat{\Sigma}_{n,X}^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant t_{1-\alpha}^* \},$$

where t_α^* is an α quantile of the distribution $\|T_n^*(h)\|^2$.

4.3. Bootstrap simultaneous prediction regions based on extreme statistics. In this subsection we present different types of bootstrap confidence regions based on the distribution of minimum and maximum of $X_{n+h} - \widehat{X}_{n+h}$. More precisely, we investigate simultaneous hybrid and studentized confidence regions. For each type we propose left-sided, right-sided and both-sided regions. First, we introduce the following notation for statistics U_{n+h}, V_{n+h} and R_{n+h}:

$$(4.17) \qquad U_{n+h} = \min_{1 \leqslant i \leqslant k} W_{n+h,i} = \min_{1 \leqslant i \leqslant k} (X_{n+h,i} - \widehat{X}_{n+h,i}),$$

$$(4.18) \qquad V_{n+h} = \max_{1 \leqslant i \leqslant k} W_{n+h,i} = \max_{1 \leqslant i \leqslant k} (X_{n+h,i} - \widehat{X}_{n+h,i}),$$

$$(4.19) \qquad R_{n+h} = \max_{1 \leqslant i \leqslant k} |W_{n+h,i}| = \max_{1 \leqslant i \leqslant k} |X_{n+h,i} - \widehat{X}_{n+h,i}|,$$

and for their bootstrap versions:

$$(4.20) \qquad U_{n+h}^* = \min_{1 \leqslant i \leqslant k} W_{n+h,i}^* = \min_{1 \leqslant i \leqslant k} (X_{n+h,i}^* - \widehat{X}_{n+h,i}),$$

$$(4.21) \qquad V_{n+h}^* = \max_{1 \leqslant i \leqslant k} W_{n+h,i}^* = \max_{1 \leqslant i \leqslant k} (X_{n+h,i}^* - \widehat{X}_{n+h,i}),$$

$$(4.22) \qquad R_{n+h}^* = \max_{1 \leqslant i \leqslant k} |W_{n+h,i}^*| = \max_{1 \leqslant i \leqslant k} |X_{n+h,i}^* - \widehat{X}_{n+h,i}|.$$

Let $\widehat{u}_{n+h,(\cdot)}^*$, $\widehat{v}_{n+h,(\cdot)}^*$ and $\widehat{r}_{n+h,(\cdot)}^*$ be Monte Carlo estimators of quantiles of bootstrap distributions of U_{n+h}^*, V_{n+h}^* and R_{n+h}^*. Then, the bootstrap prediction cubes, constructed by using extreme statistics, can be written in the following form:

$$(4.23) \qquad I_B^{UV}(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}_{n+h,\alpha/2}^*,$$
$$\widehat{X}_{n+h,i} + \widehat{v}_{n+h,1-\alpha/2}^*], \ i = 1, \dots, k\},$$

$$(4.24) \qquad I_B^U(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}_{n+h,\alpha}^*, +\infty), \ i = 1, \dots, k\},$$

$$(4.25) \qquad I_B^V(h) = \{X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{v}_{n+h,1-\alpha}^*], \ i = 1, \dots, k\},$$

$$(4.26) \qquad I_B^R(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{r}_{n+h,1-\alpha}^*,$$
$$\widehat{X}_{n+h,i} + \widehat{r}_{n+h,1-\alpha}^*], \ i = 1, \dots, k\}.$$

We can also construct studentized prediction cubes:

$$(4.27) \ I_{B-t}^{UV}(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}_{n+h,\alpha/2}^*\widehat{\sigma}_{n+h,i},$$
$$\widehat{X}_{n+h,i} + \widehat{vs}_{n+h,1-\alpha/2}^*\widehat{\sigma}_{n+h,i}], \ i = 1, \dots, k\},$$

$$(4.28) \ I_{B-t}^U(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}_{n+h,\alpha}^*\widehat{\sigma}_{n+h,i}, +\infty), i = 1, \dots, k\},$$

$$(4.29) \ I_{B-t}^V(h) = \{X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{vs}_{n+h,1-\alpha}^*\widehat{\sigma}_{n+h,i}], i = 1, \dots, k\},$$

$$(4.30) \ I_{B-t}^R(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{rs}_{n+h,1-\alpha}^*\widehat{\sigma}_{n+h,i},$$
$$\widehat{X}_{n+h,i} + \widehat{rs}_{n+h,1-\alpha}^*\widehat{\sigma}_{n+h,i}], \ i = 1, \dots, k\},$$

where $\widehat{us}_{n+h,(\cdot)}^*$, $\widehat{vs}_{n+h,(\cdot)}^*$ and $\widehat{rs}_{n+h,(\cdot)}^*$ are, respectively, Monte Carlo estimators of quantiles of bootstrap distributions of

$$US_{n+h}^* = \min_{1 \leqslant i \leqslant k} W_{n+h,i}^* = \min_{1 \leqslant i \leqslant k} \left(\frac{X_{n+h,i}^* - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}^*} \right),$$

$$VS_{n+h}^* = \max_{1 \leqslant i \leqslant k} W_{n+h,i}^* = \max_{1 \leqslant i \leqslant k} \left(\frac{X_{n+h,i}^* - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}^*} \right),$$

$$RS_{n+h}^* = \max_{1 \leqslant i \leqslant k} |W_{n+h,i}^*| = \max_{1 \leqslant i \leqslant k} \left| \frac{X_{n+h,i}^* - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}^*} \right|,$$

and US^*_{n+h}, VS^*_{n+h} and RS^*_{n+h} are the following bootstrap versions of distributions:

$$US_{n+h} = \min_{1 \leqslant i \leqslant k} W_{n+h,i} = \min_{1 \leqslant i \leqslant k} \left(\frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}} \right),$$

$$VS_{n+h} = \max_{1 \leqslant i \leqslant k} W_{n+h,i} = \max_{1 \leqslant i \leqslant k} \left(\frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}} \right),$$

$$RS_{n+h} = \max_{1 \leqslant i \leqslant k} |W_{n+h,i}| = \max_{1 \leqslant i \leqslant k} \left| \frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}} \right|.$$

5. CONSISTENCY OF BOOTSTRAP METHODS

5.1. Representation of the $VAR(\infty)$ model by the moving average model. In this subsection we use the following theorem proved in [15] (see Theorem 7.4.2). The theorem is formulated below but with changed notation, adapted to the present article.

THEOREM 5.1 ([15], Theorem 7.4.2). *If* $\det\big(\Psi(z)\big) \neq 0$ *for* $|z| \leqslant 1$ *and*

$$\sum_{j=0}^{\infty} j^\lambda \|\psi_j\| < \infty, \quad \lambda > 0,$$

then for $\Phi(z) = \Psi^{-1}(z) = \sum_{j=1}^{\infty} \phi_j z^j$ *we have*

$$\sum_{j=0}^{\infty} j^\lambda \|\phi_j\| < \infty.$$

By analogical reasoning, we can prove the converse of Theorem 5.1. Namely, we have

COROLLARY 5.1. *The following assertions are equivalent*:
 (i) $\det\big(\Phi(z)\big) \neq 0$, $|z| \leqslant 1$, $\sum_{j=0}^{\infty} j^\lambda \|\phi_j\| < \infty$, $\lambda > 0$.
 (ii) $\det\big(\Psi(z)\big) \neq 0$, $|z| \leqslant 1$, $\sum_{j=0}^{\infty} j^\lambda \|\psi_j\| < \infty$, $\lambda > 0$.

The next lemma is a multivariate version of a lemma given by Bühlmann ([8], Lemma 2.2). For the purpose of this lemma (and only for this lemma) we assume that $\Phi_n(z) = I_k - \sum_{j=1}^{\infty} \phi_{j,n} z^j$ is some deterministic approximation of the function $\Phi(z) = I_k - \sum_{j=1}^{\infty} \phi_j z^j$ (model (2.2)) and we define $\Psi_n(z) = \Phi_n^{-1}(z) = \sum_{j=0}^{\infty} \psi_{j,n} z^j$, where $\psi_{0,n} = I_k$. The lemma gives us conditions under which the function $\Psi_n(z)$ is correctly defined and also asymptotical properties of the coefficients $\psi_{j,n}$.

LEMMA 5.1. *Let* (A2) *with* $r \geqslant 1$ *hold and* $\sum_{j=1}^{\infty} j^r \|\phi_{j,n} - \phi_j\| = o(1)$, *where* $n \to \infty$. *Then*:

(i) *There exists $n_0 \in \mathbb{N}$ such that*

$$\sup_{n \geqslant n_0} \sum_{j=1}^{\infty} j^r \|\phi_{j,n}\| < \infty \qquad and \qquad \inf_{n \geqslant n_0} \inf_{|z| \leqslant 1} \big| \det\big(\Phi_n(z)\big)\big| > 0.$$

(ii) *There exists $n_1 \in \mathbb{N}$ such that, for $n \geqslant n_1$, $\Psi_n(z)$ is absolutely convergent for $|z| \leqslant 1$ and*

$$\sup_{n \geqslant n_1} \sum_{j=0}^{\infty} j^r \|\psi_{j,n}\| < \infty.$$

P r o o f. (i) is an immediate consequence of the assumptions of the lemma and Corollary 5.1.

(ii) By the formula for the inversion of the matrix we have

$$\Psi_n(z) = \Phi_n^{-1}(z) = \frac{1}{\det\big(\Phi_n(z)\big)} \mathrm{adj}\big(\Phi_n(z)\big).$$

Denoting the element of the matrix by $\phi_{j,n} = [\phi_{(sv),j,n}]_{s,v=1}^k$, $j = 1, 2, \ldots$, we get

$$\Phi_n(z) = \Big[\sum_{j=1}^{\infty} \phi_{(sv),j,n} z^j \Big]_{s,v=1}^k.$$

From (i) we have $\det\big(\Phi_n(z)\big) \neq 0$ for $|z| \leqslant 1$, $n \geqslant n_0$ and

$$\sup_{n \geqslant n_0} \sum_{j=1}^{\infty} j^r |\phi_{(sv),j,n}| < \infty \quad \text{ for all } s, v = 1, \ldots, k.$$

For some $s, v, u, w = 1, \ldots, k$ we have

$$\sum_{j=1}^{\infty} \phi_{(sv),j,n} z^j \cdot \sum_{j=1}^{\infty} \phi_{(uw),j,n} z^j = \sum_{j=1}^{\infty} (\phi_{(sv)} * \phi_{(uw)})_{j,n} z^j,$$

$$(\phi_{(sv)} * \phi_{(uw)})_{j,n} = \sum_{l=1}^{\infty} \phi_{(sv),l,n} \phi_{(uw),j-l,n}.$$

Using properties of the convolution, we get

$$\sum_{j=1}^{\infty} j^r |(\phi_{(sv)} * \phi_{(uw)})_{j,n}| \leqslant \sum_{j=1}^{\infty} j^r |\phi_{(sv),j,n}| \sum_{j=1}^{\infty} j^r |\phi_{(uw),j,n}|,$$

and further

$$\sup_{n \geqslant n_0} \sum_{j=1}^{\infty} j^r |(\phi_{(st)} * \phi_{(uw)})_{j,n}| < \infty.$$

In consequence, we infer that the coefficients of the determinant $\det\big(\Phi_n(z)\big)$ and the coefficients of all elements of the matrix $\mathrm{adj}\big(\Phi_n(z)\big)$ fulfill

$$\det\big(\Phi_n(z)\big) = \sum_{j=1}^{\infty} \check{\phi}_{j,n} z^j, \quad \sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\check{\phi}_{j,n}| < \infty,$$

$$\mathrm{adj}\big(\Phi_n(z)\big) = [\phi^0_{(sv),n}(z)]^k_{s,v=1}, \quad \phi^0_{(sv),n}(z) = \sum_{j=1}^{\infty} \phi^0_{(sv),j,n} z^j,$$

$$\sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\phi^0_{(sv),j,n}| < \infty \quad \text{for all } s,v = 1,\ldots,k.$$

To complete the proof, we have to show that $1/\det\big(\Phi_n(z)\big)$ is an analytical function for $|z| \leqslant 1$ with coefficients $\underline{\phi}_{j,n}$ satisfying

$$\frac{1}{\det\big(\Phi_n(z)\big)} = \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j, \quad \sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\underline{\phi}_{j,n}| < \infty.$$

However, this is a consequence of the Wiener theorem (see [30]) and the lemma given by Bühlmann [8] for the one-dimensional case. Finally, we get

$$\Psi_n(z) = \frac{1}{\det\big(\Phi_n(z)\big)} \mathrm{adj}\big(\Phi_n(z)\big) = \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j [\phi^0_{(sv),n}(z)]^k_{s,v=1}$$

$$= \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j \Big[\sum_{j=1}^{\infty} \phi^0_{(sv),j,n} z^j\Big]^k_{s,v=1} = \sum_{j=0}^{\infty} \psi_{j,n} z^j$$

and

$$\sup_{n\geqslant n_1} \sum_{j=0}^{\infty} j^r \|\psi_{j,n}\| < \infty. \quad \blacksquare$$

The next two theorems give us properties of the coefficients $\widehat{\psi}_{j,n}$ from the representation (3.3), in which $\widehat{\Phi}_n(z)$ replaces the function $\Phi_n(z)$ in Lemma 5.1. These two theorems are multivariate versions of the theorems given by Bühlmann for the one-dimensional case ([8], Theorems 3.1 and 3.2).

THEOREM 5.2. *Let* (LA) *hold. Then there exists a random variable* n_1 *for which*

$$\sup_{n\geqslant n_1} \sum_{j=0}^{\infty} j^r \|\widehat{\psi}_{j,n}\| < \infty \text{ almost surely.}$$

P r o o f. We will use Lemma 5.1. Thus, we have to check its assumption:

$$\sum_{j=1}^{\infty} j^r \|\widehat{\phi}_{j,n} - \phi_j\| \leqslant \sum_{j=1}^{p} j^r \|\widehat{\phi}_{j,n} - \phi_{j,n}\| + \sum_{j=1}^{p} j^r \|\phi_{j,n} - \phi_j\| + \sum_{j=p+1}^{\infty} j^r \|\phi_j\|$$

$$= S_1 + S_2 + S_3.$$

By Theorem 2.1 from [16] we obtain $S_1 = o(1)$ almost surely. Under the assumption (A2), using the Baxter inequality ([15], Theorem 6.6.12 and p. 271) and Corollary 5.1, we get $S_2 = o(1)$. Similarly we show that $S_3 = o(1)$. Finally, we get

$$\sum_{j=1}^{\infty} j^r \|\widehat{\phi}_{j,n} - \phi_j\| = o(1) \text{ almost surely.}$$

We complete the proof using Lemma 5.1. ∎

THEOREM 5.3. *Let* (LA) *hold. Then*

$$\sup_{1 \leqslant j < \infty} \|\widehat{\psi}_{j,n} - \psi_j\| = O\big((\log n/n)^{1/2}\big) + O(p^{-r}) \text{ almost surely.}$$

P r o o f. We start from showing that

(5.1) $$\Sigma - \widehat{\Sigma} = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}).$$

To prove it, we use Theorem 2.1 from [16] and the Hannan and Deistler bound ([15], Theorem 7.4.3). We get

(5.2) $$\max_{0 \leqslant j < \infty} \|\widehat{\gamma}(j) - \gamma(j)\| = O_{a.s.}\big((\log n/n)^{1/2}\big),$$

where $\widehat{\gamma}(j) = 0$ for $|j| \geqslant n$. From the Yule–Walker equations for the models (2.2) and (3.3) we have

$$\Sigma = \gamma(0) - \sum_{j=1}^{\infty} \phi_j \gamma^T(j), \quad \widehat{\Sigma} = \widehat{\gamma}(0) - \sum_{j=1}^{p} \widehat{\phi}_{j,n} \widehat{\gamma}^T(j).$$

Therefore,

$$\|\Sigma - \widehat{\Sigma}\| \leqslant \|\gamma(0) - \widehat{\gamma}(0)\| + \Big\| \sum_{j=1}^{p} \big(\widehat{\phi}_{j,n} \widehat{\gamma}^T(j) - \phi_j \gamma^T(j)\big) \Big\| + \Big\| \sum_{j=p+1}^{\infty} \phi_j \gamma^T(j) \Big\|$$

$$\leqslant \|\gamma(0) - \widehat{\gamma}(0)\|$$

$$+ \max_{1 \leqslant j \leqslant p} \|\widehat{\phi}_{j,n} - \phi_j\| \big(p \max_{1 \leqslant j \leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| + \sum_{j=1}^{p} \|\gamma(j)\| \big)$$

$$+ \max_{1 \leqslant j \leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| \sum_{j=1}^{\infty} \|\phi_j\| + \|[EX_{t,i}^2 EX_{t,j}^2]_{i,j=1}^{k}\| \sum_{j=p+1}^{\infty} \|\phi_j\|$$

$$= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r})$$

and we get (5.1).

Following the ideas of the proof of Theorem 3.2 from [8] for the one-dimensional case, let $\{Y_t\}_{t\in\mathbb{Z}}$ be a conditional process (given $X_1,\ldots,X_n$) in the form

$$(5.3) \qquad\qquad Y_t = \sum_{j=1}^{p} \widehat{\phi}_{j,n} Y_{t-j} + \eta_t,$$

where $\{\eta_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random vectors and $E\eta_t = 0$, $E\eta_t\eta_t^T = \Sigma$. Additionally, we assume that η_t and ϵ_t are independent. According to properties of the Yule–Walker estimators, $\{Y_t\}_{t\in\mathbb{Z}}$ is a causal process

$$(5.4) \qquad\qquad Y_t = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} \eta_{t-j}.$$

Denote by E^Y and Cov^Y the conditional expectation and the autocovariance function of the process $\{Y_t\}_{t\in\mathbb{Z}}$ under the condition $X_1,\ldots,X_n$. Thus, using (5.3) and (5.4) for Y_{t+i}, we have

$$E^Y Y_{t+i}\eta_t^T = \widehat{\psi}_{i,n}\Sigma = \gamma_Y(i) - \sum_{j=1}^{p} \gamma_Y(i+j)\widehat{\phi}_{j,n}^T,$$

where $\gamma_Y(j) = \mathrm{Cov}^Y(Y_{t+j}, Y_t) = \sum_{l=0}^{\infty} \widehat{\psi}_{l+j,n}\Sigma\widehat{\psi}_{l,n}^T$. We get a similar equality for ψ_i:

$$\psi_i\Sigma = \gamma(i) - \sum_{j=1}^{\infty} \gamma(i+j)\phi_j^T.$$

Thus, for $i = 0, 1, 2, \ldots$ we have

$$\|\widehat{\psi}_{i,n} - \psi_i\| \leqslant \|\Sigma^{-1}\|\Big\|\gamma_Y(i) - \sum_{j=1}^{p}\gamma_Y(i+j)\widehat{\phi}_{j,n}^T - \gamma(i) + \sum_{j=1}^{\infty}\gamma(i+j)\phi_j^T\Big\|$$

$$\leqslant \|\Sigma^{-1}\|\Big\|\sum_{j=1}^{p}\gamma_Y(i+j)(\widehat{\phi}_{j,n}^T - \phi_j^T) - \sum_{j=p+1}^{\infty}\gamma_Y(i+j)\phi_j^T\Big\|$$

$$+ \|\Sigma^{-1}\|\Big\|\sum_{j=1}^{\infty}\big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_i^T\Big\| + \|\Sigma^{-1}\|\,\|\gamma_Y(i) - \gamma(i)\|$$

$$= \|\Sigma^{-1}\|(S_{1,i} + S_{2,i} + S_{3,i}).$$

We bound the components separately. We get

$$S_{1,i} \leqslant \sum_{j=1}^{p}\|\gamma_Y(i+j)\|\,\|\widehat{\phi}_{j,n} - \phi_j\| + \sum_{j=p+1}^{\infty}\|\gamma_Y(i+j)\|\,\|\phi_j\|$$

$$\leqslant \|\Sigma\| \max_{1\leqslant j\leqslant p}\|\widehat{\phi}_{j,n} - \phi_j\|\Big(\sum_{l=0}^{\infty}\|\widehat{\psi}_{l,n}\|\Big)^2 + \|\Sigma\|\Big(\sum_{l=0}^{\infty}\|\widehat{\psi}_{l,n}\|\Big)^2\sum_{j=p+1}^{\infty}\|\phi_j\|$$

$$= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}) \quad \text{for each } i \in \mathbb{N},$$

where the last approximation is obtained by using Theorem 2.1 from [16] and Theorem 5.2. Further, using the convention $\sum_{j=l}^{i} a_j = 0$ for $i < l$, we have

$$S_{2,i} \leqslant \Big\| \sum_{j=1}^{p-i} \big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_j^T \Big\| + \Big\| \sum_{j=p-i+1}^{\infty} \big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_j^T \Big\|$$

$$\leqslant \max_{1 \leqslant j \leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| \sum_{j=0}^{\infty} \|\phi_j\| + \|\Sigma - \widehat{\Sigma}\| \Big(\sum_{j=0}^{\infty} \|\widehat{\psi}_{j,n}\| \Big)^2 \sum_{j=0}^{\infty} \|\phi_j\|$$

$$+ \|\Sigma\| \sum_{j=0}^{\infty} \|\phi_j\| \Big(\sum_{j=0}^{\infty} \|\widehat{\psi}_{j,n}\| \sum_{j=p+1}^{\infty} \|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{\infty} \|\psi_j\| \sum_{j=p+1}^{\infty} \|\psi_j\| \Big)$$

$$= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}) \quad \text{for each } i \in \mathbb{N},$$

where the last approximation is due to (5.1), (5.2), Theorem 5.2 and the assumption (A2). The component $S_{3,i}$ can be bounded analogously to the component $S_{2,i}$ and we get $S_{3,i} = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r})$ for each $i \in \mathbb{N}$. Finally, we have

$$\sup_{0 \leqslant i} \|\widehat{\psi}_{i,n} - \psi_i\| = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}). \quad \blacksquare$$

5.2. Consistency of sieve bootstrap. From the construction of the bootstrap replications algorithm we have $E^*\epsilon_t^* = 0$. The following lemma gives us the convergence of the bootstrap covariance matrix $\Sigma^* = E^*\epsilon_t^*\epsilon_t^{*T}$ to the covariance matrix Σ of the white noise $\{\epsilon_t\}_{t \in \mathbb{Z}}$. In the one-dimensional case, this fact was proved in [9] (see Lemma 5.3).

LEMMA 5.2. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$E^*\epsilon_t^*\epsilon_t^{*T} = E\epsilon_t\epsilon_t^T + o_P(1).$$

P r o o f. Although the proof of this lemma is similar to the proof of Lemma 5.3 in [9] for the one-dimensional case, we will give it for completeness. Notice the following equality holds:

$$(5.5) \qquad \widehat{\epsilon}_{t,n} = X_t - \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j} = \epsilon_t + Q_{t,n} + R_{t,n},$$

where

$$Q_{t,n} = \sum_{j=1}^{p} (\phi_{j,n} - \widehat{\phi}_{j,n}) X_{t-j},$$

$$R_{t,n} = \sum_{j=1}^{\infty} (\phi_j - \phi_{j,n}) X_{t-j}, \qquad \phi_{j,n} = 0 \text{ for } j > p.$$

The bootstrap covariance matrix has the form

$$E^* \epsilon_t^* \epsilon_t^{*T} = \frac{1}{n-p} \sum_{t=p+1}^{n} \left(\widehat{\epsilon}_{t,n} - \bar{\epsilon}_n\right)\left(\widehat{\epsilon}_{t,n} - \bar{\epsilon}_n\right)^T,$$

where $\bar{\epsilon}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\epsilon}_{t,n}$. First, we show that $\bar{\epsilon}_n = o_P(1)$. We have

$$\bar{\epsilon}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\epsilon}_{t,n} = \frac{1}{n-p} \sum_{t=p+1}^{n} (\epsilon_t + Q_{t,n} + R_{t,n}) = S_1 + S_2 + S_3.$$

From the assumption (A1) and the Markov inequality we get $S_1 = O_P(n^{-1/2})$. Using the Cauchy–Schwarz inequality and the Jensen inequality, we obtain

$$\|S_2\| = \left\| \frac{1}{n-p} \sum_{t=p+1}^{n} \sum_{j=1}^{p} (\phi_{j,n} - \widehat{\phi}_{j,n}) X_{t-j} \right\|$$

$$\leqslant \big(\sum_{j=1}^{p} \|\phi_{j,n} - \widehat{\phi}_{j,n}\|^2 \big)^{1/2} \left(\frac{1}{n-p} \sum_{t=p+1}^{n} \sum_{j=1}^{p} \|X_{t-j}\|^2 \right)^{1/2}.$$

By Theorem 2.1 from [16] we have

$$\sum_{j=1}^{p} \|\phi_{j,n} - \widehat{\phi}_{j,n}\|^2 \leqslant p \max_{1 \leqslant j \leqslant p} \|\phi_{j,n} - \widehat{\phi}_{j,n}\|^2 = o_{a.s.}\big((\log n/n)^{1/2}\big).$$

Under the assumption (A1) and using the Markov inequality, we have

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \sum_{j=1}^{p} \|X_{t-j}\|^2 = O_P\big(p(n)\big),$$

and further

$$S_2 = o_{a.s.}\big((\log n/n)^{1/4}\big) O_P\big(p(n)^{1/2}\big) = o_P(1).$$

To bound S_3, we use the Baxter inequality (see [15], Theorem 6.6.12 and p. 271) and get

$$E\,\|S_3\| \leqslant \sum_{j=1}^{\infty} \|\phi_j - \phi_{j,n}\| \, E\,\|X_{t-j}\| \leqslant \text{const} \cdot E\,\|X_t\| \sum_{j=p+1}^{\infty} \|\phi_j\|.$$

Using Corollary 5.1 and the Markov inequality, we see that $S_3 = o_P(1)$. Thus, we have $\bar{\epsilon}_n = o_P(1)$. By formula (5.5) for $\widehat{\epsilon}_{t,n}$ we can write

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\epsilon}_{t,n} \widehat{\epsilon}_{t,n}^T = \frac{1}{n-p} \sum_{t=p+1}^{n} (\epsilon_t \epsilon_t^T + Q_{t,n} Q_{t,n}^T + R_{t,n} R_{t,n}^T + \epsilon_t Q_{t,n}^T$$

$$+ \epsilon_t R_{t,n}^T + Q_{t,n} \epsilon_t^T + Q_{t,n} R_{t,n}^T + R_{t,n} \epsilon_t^T + R_{t,n} Q_{t,n}^T).$$

Under the assumption (A1) we obtain $\frac{1}{n-p} \sum_{t=p+1}^{n} \epsilon_t \epsilon_t^T = E \epsilon_t \epsilon_t^T + O_P(n^{-1/2})$. Using the Cauchy–Schwarz inequality, we have

$$\left\| \frac{1}{n-p} \sum_{t=p+1}^{n} Q_{t,n} Q_{t,n}^T \right\| \leqslant \sum_{j=1}^{p} \| \phi_{j,n} - \widehat{\phi}_{j,n} \|^2 \frac{1}{n-p} \sum_{t=p+1}^{n} \sum_{j=1}^{p} \| X_{t-j} \|^2 = o_P(1).$$

From the Baxter inequality we get

$$E \left\| \frac{1}{n-p} \sum_{t=p+1}^{n} R_{t,n} R_{t,n}^T \right\| \leqslant \sum_{i,j=1}^{\infty} \| \phi_j - \phi_{j,n} \| \, \| \phi_i - \phi_{i,n} \| \, E \, \| X_{t-j} \| \, \| X_{t-i} \|$$

$$\leqslant \text{const} \cdot E \, \| X_t \|^2 \, \Big(\sum_{j=p+1}^{\infty} \| \phi_j \| \Big)^2 = O(p^{-2r}).$$

Thus $\frac{1}{n-p} \sum_{t=p+1}^{n} R_{t,n} R_{t,n}^T = o_P(1)$. The remaining components can be bounded analogously, by using the Cauchy–Schwarz inequality. Finally, we obtain

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\epsilon}_{t,n} \widehat{\epsilon}_{t,n}^T = E \epsilon_t \epsilon_t^T + o_P(1),$$

which completes the proof of the lemma. ∎

By similar arguments to those given above, we obtain

COROLLARY 5.2. *Let us assume that* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$E^*(\epsilon_t^* \epsilon_t^{*T})^2 = E(\epsilon_t \epsilon_t^T)^2 + o_P(1).$$

In the next lemma we prove the consistency of the bootstrap replications ϵ_t^*. It is a multivariate version of Lemma 5.5 given by Bühlmann [9] for the one-dimensional case.

LEMMA 5.3. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

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

P r o o f. Write $F_{\epsilon,n}(x) = \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{1}\{\epsilon_t \leqslant x\}$ and $F_\epsilon(x) = P(\epsilon_t \leqslant x)$. Let $\widehat{F}_{\epsilon,n}$ be given by (3.1) and $d_2(\cdot, \cdot)$ be the Mallows metric. Consequently, we have (see [6])

$$d_2(F_{\epsilon,n}, F_\epsilon) = o(1) \text{ almost surely.}$$

To complete the proof, we have to show that

$$d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1).$$

By the definition of the Mallows metric and (5.5) we obtain

$$d_2^2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) \leqslant \frac{1}{n-p} \sum_{t=p+1}^{n} \|\widetilde{\epsilon}_{t,n} - \epsilon_t\|^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} \|\widehat{\epsilon}_{t,n} - \epsilon_t - \bar{\epsilon}_n\|^2$$
$$= \frac{1}{n-p} \sum_{t=p+1}^{n} \|Q_{t,n} + R_{t,n} - \bar{\epsilon}_n\|^2.$$

Using bounds from the proof of Lemma 5.2, we have $d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1)$, and as a consequence

$$d_2(\widehat{F}_{\epsilon,n}, F_\epsilon) \leqslant d_2(F_{\epsilon,n}, F_\epsilon) + d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1),$$

which completes the proof of the lemma. ■

The following lemma gives us asymptotic bounds for some sample bootstrap estimators.

LEMMA 5.4. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$(5.6) \qquad\qquad \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^* = O_{P^*}(n^{-1/2}) \text{ in probability,}$$

$$(5.7) \qquad\qquad \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^* \epsilon_{t+s}^{*T} = O_{P^*}(n^{-1/2}) \text{ in probability,}$$

$$(5.8) \qquad\qquad \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^* \epsilon_t^{*T} = \Sigma^* + O_{P^*}(n^{-1/2}) \text{ in probability,}$$

where $\Sigma^* = E^* \epsilon_t^* \epsilon_t^{*T}$ *and* $s \neq 0$.

P r o o f. Let $\epsilon, \eta > 0$. According to Lemma 5.2 we have

$$E^* \|\epsilon_t^*\|^2 = E^* \sum_{i=1}^{k} \epsilon_{t,i}^{*2} \xrightarrow{P} E \sum_{i=1}^{k} \epsilon_{t,i}^2 = E\|\epsilon_t\|^2.$$

So, there exists some constant M_η such that

$$P(E^* \|\epsilon_t^*\|^2 > M_\eta) < \eta.$$

Let $\delta_{\epsilon,\eta} = \sqrt{M_\eta/\epsilon}$. Thus, (5.6) is a consequence of the following inequality:

$$P\left(P^*\left(n^{1/2} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^* \right\| > \delta_{\epsilon,\eta} \right) > \epsilon \right) \leqslant P\left(\frac{E^* \|\epsilon_t^*\|^2}{\delta_{\epsilon,\eta}^2} > \epsilon \right) < \eta.$$

The bounds (5.7) and (5.8) can be shown analogously, by using Corollary 5.2. ■

In the next lemma we prove the convergence of the bootstrap replications $\widehat{\phi}_j^*$ of the Yule–Walker estimators. It is a multivariate version of the result given by Alonso et al. [1].

LEMMA 5.5. *Let* (LA) *hold. Then*

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

P r o o f. We have the following bounds:

$$
\begin{aligned}
\|\widehat{\Phi}_p^* - \Phi_{p,n}\|_{kp \times p, \infty} &= \|\widehat{\Gamma}_p^{*-1}\widehat{\gamma}_p^* - \Gamma_p^{-1}\gamma_p\|_{kp \times p, \infty} \\
&= \|(\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1})\widehat{\gamma}_p^* - \Gamma_p^{-1}(\gamma_p - \widehat{\gamma}_p^*)\|_{kp \times p, \infty} \\
&\leqslant \|\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1}\|_{kp \times kp, \infty}\|\widehat{\gamma}_p^*\|_{kp \times p, \infty} \\
&\quad + \|\Gamma_p^{-1}\|_{kp \times kp, \infty}\|\gamma_p - \widehat{\gamma}_p^*\|_{kp \times p, \infty},
\end{aligned}
$$

where $\widehat{\Phi}_p^* = (\widehat{\phi}_1^*, \ldots, \widehat{\phi}_p^*)^T$, $\Phi_{p,n} = (\phi_{1,n}, \ldots, \phi_{p,n})^T$, $\widehat{\Gamma}_p^* = [\widehat{\gamma}^*(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p^* = (\widehat{\gamma}^*(1), \ldots, \widehat{\gamma}^*(p))^T$, $\Gamma_p = [\gamma(i-j)]_{i,j=1}^p$, and $\gamma_p = (\gamma(1), \ldots, \gamma(p))^T$. Hannan and Deistler ([15], Theorem 6.6.11) proved that the norm of the matrix Γ_p^{-1} is uniformly bounded with respect to p, i.e. $\sup_{0 < p < \infty} \|\Gamma_p^{-1}\|_{kp \times kp, \infty} \leqslant C < \infty$. Thus, we get

$$
\begin{aligned}
\|\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1}\|_{kp \times kp, \infty} &= \|\widehat{\Gamma}_p^{*-1}(\widehat{\Gamma}_p^* - \Gamma_p)\Gamma_p^{-1}\|_{kp \times kp, \infty} \\
&\leqslant \|\widehat{\Gamma}_p^{*-1}\|_{kp \times kp, \infty}\|\widehat{\Gamma}_p^* - \Gamma_p\|_{kp \times kp, \infty}\|\Gamma_p^{-1}\|_{kp \times kp, \infty} \\
&\leqslant (C + \|\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1}\|_{kp \times kp, \infty})C\|\widehat{\Gamma}_p^* - \Gamma_p\|_{kp \times kp, \infty}.
\end{aligned}
$$

Further, we have

$$\frac{\|\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1}\|_{kp \times kp, \infty}}{(C + \|\widehat{\Gamma}_p^{*-1} - \Gamma_p^{-1}\|_{kp \times kp, \infty})C} \leqslant \|\widehat{\Gamma}_p^* - \Gamma_p\|_{kp \times kp, \infty}.$$

So, we have to show that $\|\widehat{\Gamma}_p^* - \Gamma_p\|_{kp \times kp, \infty} = o_{P^*}(1)$ in probability.

Notice that $\|\widehat{\Gamma}_p^* - \Gamma_p\|_{kp \times kp, \infty} \leqslant p \max_{0 \leqslant i \leqslant p} \|\widehat{\gamma}^*(i) - \gamma(i)\|$. We have

$$p\|\widehat{\gamma}^*(i) - \gamma(i)\| \leqslant p\|\widehat{\gamma}^*(i) - \gamma^*(i)\| + p\|\gamma^*(i) - \gamma(i)\| = S_1 + S_2.$$

The component S_1 contains a sample bootstrap autocovariance function, thus we are able to bound it only for $|i| \leqslant p$. We have

$$
\begin{aligned}
S_1 &= p\left\|\frac{1}{n}\sum_{t=1}^{n-i} X_{t+i}^* X_t^{*T} - \gamma^*(i)\right\| \\
&\leqslant p\left\|\sum_{j,l=0}^{\infty} \widehat{\psi}_{j,n}\frac{1}{n}\sum_{t=1}^{n-i} \epsilon_{t+i-j}^* \epsilon_{t-l}^{*T}\widehat{\psi}_{l,n}^T - \sum_{j=0}^{\infty} \widehat{\psi}_{j+i,n}\Sigma^*\widehat{\psi}_{j,n}^T\right\| \\
&= o_{P^*}(1) \text{ in probability,}
\end{aligned}
$$

where the obtained bounds are a consequence of Theorem 5.2 and Lemma 5.4. Notice that with $p(n) = o\big((\log n/n)^{-1/(2r+2)}\big)$ in assumption (B) of Lemma 5.2 we get

$$(5.9) \qquad\qquad p\,(\Sigma - \Sigma^*) = o_P(1).$$

We bound the component S_2 for each $i \in \mathbb{Z}$. We get

$$
\begin{aligned}
S_2 &= p\Big\| \sum_{j=0}^{\infty} \widehat{\psi}_{j+i,n} \Sigma^* \widehat{\psi}_{j,n}^T - \sum_{j=0}^{\infty} \psi_{j+i} \Sigma \psi_j^T \Big\| \\
&\leqslant \sum_{j=0}^{\infty} p\|\widehat{\psi}_{j+i,n} - \psi_{j+i}\|\|\Sigma^*\|\|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{\infty} \|\psi_{j+i}\| p\|\Sigma^* - \Sigma\|\|\widehat{\psi}_{j,n}\| \\
&\quad + \sum_{j=0}^{\infty} \|\psi_{j+i}\|\|\Sigma\| p\|\widehat{\psi}_{j,n} - \psi_j\| \\
&= o_P(1),
\end{aligned}
$$

where the bounds are obtained by Theorems 5.2 and 5.3 and formula (5.9). Finally, we have

$$(5.10) \qquad\qquad p \max_{0 \leqslant i \leqslant p} \|\widehat{\gamma}^*(i) - \gamma(i)\| = o_{P^*}(1) \text{ in probability,}$$

which completes the proof of the lemma. ∎

Moreover, from Lemma 5.5, Theorem 2.1 in [16] and the Baxter inequality ([15], Theorem 6.6.12 and p. 271) we obtain

COROLLARY 5.3. *Let* (LA) *hold. Then*

$$\max_{1 \leqslant j \leqslant p} \|\widehat{\phi}_j^* - \phi_j\| \xrightarrow{P^*} 0 \text{ in probability.}$$

In the next lemmas we present multivariate generalizations of the results given by Różański and Zagdański for the one-dimensional case (see [24]) and concerning the prediction error $X_{n+h} - \widehat{X}_{n+h}$. Since the proofs of these lemmas are similar to the proofs for the one-dimensional case, we omit them.

LEMMA 5.6. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$X_{n+h} - \widehat{X}_{n+h} = O_P(1).$$

LEMMA 5.7. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$
\begin{aligned}
X_{n+h} - \widehat{X}_{n+h} &= D_{1,h}(\Phi_{h-1})\epsilon_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{n+h-1} \\
&\quad + \epsilon_{n+h} + o_P(1),
\end{aligned}
$$

where $D_{1,h}(\cdot), \ldots, D_{h-1,h}(\cdot)$ *are some continuous functions and*

$$\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T.$$

In the following, we prove analogous results for the bootstrap prediction error $X_{n+h}^* - \widehat{X}_{n+h}$ approximating the prediction error $X_{n+h} - \widehat{X}_{n+h}$.

LEMMA 5.8. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$X_{n+h}^* - \widehat{X}_{n+h} = O_{P^*}(1) \text{ in probability.}$$

P r o o f. We give the proof by induction on h. For $h = 1$ we have

$$
\begin{aligned}
X_{n+1}^* - \widehat{X}_{n+1} &= \sum_{j=1}^{p} \widehat{\phi}_j^* X_{n+1-j} + \epsilon_{n+1}^* - \sum_{j=1}^{p} \widehat{\phi}_j X_{n+1-j} \\
&= \sum_{j=1}^{p} X_{n+1-j} (\widehat{\phi}_j^* - \widehat{\phi}_j) + \epsilon_{n+1}^* \\
&= \epsilon_{n+1}^* + o_{P^*}(1) = O_{P^*}(1) \text{ in probability.}
\end{aligned}
$$

Let us assume that for all l such that $1 \leqslant l \leqslant h - 1$ we have $X_{n+l}^* - \widehat{X}_{n+l} = O_{P^*}(1)$ in probability and we prove it for h. We have

$$
\begin{aligned}
X_{n+h}^* - \widehat{X}_{n+h} &= \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j^* X_{n+h-j}^* + \sum_{j=h}^{p} \widehat{\phi}_j^* X_{n+h-j} + \epsilon_{n+h}^* \Big) \\
&\quad - \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j \widehat{X}_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}_j X_{n+h-j} \Big) \\
&= \sum_{j=1}^{h-1} \widehat{\phi}_j^* (X_{n+h-j}^* - \widehat{X}_{n+h-j}) + \sum_{j=1}^{h-1} \widehat{X}_{n+h-j} (\widehat{\phi}_j^* - \widehat{\phi}_j) \\
&\quad + \sum_{j=h}^{p} X_{n+h-j} (\widehat{\phi}_j^* - \widehat{\phi}_j) + \epsilon_{n+h}^* \\
&= O_{P^*}(1) \text{ in probability,}
\end{aligned}
$$

where the last bounds are a consequence of Lemma 5.5, Corollary 5.3, the Baxter inequality ([15], Theorem 6.6.12 and p. 271) and the induction assumption. ∎

LEMMA 5.9. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$
\begin{aligned}
X_{n+h}^* - \widehat{X}_{n+h} = D_{1,h}(\Phi_{h-1}) \epsilon_{n+1}^* + \ldots + D_{h-1,h}(\Phi_{h-1}) \epsilon_{n+h-1}^* \\
+ \epsilon_{n+h}^* + o_{P^*}(1) \text{ in probability,}
\end{aligned}
$$

where $D_{1,h}(\cdot), \ldots, D_{h-1,h}(\cdot)$ *are the same continuous functions as in Lemma* 5.7 *and* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. We give the proof by induction on h. For $h = 1$ we have

$$X_{n+1}^* - \widehat{X}_{n+1} = \sum_{j=1}^{p} \widehat{\phi}_j^* X_{n+1-j} + \epsilon_{n+1}^* - \sum_{j=1}^{p} \widehat{\phi}_j X_{n+1-j}$$

$$= \sum_{j=1}^{p} X_{n+1-j}(\widehat{\phi}_j^* - \widehat{\phi}_j) + \epsilon_{n+1}^*$$

$$= \epsilon_{n+1}^* + o_{P*}(1) \text{ in probability.}$$

Let us assume that for all l such that $1 \leqslant l \leqslant h - 1$ we have

$$X_{n+l}^* - \widehat{X}_{n+l} = \widetilde{D}_{1,l}(\Phi_{l-1})\epsilon_{n+1}^* + \ldots + \widetilde{D}_{l-1,l}(\Phi_{l-1})\epsilon_{n+l-1}^*$$
$$+ \epsilon_{n+l}^* + o_{P*}(1) \text{ in probability.}$$

We prove the equality for h. We have

$$X_{n+h}^* - \widehat{X}_{n+h} = \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j^* X_{n+h-j}^* + \sum_{j=h}^{p} \widehat{\phi}_j^* X_{n+h-j} + \epsilon_{n+h}^* \Big)$$

$$- \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j \widehat{X}_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}_j X_{n+h-j} \Big)$$

$$= \sum_{j=1}^{h-1} \widehat{\phi}_j^* (X_{n+h-j}^* - \widehat{X}_{n+h-j}) + \sum_{j=1}^{h-1} \widehat{X}_{n+h-j}(\widehat{\phi}_j^* - \widehat{\phi}_j)$$

$$+ \sum_{j=h}^{p} X_{n+h-j}(\widehat{\phi}_j^* - \widehat{\phi}_j) + \epsilon_{n+h}^*$$

$$= \sum_{j=1}^{h-1} \phi_j (X_{n+h-j}^* - \widehat{X}_{n+h-j}) + o_{P*}(1) + \epsilon_{n+h}^* \text{ in probability,}$$

where the last bounds are obtained by Corollary 5.3, Lemma 5.8 and the Baxter inequality ([15], Theorem 6.6.12 and p. 271). We complete the proof of the lemma using the induction assumption. ∎

In the same way as in the one-dimensional case (see [24]) we can prove the following useful lemma.

LEMMA 5.10. *Assume that*

$$|P^* (X_n^* \leqslant u) - P (X \leqslant u)| \xrightarrow{P} 0$$

for some continuity point u *of the cumulative distribution function* F_X, *where* $\leqslant$ *means the relation of product order (partial order) in* $\mathbb{R}^k$,

$$Y_n^* \xrightarrow{P^*} 0 \text{ in probability,}$$

$$V_n \xrightarrow{P} 0.$$

Then

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

The next theorem gives us consistency of the hybrid bootstrap. An analogous result for the one-dimensional case was given by Różański and Zagdański (see [24]).

THEOREM 5.4. *Let* (LA) *hold. Then for $h \in \mathbb{N}$ we have*

$$|P^*(X_{n+h}^* - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)| = o_P(1)$$

for each u being a continuity point of the cumulative distribution function of the random vector $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$, where $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. Let $h \in \mathbb{N}$. According to Lemmas 5.7 and 5.9 we can write

$$P^*(X_{n+h}^* - \widehat{X}_{n+h}^* \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)$$
$$= P^*\big(D_{1,h}(\Phi_{h-1})\epsilon_{n+1}^* + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{n+h-1}^* + \epsilon_{n+h}^* + o_{P^*}(1) \leqslant u\big)$$
$$- P\big(D_{1,h}(\Phi_{h-1})\epsilon_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1) \leqslant u\big)$$
$$= P^*\big(D_{1,h}(\Phi_{h-1})\epsilon_1^* + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1}^* + \epsilon_h^* + o_{P^*}(1) \leqslant u\big)$$
$$- P\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h + o_P(1) \leqslant u\big).$$

Using the independence of ϵ_t, conditional independence of ϵ_t^* and Lemmas 5.3 and 5.10 we complete the proof of the theorem. ∎

By Theorem 5.4 and Remark 4.1 we have

COROLLARY 5.4. *Let* (LA) *hold. Then for $h \in \mathbb{N}$ we have*

$$|P^*(X_{n+h}^* - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \mathrm{Proj}_n X_{n+h} \leqslant u)| = o_P(1)$$

for each u being a continuity point of the cumulative distribution function of the random vector $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$, where $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$ and $\mathrm{Proj}_n X_{n+h}$ is the best linear h-step predictor (in the mean square sense) of X_{n+h}.

COROLLARY 5.5. *Let* (LA) *hold. Additionally, assume that the cumulative distribution function of the distribution ϵ_t is continuous. Then for $h \in \mathbb{N}$ we have*

$$\sup_{u \in \mathbb{R}^k} |P^*(X_{n+h}^* - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)| = o_P(1)$$

and

$$\sup_{u \in \mathbb{R}^k} |P^*(X_{n+h}^* - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \mathrm{Proj}_n X_{n+h} \leqslant u)| = o_P(1).$$

To prove the consistency of bootstrap-t, we have to show the consistency of the estimator $\widehat{\Sigma}_{n,X}(h)$ (estimator of the mean squared prediction error (4.1)).

LEMMA 5.11. *Let* (LA) *hold. Then*

(5.11)
$$\widehat{\Sigma}_{n,X}(h) \xrightarrow{P} \Sigma_X(h),$$

(5.12)
$$\widehat{\Sigma}_{n,X}^{-1}(h) \xrightarrow{P} \Sigma_X^{-1}(h).$$

P r o o f. We consider an estimator of the mean squared prediction error of the form (4.2):

$$\Sigma_X(h) = \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T, \quad \widehat{\Sigma}_{n,X}(h) = \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T.$$

Thus

$$\Big\| \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T - \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T \Big\| \leqslant \sum_{j=0}^{h-1} \|\widehat{\psi}_{j,n} - \psi_j\| \|\widehat{\Sigma}\| \|\widehat{\psi}_{j,n}\|$$

$$+ \sum_{j=0}^{h-1} \|\psi_j\| \|\widehat{\Sigma} - \Sigma\| \|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma\| \|\widehat{\psi}_{j,n} - \psi_j\|$$

$$= o_P(1),$$

where the bounds are obtained by using (5.1) and Theorems 5.2 and 5.3. From the consistency of $\widehat{\Sigma}_{n,X}(h)$ it follows that $P(\widehat{\Sigma}_{n,X}(h)$ is invertible$) \to 1$ as $n \to \infty$, which together with the continuity of $\widehat{\Sigma}_X^{-1}(h)$ as a function of elements of $\widehat{\Sigma}_X(h)$ imply (5.12) (see [25], Theorem 5.18, p. 188). ■

In the next lemma we prove the convergence of the bootstrap replication $\widehat{\psi}_j^*$. It is a multivariate version of the result given by Zagdański [29].

LEMMA 5.12. *Let* (LA) *hold. Then*

$$\sup_{0 \leqslant j} \|\widehat{\psi}_j^* - \psi_j\| = o_{P^*}(1) \text{ in probability.}$$

P r o o f. We use a similar method to that given in the proof of Theorem 5.3. We can write analogous equations:

$$\psi_j \Sigma = \gamma(j) - \sum_{i=1}^{\infty} \gamma(j+i) \phi_i^T, \quad \widehat{\psi}_j^* \Sigma^* = \gamma^*(j) - \sum_{i=1}^{p} \gamma^*(j+i) \widehat{\phi}_i^{*T}.$$

For $j \in \mathbb{N}$ we have

$$\|\widehat{\psi}_j^* - \psi_j\| \leqslant \|\Sigma^{-1} - \Sigma^{*-1}\| \big\| \gamma(j) - \sum_{i=1}^{\infty} \gamma(j+i)\phi_i^T \big\|$$
$$+ \|\Sigma^{*-1}\| \|\gamma^*(j) - \gamma(j)\|$$
$$+ \|\Sigma^{*-1}\| \big\| \sum_{i=1}^{\infty} \gamma(j+i)\phi_i^T - \sum_{i=1}^{p} \gamma^*(j+i)\widehat{\phi}_i^{*T} \big\|$$
$$= S_1 + S_2 + S_3.$$

We treat the components S_i individually. We have

$$\|\Sigma^{-1} - \Sigma^{*-1}\| = \|\Sigma^{*-1}(\Sigma^* - \Sigma)\Sigma^{-1}\|$$
$$\leqslant \|\Sigma^{*-1}\| \|\Sigma^* - \Sigma\| \|\Sigma^{-1}\|$$
$$\leqslant (\|\Sigma^{-1}\| + \|\Sigma^{-1} - \Sigma^{*-1}\|) \|\Sigma^* - \Sigma\| \|\Sigma^{-1}\|.$$

Further we get

$$(5.13) \qquad \frac{\|\Sigma^{-1} - \Sigma^{*-1}\|}{(C + \|\Sigma^{-1} - \Sigma^{*-1}\|)\,C} \leqslant \|\Sigma^* - \Sigma\| = o_P(1),$$

where $\|\Sigma^{-1}\| \leqslant C$. Thus $S_1 = o_P(1)$ by Lemma 5.2 and the assumption (A2). We bound the components S_2 and S_3 using the same method. Thus, we present the calculations only for S_3:

$$S_3 = \|\Sigma^{*-1}\| \big\| \sum_{i=1}^{p} \gamma(j+i)(\phi_i^T - \widehat{\phi}_i^{*T}) + \sum_{i=p+1}^{\infty} \gamma(j+i)\phi_i^T$$
$$+ \sum_{i=1}^{p} \big(\gamma(j+i) - \gamma^*(j+i)\big)\widehat{\phi}_i^{*T} \big\|$$
$$\leqslant \|\Sigma^{*-1}\| \big(\max_{1 \leqslant i \leqslant p} \|\phi_i - \widehat{\phi}_i^*\| \sum_{i=1}^{\infty} \|\gamma(i)\| + \big\|[EX_{t,i}^2 EX_{t,j}^2]_{i,j=1}^k\big\| \sum_{i=p+1}^{\infty} \|\phi_i\|$$
$$+ p \max_{1 \leqslant i \leqslant p} \|\gamma(i) - \gamma^*(i)\| \max_{1 \leqslant i \leqslant p} \|\widehat{\phi}_i^*\| \big)$$
$$= o_{P^*}(1),$$

where the bounds are a consequence of (5.10), (5.13) and Corollary 5.3. ∎

LEMMA 5.13. *Let* (LA) *hold. Then*

$$(5.14) \qquad \widehat{\Sigma}_X^*(h) \xrightarrow{P^*} \Sigma_X(h) \text{ in probability},$$
$$(5.15) \qquad \widehat{\Sigma}_X^{*-1}(h) \xrightarrow{P^*} \Sigma_X^{-1}(h) \text{ in probability}.$$

P r o o f. We consider the bootstrap estimator of the mean squared prediction error in the form

$$\widehat{\Sigma}_X^*(h) = \sum_{j=0}^{h-1} \widehat{\psi}_j^* \Sigma^* \widehat{\psi}_j^{*T}.$$

Thus, using (5.10) and Lemma 5.12, we obtain

$$\Big\| \sum_{j=0}^{h-1} \widehat{\psi}_j^* \Sigma^* \widehat{\psi}_j^{*T} - \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T \Big\| \leqslant \sum_{j=0}^{h-1} \|\widehat{\psi}_j^* - \psi_j\| \|\Sigma^*\| \|\widehat{\psi}_j^*\|$$
$$+ \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma^* - \Sigma\| \|\widehat{\psi}_j^*\| + \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma\| \|\widehat{\psi}_j^* - \psi_j\|$$
$$= o_{P^*}(1),$$

which proves (5.14). And (5.15), in the same way as (5.12), follows from the continuity of $\widehat{\Sigma}_X^{*-1}(h)$ as a function of elements of $\widehat{\Sigma}_X^*(h)$ (see [25], Theorem 5.18, p. 188). ∎

REMARK 5.1. *All proofs of the theorems and lemmas above remain valid for* $p(n)$ *chosen as in Step* 1 *of the sieve bootstrap algorithm.*

The next theorem gives us consistency of bootstrap-t. An analogous result for the one-dimensional case was given by Zagdański (see [29]).

THEOREM 5.5. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$ *we have*

$$\big| P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big)$$
$$- P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\big| = o_P(1)$$

for each u *being a continuity point of the cumulative distribution function of the random vector* $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$, *where* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. We have the convergence

$$X_{n+h} - \widehat{X}_{n+h} \xrightarrow{D} D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h.$$

Using the multidimensional version of Slutsky's theorem ([20], Proposition C, p. 683) and Lemma 5.11, we have

$$\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h})$$
$$\xrightarrow{D} \Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big).$$

Similarly, using the conditional independence of ϵ_t^* and Lemma 5.3, we obtain convergence for the bootstrap prediction error

$$X_{n+h}^* - \widehat{X}_{n+h} \xrightarrow{D^*} D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h \text{ in probability.}$$

By the Conditional Slutsky's Theorem ([19], p. 77) and Lemma 5.13 we get

$$\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h})$$
$$\xrightarrow{D^*} \Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big) \text{ in probability.}$$

Finally, we show that

$$\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big)$$
$$- P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\Big| = o_P(1)$$

for some u being a continuity point of the cumulative distribution function of the random vector $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$. ∎

From Corollary 5.4, Lemmas 5.11, 5.13 and Theorem 5.5 we deduce the following corollaries.

COROLLARY 5.6. *Let* (LA) *hold. Then for $h \in \mathbb{N}$ we have*

$$\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big)$$
$$- P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \text{Proj}_n X_{n+h}) \leqslant u\big)\Big| = o_P(1)$$

for each u being a continuity point of the cumulative distribution function of the random vector $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$, where $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

COROLLARY 5.7. *Let* (LA) *hold. Additionally, assume that the cumulative distribution function of the distribution ϵ_t is continuous. Then for $h \in \mathbb{N}$ we have*

$$\sup_{u \in \mathbb{R}^k} \Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big)$$
$$- P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\Big| = o_P(1)$$

and

$$\sup_{u \in \mathbb{R}^k} \Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big)$$
$$- P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \text{Proj}_n X_{n+h}) \leqslant u\big)\Big| = o_P(1).$$

5.3. Consistency of extreme statistics. By Lemma 5.7 and the continuous mapping theorem we conclude that U_{n+h}, V_{n+h} and R_{n+h} given by formulas (4.17)–(4.19) are convergent in distribution to some random variables.

LEMMA 5.14. *Assume that* (LA) *hold and that ϵ_t has a continuous distribution function. Then, as $n \to \infty$, $U_{n+h} \xrightarrow{D} U$, $V_{n+h} \xrightarrow{D} V$ and $R_{n+h} \xrightarrow{D} R$, where U, V and R are random variables with continuous distribution functions.*

In the next lemma we show that distributions of U_{n+h}^*, V_{n+h}^* and R_{n+h}^* given by formulas (4.20)–(4.22) are close to their corresponding non-bootstrap distributions.

LEMMA 5.15. *Assuming that* (LA) *hold, we have $U_{n+h}^* \xrightarrow{D^*} U$, $V_{n+h}^* \xrightarrow{D^*} V$ and $R_{n+h}^* \xrightarrow{D^*} R$ in probability.*

P r o o f. By Lemma 5.7, Lemma 10 from [5], and Theorem 5.4 we conclude that, as $n \to \infty$,

$$W_{n+h} = X_{n+h} - \widehat{X}_{n+h} \xrightarrow{D} W = D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h,$$

and the sequence of bootstrap distributions of W_{n+h}^* weakly approaches in probability the sequence of distributions of W_{n+h}. In consequence, $W_{n+h}^* \xrightarrow{D^*} W$ in probability. Since the mappings defining U_{n+h}^*, V_{n+h}^*, R_{n+h}^* are continuous functions, the version of the continuous mapping theorem (see [4]) implies $U_{n+h}^* \xrightarrow{D^*} U$ in probability, $V_{n+h}^* \xrightarrow{D^*} V$ in probability and $R_{n+h}^* \xrightarrow{D^*} R$ in probability, which completes the proof. ■

Using Lemmas 5.11, 5.13, and Theorem 5.5 we can prove analogous results for asymptotic distributions of US_{n+h}, VS_{n+h}, RS_{n+h} and their bootstrap versions US_{n+h}^*, VS_{n+h}^*, RS_{n+h}^* in the same way as in the proofs of the lemmas above.

LEMMA 5.16. *With* (LA) *we have, as $n \to \infty$,*

$$US_{n+h} \xrightarrow{D} US, \quad VS_{n+h} \xrightarrow{D} VS \quad and \quad RS_{n+h} \xrightarrow{D} RS.$$

LEMMA 5.17. *With* (LA)

$$US_{n+h}^* \xrightarrow{D^*} US, \quad VS_{n+h}^* \xrightarrow{D^*} VS \quad and \quad RS_{n+h}^* \xrightarrow{D^*} RS \quad in \ probability.$$

In the construction of simultaneous confidence intervals we should replace unknown quantiles with their bootstrap equivalences. Therefore, we show that bootstrap quantiles of U_{n+h}^*, V_{n+h}^* and R_{n+h}^* are approximations of the quantiles of U, V and R. This fact has been proved in [24] (see Lemma 5.2). For the clearance let us reformulate their lemma with notation stated in Lemma 5.15.

LEMMA 5.18. *With* (LA)

$$u^*_{n+h,\alpha} = u_\alpha + o_P(1), \quad v^*_{n+h,\alpha} = v_\alpha + o_P(1), \quad r^*_{n+h,\alpha} = r_\alpha + o_P(1).$$

In practice, we also do not know the exact bootstrap distributions of U^*_{n+h}, V^*_{n+h} and R^*_{n+h}. Therefore, using the result by Shi et al. [26], we apply the Monte Carlo method to approximate corresponding quantiles. In consequence, we obtain the following lemmas.

LEMMA 5.19. *With* (LA)

$$\widehat{u}^*_{n+h,\alpha} = u_\alpha + o_P(1),$$
$$\widehat{v}^*_{n+h,\alpha} = v_\alpha + o_P(1),$$
$$\widehat{r}^*_{n+h,\alpha} = r_\alpha + o_P(1).$$

LEMMA 5.20. *With* (LA) *and for a continuous distribution function of* ϵ_t

$$\widehat{us}^*_{n+h,(\cdot)} = us_{(\cdot)} + o_P(1),$$
$$\widehat{vs}^*_{n+h,(\cdot)} = vs_{(\cdot)} + o_P(1),$$
$$\widehat{rs}^*_{n+h,(\cdot)} = rs_{(\cdot)} + o_P(1),$$

where $\widehat{us}^*_{n+h,(\cdot)}$, $\widehat{vs}^*_{n+h,(\cdot)}$ *and* $\widehat{rs}^*_{n+h,(\cdot)}$ *are Monte Carlo bootstrap estimators of quantiles of* US^*_{n+h}, VS^*_{n+h}, RS^*_{n+h}, *and thus the approximations of quantiles of* US_{n+h}, VS_{n+h}, RS_{n+h}.

6. CONSISTENCY OF BOOTSTRAP PREDICTION REGIONS

In this section we present results about the consistency of the bootstrap prediction regions for the stationary time series models. Theorem 6.1 is a multivariate version of the result given by Różański and Zagdański (see [24]), and Theorem 6.3 is a multivariate version of the result given by Zagdański (see [29]).

To prove the consistency of the prediction regions we will use auxiliary results about convergence of quantiles for a weakly convergent sequence of the cumulative distribution function. The first lemma was given by Politis et al. [23] and the second one, the modification for the conditional case, was given by Różański and Zagdański [24].

LEMMA 6.1 ([23], Lemma 1.2.1). *Let* $\{F_n\}$ *be a sequence of cumulative distribution functions which converges to* F *in a weak sense and assume that* F *is continuous and strictly increasing at* $y = F^{-1}(\alpha)$ *(for* $x \in (0,1)$, $F^{-1}(x) = \inf\{y : x \leqslant F(y)\}$*). Then*

$$F^{-1}_n(\alpha) \to F^{-1}(\alpha).$$

LEMMA 6.2 ([24]). *Let $\{F_n^*\}$ be a sequence of cumulative distribution functions which converges to F in a weak sense (i.e. $F_n^* \Rightarrow F$ in probability) and assume that F is continuous and strictly increasing at $y = F^{-1}(\alpha)$. Then*

$$F_n^{*-1}(\alpha) \xrightarrow{P} F^{-1}(\alpha).$$

In Subsection 4.3 we have presented the hybrid bootstrap prediction cube given by (4.10). In practice we do not know the distribution of the random vector $H_n^*(h)$, and in consequence we are not able to compute its quantiles. Thus, we define the modified hybrid bootstrap prediction cube in the form

$$(6.1) \quad \widehat{I}_B(h) = \{X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}_{\alpha/(2k),j}^*,$$
$$\widehat{X}_{n+h,j} + \widehat{q}_{1-\alpha/(2k),j}^*], \; j = 1, \ldots, k\},$$

where $\widehat{q}_{\alpha/(2k),j}^*, \widehat{q}_{1-\alpha/(2k),j}^*$ are Monte Carlo approximations of quantiles $q_{\alpha/(2k),j}^*$, $q_{1-\alpha/(2k),j}^*$ (computed by using B bootstrap replications).

REMARK 6.1. *The replacement of the quantiles is made due to the result given by Shi et al. [26], i.e.*

$$q_{\alpha,j}^* - \widehat{q}_{\alpha,j}^* = o_P(1), \quad j = 1, \ldots, k, \; \alpha \in (0,1).$$

THEOREM 6.1. *Let* (LA) *hold. Additionally, assume that $c_{\alpha/(2k),j}$, $c_{1-\alpha/(2k),j}$ are continuity points of the cumulative distribution function of the random variables H_j, which are the jth coordinates of the random vector $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$. Then for $h \in \mathbb{N}$ we have*

$$P\big(X_{n+h} \in \widehat{I}_B(h)\big) \geqslant 1 - \alpha \quad as \; n \to \infty.$$

P r o o f. Using Theorem 5.4 we get for $1 \leqslant j \leqslant k$

$$X_{n+h,j} - \widehat{X}_{n+h,j} \xrightarrow{D} H_j.$$

By the result given by Różański and Zagdański [24] we have for $1 \leqslant j \leqslant k$

$$P(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}_{\alpha/(2k),j}^*, \widehat{X}_{n+h,j} + \widehat{q}_{1-\alpha/(2k),j}^*]) \to 1 - \alpha/k.$$

Using the Bonferroni inequality, we obtain

$$P\big(X_{n+h} \in \widehat{I}_B(h)\big)$$
$$\geqslant 1 - \sum_{j=1}^{k} P(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}_{\alpha/(2k),j}^*, \widehat{X}_{n+h,j} + \widehat{q}_{1-\alpha/(2k),j}^*]^c) \to 1 - \alpha. \quad \blacksquare$$

In a similar way, we define $\widehat{\mathcal{E}}_B(h)$ as an equivalent of $\mathcal{E}_B(h)$ given by (4.11) in the form

$$(6.2) \qquad \widehat{\mathcal{E}}_B(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T(X_{n+h} - \widehat{X}_{n+h}) \leqslant \widehat{q}^*_{1-\alpha}\},$$

where $\widehat{q}^*_{1-\alpha}$ is the Monte Carlo approximation of quantile $q^*_{1-\alpha}$ (computed by using B bootstrap replications). The consistency of this hybrid prediction region is given in the next theorem.

THEOREM 6.2. *Let* (LA) *hold. Additionally, assume that $c_{1-\alpha}$ is a continuity point of the cumulative distribution function of the random variable* $\|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2$. *Then for $h \in \mathbb{N}$*

$$P\big(X_{n+h} \in \widehat{\mathcal{E}}_B(h)\big) \to 1 - \alpha.$$

P r o o f. Using continuity of the function $\|\cdot\|^2$ and Theorem 5.4, we have

$$\|H_n(h)\|^2 \xrightarrow{D} \|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2,$$

$$\|H_n^*(h)\|^2 \xrightarrow{D^*} \|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2 \text{ in probability.}$$

Let $q_{1-\alpha}$ be the $1 - \alpha$ quantile of the distribution $\|H_n(h)\|^2$. By Lemmas 6.1, 6.2 and the consistency of the bootstrap sample quantiles (Remark 6.1) we have

$$q_{1-\alpha} - c_{1-\alpha} = o(1), \quad q^*_{1-\alpha} - q_{1-\alpha} = o_P(1), \quad \widehat{q}^*_{1-\alpha} - q^*_{1-\alpha} = o_P(1).$$

Notice that

$$P\big(X_{n+h} \in \widehat{\mathcal{E}}_B(h)\big) = P\left(\|H_n(h)\|^2 \leqslant \widehat{q}^*_{1-\alpha}\right)$$
$$= P\left(\|H_n(h)\|^2 \leqslant (\widehat{q}^*_{1-\alpha} - q_{1-\alpha}) + (q^*_{1-\alpha} - q_{1-\alpha}) + (q_{1-\alpha} - c_{1-\alpha}) + c_{1-\alpha}\right)$$
$$= P\big(\|H_n(h)\|^2 + o_P(1) \leqslant c_{1-\alpha}\big).$$

We use the Slutsky theorem to complete the proof:

$$P\left(\|H_n(h)\|^2 + o_P(1) \leqslant c_{1-\alpha}\right)$$
$$\to P\big(\|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2 \leqslant c_{1-\alpha}\big) = 1 - \alpha. \quad \blacksquare$$

We define the modified bootstrap-t prediction cube $\widehat{I}_{B-t}(h)$ and the modified bootstrap-t prediction ellipse $\widehat{\mathcal{E}}_{B-t}(h)$:

$$(6.3) \quad \widehat{I}_{B-t}(h) = \{X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h)\widehat{t}^*_{\alpha/(2k),j},$$
$$\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h)\widehat{t}^*_{1-\alpha/(2k),j}], \ j = 1, \ldots, k\},$$

(6.4) $\widehat{\mathcal{E}}_{B-t}(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T \widehat{\Sigma}_{n,X}^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant \widehat{t}_{1-\alpha}^*\},$

where $\widehat{t}_{\alpha/(2k),j}^*, \widehat{t}_{1-\alpha/(2k),j}^*$ are the Monte Carlo approximations of the quantiles $t_{\alpha/(2k),j}^*, t_{1-\alpha/(2k),j}^*$, and $\widehat{t}_{1-\alpha}^*$ is the Monte Carlo approximation of the quantile $t_{1-\alpha}^*$ (all the approximations are computed by using B bootstrap replications).

THEOREM 6.3. *Let* (LA) *hold. Additionally, assume that* $d_{\alpha/(2k),j}, d_{1-\alpha/(2k),j}$ *are continuity points of the cumulative distribution functions of the random variables* $T_j = H_j/\sigma_{X,j}(h)$. *Then for* $h \in \mathbb{N}$ *we have*

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

P r o o f. Using Theorem 5.4, Lemma 5.11 and the Slutsky theorem, we get for $1 \leqslant j \leqslant k$

$$\frac{X_{n+h,j} - \widehat{X}_{n+h,j}}{\widehat{\sigma}_{X,j}(h)} \xrightarrow{D} T_j.$$

By the result given by Zagdański [29] we have for $1 \leqslant j \leqslant k$

$$P\big(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{t}_{\alpha/(2k),j}^* \widehat{\sigma}_{X,j}(h), \widehat{X}_{n+h,j} + \widehat{t}_{1-\alpha/(2k),j}^* \widehat{\sigma}_{X,j}(h)]\big) \to 1 - \alpha/k.$$

Using the Bonferroni inequality, we obtain

$$P\big(X_{n+h} \in \widehat{I}_{B-t}(h)\big)$$
$$\geqslant 1 - \sum_{j=1}^{k} P\big(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{t}_{\alpha/(2k),j}^* \widehat{\sigma}_{X,j}(h), \widehat{X}_{n+h,j} + \widehat{t}_{1-\alpha/(2k),j}^* \widehat{\sigma}_{X,j}(h)]^c\big)$$
$$\to 1 - \alpha. \quad \blacksquare$$

THEOREM 6.4. *Let* (LA) *hold. Assume that* $d_{1-\alpha}$ *is a continuity point of the cumulative distribution function of the random variable*

$$\big\|\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)\big\|^2.$$

Then for $h \in \mathbb{N}$
$$P\big(X_{n+h} \in \widehat{\mathcal{E}}_{B-t}(h)\big) \to 1 - \alpha.$$

P r o o f. The proof is analogous to the proof of Theorem 6.2. $\blacksquare$

Now, we will prove the consistency of simultaneous bootstrap prediction intervals. First, assume that we know the distribution of U, V and R. By the Slutsky lemma, we obtain the following theorem.

THEOREM 6.5. *With the assumptions from Lemma* 5.14 *we get*

$$(6.5) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + u_{\alpha/2}, \widehat{X}_{n+h,i} + v_{1-\alpha/2}]\big) \geqslant 1 - \alpha,$$

$$(6.6) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + u_{\alpha}, +\infty)\big) = 1 - \alpha,$$

$$(6.7) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + v_{1-\alpha}]\big) = 1 - \alpha,$$

$$(6.8) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} - r_{1-\alpha}, \widehat{X}_{n+h,i} + r_{1-\alpha}]\big) = 1 - \alpha,$$

where $u_{(\cdot)}$, $v_{(\cdot)}$ *and* $r_{(\cdot)}$ *are quantiles of corresponding distributions of* U, V *and* R.

Further, we can formulate the main theorem about consistency of bootstrap prediction simultaneous intervals with theoretical quantiles in formulas (6.5)–(6.8) replaced by their Monte Carlo approximations by using the result given by Shi et al. [26]:

THEOREM 6.6. *Assume that* (LA) *hold and that* ϵ_t *has a continuous distribution. Then for each* u *being the continuity point of limited distribution we have simultaneous prediction intervals satisfying the following*:

$$(6.9) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha/2},$$
$$\widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha/2}]\big) \geqslant 1 - \alpha,$$

$$(6.10) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha}, +\infty)\big) = 1 - \alpha,$$

$$(6.11) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha}]\big) = 1 - \alpha,$$

$$(6.12) \quad \lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{r}^*_{n+h,1-\alpha},$$
$$\widehat{X}_{n+h,i} + \widehat{r}^*_{n+h,1-\alpha}]\big) = 1 - \alpha.$$

P r o o f. For the formula (6.9) we have the following relations:

$$1 - P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha/2}, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha/2}]\big)$$
$$= 1 - P\big(\min_{1\leqslant i\leqslant k}(X_{n+h,i} - \widehat{X}_{n+h,i}) \geqslant \widehat{u}^*_{n+h,\alpha/2}$$
$$\wedge \max_{1\leqslant i\leqslant k}(X_{n+h,i} - \widehat{X}_{n+h,i}) \leqslant \widehat{v}^*_{n+h,1-\alpha/2}\big)$$
$$= P\big(U_{n+h} < \widehat{u}^*_{n+h,\alpha/2} \vee V_{n+h} > \widehat{v}^*_{n+h,1-\alpha/2}\big)$$
$$\leqslant P\big(U_{n+h} < \widehat{u}^*_{n+h,\alpha/2}\big) + P\big(V_{n+h} > \widehat{v}^*_{1-\alpha/2}\big)$$
$$\approx P\big(U < o_P(1) + u_{\alpha/2}\big) + P\big(V > o_P(1) + v_{1-\alpha/2}\big)$$
$$\to \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha \quad \text{as } n \to \infty,$$

where the above approximation is a consequence of Lemma 5.19, and the last convergence follows from the Slutsky lemma. For formulas (6.10)–(6.12) the reasoning is analogous. ∎

As in the case of hybrid intervals defined in Theorem 6.6 we can construct studentized bootstrap intervals and prove the following theorem.

THEOREM 6.7. *With the assumptions from Lemma* 5.14 *we have the following simultaneous prediction intervals*:

$$(6.13) \quad \lim_{n \to \infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in \big[\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha/2}\widehat{\sigma}_{n+h},$$
$$\widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha/2}\widehat{\sigma}_{n+h}\big]\big) \geqslant 1 - \alpha,$$

$$(6.14) \quad \lim_{n \to \infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in \big[\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha}\widehat{\sigma}_{n+h}, +\infty\big)\big) = 1 - \alpha,$$

$$(6.15) \quad \lim_{n \to \infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in \big(-\infty, \widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h}\big]\big) = 1 - \alpha,$$

$$(6.16) \quad \lim_{n \to \infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in \big[\widehat{X}_{n+h,i} - \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h},$$
$$\widehat{X}_{n+h,i} + \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h}\big]\big) = 1 - \alpha,$$

where $\widehat{us}^*_{n+h,(\cdot)}$, $\widehat{vs}^*_{n+h,(\cdot)}$ *and* $\widehat{rs}^*_{n+h,(\cdot)}$ *are Monte Carlo estimators of quantiles of bootstrap distributions of* US^*_{n+h}, VS^*_{n+h} *and* RS^*_{n+h}, *respectively*.

REMARK 6.2. *It is worth noting that by Remark* 4.1 *and Corollaries* 5.4–5.7 *all the constructed bootstrap prediction intervals and regions are asymptotically equivalent to corresponding prediction intervals and regions based on the best linear mean squared prediction of* X_{n+h}.

7. SIMULATIONS

In this section we investigate how the presented procedures work on simulated data. We consider the following $VARMA(5, 4)$ model:

$$X_t = \begin{bmatrix} -0.91 & 0.01 \\ 0.37 & -0.90 \end{bmatrix} X_{t-1} + \begin{bmatrix} -0.37 & 0.12 \\ 0.42 & -0.49 \end{bmatrix} X_{t-2}$$

$$+ \begin{bmatrix} -0.18 & 0.10 \\ 0.30 & 0.18 \end{bmatrix} X_{t-3} + \begin{bmatrix} -0.12 & 0.08 \\ 0.14 & 0.24 \end{bmatrix} X_{t-4}$$

$$+ \begin{bmatrix} 0.17 & -0.02 \\ 0.18 & 0.36 \end{bmatrix} X_{t-5} + \begin{bmatrix} -0.91 & 0.01 \\ 0.37 & -0.90 \end{bmatrix} \epsilon_{t-1}$$

$$+ \begin{bmatrix} -0.37 & 0.12 \\ 0.42 & -0.49 \end{bmatrix} \epsilon_{t-2} + \begin{bmatrix} -0.18 & 0.10 \\ 0.30 & 0.18 \end{bmatrix} \epsilon_{t-3}$$

$$+ \begin{bmatrix} -0.12 & 0.08 \\ 0.14 & 0.24 \end{bmatrix} \epsilon_{t-4} + \epsilon_t.$$

In the model $VARMA(5,4)$ we used the following distributions for the noise process ϵ_t:

(N) normal distribution $\mathcal{N}(0,\Sigma)$,

(T) t-Student distribution $\mathcal{T}(5)$,

(χ^2) χ-square distribution $\chi^2(5)$,

(M) mixture of the normal distributions $0.1\mathcal{N}([9,9]^T,\Sigma) + 0.9\mathcal{N}([-1,-1]^T,\Sigma)$.

The observations of the noise process ϵ_t for each of these distributions have been scaled (observations from χ^2 have been centered). Thus for each considered distribution we have the mean $E\epsilon_t = [0,0]^T$ and the covariance matrix

$$E\epsilon_t\epsilon_t^T = \Sigma = \left[\begin{array}{cc} 1.0 & 0.5 \\ 0.5 & 1.0 \end{array}\right].$$

On the base of the simulation results, we compare sample coverage of the prediction regions, which were computed by using the Box–Jenkins method (this method assumes normality of ϵ_t), with the sample coverage of the bootstrap prediction regions. We check performance of each method for different distributions of ϵ_t. We use t-Student distribution (T) as a heavy tailed distribution, χ-square distribution (χ^2) as a nonsymmetric distribution, and mixture (M) of the normal distributions as a bimodal distribution.

In simulations we used parameters:
- confidence level $1-\alpha = 90\%$,
- forecast horizon $h = 1,2,3,4,5$,
- number of observations $n = 50, 200$,
- number of bootstrap replications $B = 1000$,
- number of Monte Carlo repetitions $N = 1000$.

We have constructed three types of the prediction regions. The first type of the prediction regions is constructed by using the Bonferroni inequality and the prediction regions have cubical shape. The Box–Jenkins prediction cube is given as in equation (4.4) and the bootstrap prediction cubes have forms of hybrid bootstrap (6.1) and bootstrap-t (6.3).

The prediction regions of the second type have elliptical shape. The Box–Jenkins prediction ellipse is given by (4.3) and the bootstrap prediction ellipses are given by hybrid bootstrap (6.2) and bootstrap-t (6.4).

The third type of prediction regions is constructed by using extreme statistics, and the prediction regions have cubical shape. We have constructed the hybrid bootstrap prediction cubes $I_B^{UV}(h)$ (see (4.23)), $I_B^R(h)$ (see (4.26)) and bootstrap-t prediction cubes $I_{B-t}^{UV}(h)$ (see (4.27)), $I_{B-t}^R(h)$ (see (4.30)).

In Tables 1 and 3, we present empirical coverage of the prediction cubes, and in Table 2, we present empirical coverage of the prediction ellipses for different number of observations $n = 50, 200$. In the brackets, next to the empirical coverage, we present the mean area of the prediction regions.

The area of the prediction ellipse, constructed by using the Box–Jenkins method, has been calculated via the formula (see [17])

$$(7.1) \qquad V\big(\mathcal{E}_{B-J}(h)\big) = \frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}+1\right)} \left(\chi^2_{1-\alpha}(k)\right)^{k/2} \left(\det\big(\Sigma_X(h)\big)\right)^{1/2},$$

where $\Sigma_X(h)$ is replaced by $\widehat{\Sigma}_X(h)$.

The area of the bootstrap-t prediction ellipse has been calculated by using (7.1) with $\Sigma_X(h)$ replaced by $\widehat{\Sigma}_X(h)$ and the quantile $\chi^2_{1-\alpha}(k)$ replaced by $\widehat{t}^{\,*}_{1-\alpha}$.

It is worth noting the better performance of the bootstrap prediction regions in comparison with the performance of the prediction regions constructed via the Box–Jenkins method.

In all cases we observe that empirical coverage of bootstrap-t prediction regions is larger than empirical coverage of hybrid bootstrap prediction regions but bootstrap-t prediction regions have larger areas.

The empirical coverage of the bootstrap prediction cubes is similar to empirical coverage of the bootstrap prediction ellipses. However, areas of the bootstrap prediction ellipses are smaller than areas of the bootstrap prediction cubes.

The bootstrap prediction regions constructed with extreme statistics are more stable than the bootstrap prediction regions constructed by using the Bonferroni inequality.

TABLE 1. The empirical coverage of the bootstrap prediction cubes for the model $VARMA(5,4)$.

Distribution	h	$n = 50$			$n = 200$		
		Box–Jenkins	hybrid bootstrap	bootstrap-t	Box–Jenkins	hybrid bootstrap	bootstrap-t
N	1	77.2 (3.1)	83.1 (3.8)	89.0 (4.7)	86.1 (3.1)	88.1 (3.4)	90.3 (3.6)
	2	75.6 (5.2)	80.1 (6.2)	86.6 (8.0)	85.2 (5.5)	86.7 (5.8)	88.9 (6.4)
	3	77.0 (6.0)	82.0 (7.1)	87.8 (9.1)	87.2 (6.3)	88.1 (6.7)	89.9 (7.4)
	4	79.3 (6.7)	82.7 (7.7)	87.4 (9.7)	87.7 (7.0)	89.1 (7.4)	91.3 (8.2)
	5	79.3 (7.2)	82.3 (8.2)	87.5 (10.3)	87.7 (7.7)	89.2 (8.1)	90.6 (8.9)
T	1	78.0 (3.1)	82.6 (3.9)	88.0 (4.9)	86.7 (3.1)	88.4 (3.4)	90.6 (3.8)
	2	77.0 (5.2)	82.7 (6.4)	88.9 (8.3)	85.6 (5.6)	87.3 (6.0)	89.2 (6.7)
	3	77.7 (6.1)	82.6 (7.3)	87.1 (9.4)	85.8 (6.4)	87.6 (6.8)	89.4 (7.6)
	4	80.1 (6.7)	83.4 (7.8)	89.0 (10.1)	87.4 (7.1)	88.4 (7.6)	90.4 (8.5)
	5	80.1 (7.3)	83.7 (8.4)	89.1 (10.7)	87.4 (7.8)	88.3 (8.3)	90.6 (9.2)
χ^2	1	75.6 (3.1)	81.7 (3.9)	86.6 (4.8)	85.4 (3.1)	86.5 (3.5)	88.7 (3.8)
	2	77.7 (5.2)	80.7 (6.3)	87.3 (8.2)	84.2 (5.5)	85.5 (5.9)	87.9 (6.6)
	3	76.0 (6.1)	81.3 (7.2)	87.2 (9.2)	84.3 (6.3)	85.2 (6.8)	87.7 (7.5)
	4	76.9 (6.7)	81.7 (7.8)	86.9 (9.9)	85.9 (7.0)	86.9 (7.5)	88.3 (8.3)
	5	76.9 (7.3)	80.3 (8.3)	85.2 (10.5)	85.9 (7.7)	86.4 (8.2)	88.2 (9.1)
M	1	74.7 (3.0)	80.8 (4.0)	85.6 (5.1)	81.1 (3.1)	86.0 (4.0)	87.6 (4.4)
	2	73.4 (5.2)	77.8 (6.5)	83.6 (8.6)	81.0 (5.5)	83.5 (6.3)	85.2 (7.0)
	3	72.8 (6.1)	77.1 (7.4)	82.3 (9.6)	80.2 (6.4)	83.4 (7.1)	85.2 (7.9)
	4	76.2 (6.7)	79.4 (8.0)	84.9 (10.3)	80.2 (6.4)	83.4 (7.1)	85.2 (7.9)
	5	76.2 (7.4)	78.2 (8.6)	83.6 (11.0)	83.4 (7.8)	85.6 (8.6)	86.5 (9.5)

TABLE 2. The empirical coverage of the bootstrap
prediction ellipses for the model $VARMA(5,4)$.

Distribution	h	$n=50$			$n=200$		
		Box–Jenkins	hybrid bootstrap	bootstrap-t	Box–Jenkins	hybrid bootstrap	bootstrap-t
N	1	75.6 (2.8)	85.3 (5.7)	88.7 (4.4)	84.6 (2.9)	87.6 (5.1)	88.9 (3.3)
	2	75.0 (4.8)	84.5 (8.8)	86.3 (7.6)	84.3 (5.1)	87.9 (7.7)	88.5 (5.9)
	3	74.0 (5.5)	83.9 (9.4)	86.0 (8.5)	85.9 (5.7)	89.4 (8.3)	89.1 (6.7)
	4	78.6 (6.2)	85.0 (10.1)	87.6 (9.2)	86.0 (6.5)	88.9 (9.3)	89.6 (7.5)
	5	78.6 (6.7)	84.8 (11.0)	88.2 (9.8)	86.0 (7.2)	89.7 (10.4)	90.0 (8.3)
T	1	77.2 (2.8)	86.6 (5.7)	88.1 (4.5)	86.7 (2.9)	89.4 (5.0)	90.1 (3.4)
	2	75.7 (4.8)	85.6 (9.0)	88.6 (7.8)	85.8 (5.1)	87.8 (7.8)	89.0 (6.1)
	3	76.9 (5.6)	85.3 (9.7)	87.6 (8.8)	86.4 (5.8)	87.8 (8.4)	88.7 (6.8)
	4	77.7 (6.2)	85.5 (10.3)	89.1 (9.4)	85.8 (6.6)	88.9 (9.4)	89.4 (7.7)
	5	77.7 (6.8)	85.4 (11.3)	89.3 (10.1)	85.8 (7.3)	88.8 (10.5)	88.8 (8.5)
χ^2	1	74.4 (2.8)	83.1 (5.7)	86.8 (4.5)	85.1 (2.9)	87.1 (5.1)	88.6 (3.4)
	2	74.7 (4.8)	86.0 (8.9)	87.3 (7.8)	83.9 (5.1)	87.4 (7.8)	88.3 (6.0)
	3	75.7 (5.5)	84.6 (9.5)	87.6 (8.6)	84.0 (5.7)	88.2 (8.4)	87.8 (6.8)
	4	75.1 (6.2)	82.8 (10.2)	87.1 (9.3)	85.4 (6.5)	86.7 (9.3)	89.0 (7.6)
	5	75.1 (6.8)	83.4 (11.1)	85.5 (9.9)	85.4 (7.2)	86.3 (10.5)	88.7 (8.4)
M	1	73.3 (2.8)	82.8 (5.8)	86.8 (4.8)	80.7 (2.9)	88.3 (5.8)	88.3 (4.0)
	2	71.5 (4.8)	81.7 (9.0)	83.9 (8.1)	81.1 (5.1)	86.0 (8.3)	85.5 (6.6)
	3	72.3 (5.6)	83.1 (9.6)	83.8 (9.0)	79.8 (5.8)	84.0 (8.9)	85.2 (7.3)
	4	75.0 (6.2)	83.5 (10.3)	84.1 (9.6)	82.7 (6.6)	88.1 (9.8)	86.7 (8.1)
	5	75.0 (6.8)	82.7 (11.2)	84.6 (10.3)	82.7 (7.3)	88.3 (10.9)	87.6 (8.8)

TABLE 3. The empirical coverage of the bootstrap prediction cubes
constructed by extreme statistics for the model $VARMA(5,4)$.

Distri-bution	h	$n=50$				$n=200$			
		UV	R	UV-t	R-t	UV	R	UV-t	R-t
N	1	85.6 (6.6)	85.6 (6.7)	88.5 (4.7)	89.1 (4.7)	88.1 (6.0)	88.0 (6.0)	90.2 (3.6)	89.7 (3.5)
	2	85.1 (10.0)	85.1 (10.3)	86.3 (7.9)	87.6 (8.0)	88.2 (8.9)	88.8 (9.0)	88.5 (6.3)	88.8 (6.3)
	3	83.7 (10.6)	84.5 (10.8)	87.6 (8.9)	88.3 (9.0)	89.0 (9.5)	89.6 (9.6)	89.8 (7.2)	89.4 (7.2)
	4	84.9 (11.4)	86.0 (11.7)	87.0 (9.6)	87.6 (9.6)	88.4 (10.7)	88.7 (10.7)	90.9 (8.0)	90.7 (8.0)
	5	84.5 (12.4)	86.1 (12.7)	87.7 (10.2)	88.8 (10.3)	90.6 (12.0)	90.6 (12.1)	90.6 (8.7)	90.5 (8.7)
T	1	86.2 (6.6)	87.4 (6.8)	87.5 (4.8)	87.5 (4.7)	89.0 (5.9)	89.2 (6.0)	90.0 (3.7)	89.3 (3.6)
	2	84.6 (10.3)	85.6 (10.5)	87.4 (8.1)	88.1 (8.2)	87.9 (9.0)	88.1 (9.0)	88.5 (6.4)	88.6 (6.4)
	3	84.7 (10.9)	85.4 (11.2)	86.8 (9.2)	87.7 (9.2)	87.7 (9.5)	87.7 (9.6)	88.8 (7.3)	89.1 (7.3)
	4	85.4 (11.7)	85.8 (11.9)	87.7 (9.9)	88.7 (9.9)	89.4 (10.7)	89.5 (10.8)	90.0 (8.2)	89.8 (8.1)
	5	85.4 (12.8)	85.6 (13.1)	88.7 (10.6)	88.7 (10.6)	89.0 (12.1)	89.6 (12.1)	89.8 (8.9)	89.4 (8.9)
χ^2	1	82.7 (6.6)	83.7 (6.7)	86.2 (4.8)	87.1 (4.8)	87.6 (6.0)	87.9 (6.0)	88.6 (3.7)	88.4 (3.6)
	2	85.0 (10.1)	86.4 (10.4)	87.1 (8.0)	87.0 (8.1)	87.2 (9.0)	87.8 (9.0)	88.0 (6.3)	87.4 (6.3)
	3	83.5 (10.7)	85.1 (10.9)	86.7 (9.0)	87.4 (9.0)	87.8 (9.5)	88.2 (9.5)	87.6 (7.2)	87.6 (7.2)
	4	82.5 (11.5)	84.2 (11.7)	86.7 (9.7)	86.9 (9.8)	86.1 (10.6)	87.3 (10.7)	88.0 (8.1)	88.3 (8.1)
	5	82.7 (12.6)	83.8 (12.8)	85.1 (10.4)	85.8 (10.4)	86.0 (12.0)	86.5 (12.1)	88.2 (8.9)	88.2 (8.8)
M	1	82.8 (6.8)	84.0 (6.9)	84.9 (5.0)	86.1 (5.0)	87.7 (6.7)	88.1 (6.8)	87.4 (4.3)	87.3 (4.2)
	2	82.5 (10.3)	83.0 (10.5)	83.8 (8.4)	85.0 (8.4)	86.8 (9.5)	86.5 (9.5)	84.6 (6.7)	85.0 (6.7)
	3	83.7 (10.9)	84.4 (11.1)	82.3 (9.4)	82.8 (9.4)	84.2 (10.0)	84.2 (10.0)	84.5 (7.6)	85.1 (7.6)
	4	83.1 (11.6)	84.1 (11.8)	84.6 (10.0)	84.9 (10.1)	88.5 (11.1)	88.2 (11.1)	86.1 (8.5)	86.3 (8.4)
	5	82.9 (12.8)	83.7 (13.0)	84.1 (10.8)	83.8 (10.8)	88.4 (12.5)	88.9 (12.6)	86.4 (9.2)	86.6 (9.2)

REFERENCES

[1] A. M. Alonso, D. Peña, and J. Romo, *On sieve bootstrap prediction intervals*, Statist. Probab. Lett. 65 (1) (2003), pp. 13–20.

[2] A. M. Alonso, D. Peña, and J. Romo, *Introducing model uncertainty by moving blocks bootstrap*, Statist. Papers 47 (2) (2006), pp. 167–179.

[3] T. W. Anderson, *The Statistical Analysis of Time Series*, Wiley, New York 1971.

[4] Yu. K. Belyaev and O. Seleznjev, *Approaching in distribution with applications to re-sampling of stochastic processes*, Scand. J. Stat. 27 (2) (2000), pp. 371–384.

[5] Yu. K. Belyaev and S. Sjöstedt-de Luna, *Weakly approaching sequences of random distributions*, J. Appl. Probab. 37 (3) (2000), pp. 807–822.

[6] P. J. Bickel and D. A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. 9 (6) (1981), pp. 1196–1217.

[7] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, second edition, Springer, New York 1991.

[8] P. Bühlmann, *Moving-average representation of autoregressive approximations*, Stochastic Process. Appl. 60 (2) (1995), pp. 331–342.

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

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

[11] T. Di Battista, S. A. Gattone, and M. Granturco, *VAR-models for predicting biodiversity*, J. Math. Stat. 1 (4) (2005), pp. 312–315.

[12] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York 1993.

[13] D. Fresoli, E. Ruiz, and L. Pascual, *Bootstrap multi-step forecasts of non-Gaussian VAR models*, Int. J. Forecast. 31 (2015), pp. 834–848.

[14] U. Grenander, *Abstract Inference*, Wiley Ser. Probab. Math. Statist., Wiley, New York 1981.

[15] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, New York 1988.

[16] E. J. Hannan and L. Kavalieris, *Regression, autoregression models*, J. Time Series Anal. 7 (1) (1986), pp. 27–49.

[17] J. H. Kim, *Asymptotic and bootstrap prediction regions for vector autoregression*, Int. J. Forecast. 15 (1999), pp. 393–403.

[18] J.-P. Kreiss, *Bootstrap procedures for AR(∞)-processes*, in: *Bootstrapping and Related Techniques*, K. H. Jöckel, G. Rothe, and W. Sendler (Eds.), Springer, New York 1992, pp. 107–113.

[19] S. N. Lahiri, *Resampling Methods for Dependent Data*, Springer, New York 2003.

[20] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin 2005.

[21] M. Meyer and J. Kreiss, *On the vector autoregressive sieve bootstrap*, J. Time Series Anal. 36 (3) (2015), pp. 377–397.

[22] S. Mirmirani and H. C. Li, *A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil*, Adv. Econom. 19 (2004), pp. 203–223.

[23] D. N. Politis, J. P. 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. Stat. 24 (1) (2004), pp. 5–40.

[25] J. R. Schott, *Matrix Analysis for Statistics*, third edition, Wiley, Hoboken, NJ, 2017.

[26] X. Q. Shi, C. J. Wu, and J. H. Chen, *Weak and strong representations for quantile processes from finite populations with application to simulation size in resampling inference*, Canad. J. Statist. 18 (2) (1990), pp. 141–148.

[27] R. S. Tsay, *Analysis of Financial Time Series*, third edition, Wiley, Hoboken, NJ, 2010.

[28] R. S. Tsay, *Multivariate Time Series Analysis: With R and Financial Applications*, Wiley, Hoboken, NJ, 2014.

[29] A. Zagdański, *On the construction and properties of bootstrap-t prediction intervals for stationary time series*, Probab. Math. Statist. 25 (1) (2005), pp. 133–153.

[30] A. Zygmund, *Trigonometric Series*, Vol. I, Cambridge University Press, London–New York 1968.

Roman Różański and Adam Zagdański
(corresponding authors)
Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology
27 Wybrzeże Wyspiańskiego
50-370 Wrocław, Poland
E-mail: Roman.Rozanski@pwr.edu.pl
E-mail: Adam.Zagdanski@pwr.edu.pl

Grzegorz Chłapiński
KRUK S.A.

Marcin Hławka
Krajowy Rejestr Długów
Biuro Informacji Gospodarczej SA

Krzysztof Jamróz
Santander Bank Polska S.A.

Maciej Kawecki
IT Consulting Maciej Kawecki