

VALID RESAMPLING OF HIGHER ORDER STATISTICS USING THE LINEAR PROCESS BOOTSTRAP AND AUTOREGRESSIVE SIEVE BOOTSTRAP

CARSTEN JENTSCH AND DIMITRIS N. POLITIS

ABSTRACT. We show that the linear process bootstrap (LPB) and the autoregressive sieve bootstrap (AR sieve) are in general not valid for statistics whose large-sample distribution depends on moments of order higher than two, irrespective of whether the data come from a linear time series or not. Inspired by the block-of-blocks bootstrap, we circumvent this non-validity by applying the LPB and AR sieve to suitably blocked data and not to the original data itself. In a simulation study, we compare the LPB, AR sieve and moving block bootstrap applied directly and to blocked data.

1. INTRODUCTION

Over the last decades several techniques for bootstrapping dependent and stationary time series data have been proposed and many papers have been published on this area of research. An overview can be found in the monograph of Lahiri (2003) or the recent review paper by Kreiss and Paparoditis (2011) and the references therein. After having defined a new general bootstrap scheme, usually the first statistic under consideration is the sample mean which appears to be the most simple quantity of interest. Unfortunately, many papers do not go far beyond the sample mean.

Of course other statistics of higher order are of considerable interest. Important examples in time series analysis are estimates for autocovariances, autocorrelations and the spectral density. But also estimates for more general higher order statistics are of interest as well. Many of such statistics are contained in the following broad class of functions of generalized means which has been considered by Künsch (1989), Politis and Romano (1992), Bühlmann (1997) and more recently by Kreiss and Paparoditis (2011). For functions $g : \mathbb{R}^m \rightarrow \mathbb{R}^d$ and $w : \mathbb{R}^d \rightarrow \mathbb{R}$ let

$$T_n = w \left\{ \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t, \dots, X_{t+m-1}) \right\}, \quad (1.1)$$

where $m \in \mathbb{N}$ (fixed) and the functions w and g fulfill the following smoothness assumptions (Bühlmann, 1997, p. 131):

- (A) w has continuous partial derivatives for all y in a neighborhood of $\vartheta = \mathbb{E}g(X_t, \dots, X_{t+m-1})$ and the differentials $\sum_{i=1}^m \partial w / \partial x_i |_{x=\vartheta} y_i$ do not vanish. The function g has partial derivatives of order $h \in \mathbb{N}$, which satisfy a Lipschitz condition.

Date: May 4, 2012.

2000 Mathematics Subject Classification. 62G09; 62M10.

Key words and phrases. AR sieve bootstrap, block bootstrap, generalized means, linear process bootstrap, resampling of blocks, sample autocovariances

Carsten Jentsch, Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany. cjentsch@mail.uni-mannheim.de

Dimitris N. Politis, Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA. politis@math.ucsd.edu.

The class of estimators defined above is quite rich and contains as special cases versions of sample autocovariances, sample autocorrelations, sample partial autocorrelations, Yule–Walker estimators, and the sample mean.

Example 1.1 (Special cases of T_n).

- (i) If $m = 1$ and $w(x) = g(x) = x$, T_n becomes the sample mean $\bar{X} = n^{-1} \sum_{t=1}^n X_t$.
- (ii) If $m = 2$, $w(x) = x$ and $g(x, y) = yx$, T_n becomes a version of the sample autocovariance at lag $h = 1$ for centered observations X_1, \dots, X_n , i.e.

$$\widehat{C}(1) = \frac{1}{n-1} \sum_{t=1}^{n-1} X_{t+1} X_t.$$

- (iii) If $m = 2$, $w(x, y) = x/y$ and $g(x, y) = (yx, x^2)$, T_n becomes a version of the sample autocorrelation at lag $h = 1$ for centered observations X_1, \dots, X_n , i.e.

$$\widehat{R}(1) = \frac{\widehat{C}(1)}{\widehat{C}(0)} = \frac{\frac{1}{n-1} \sum_{t=1}^{n-1} X_{t+1} X_t}{\frac{1}{n-1} \sum_{t=1}^{n-1} X_t^2}.$$

As emphasized in Kreiss and Paparoditis (2011), under appropriate mixing or weak dependence conditions, it holds that $\sqrt{n}\{T_n - w(\vartheta)\}$ has a limiting normal distribution, i.e.

$$\sqrt{n}\{T_n - w(\vartheta)\} \xrightarrow{D} \mathcal{N}(0, \tau^2) \quad (1.2)$$

as $n \rightarrow \infty$, where

$$\tau^2 = \left[\frac{\partial w}{\partial x}(x)|_{x=\vartheta} \right]^2 \sum_{h=-\infty}^{\infty} \text{Cov} \{g(X_t, \dots, X_{t+m-1}), g(X_{t+h}, \dots, X_{t+h+m-1})\}. \quad (1.3)$$

Suppose we observe univariate data X_1, \dots, X_n from some stationary process $\{X_t, t \in \mathbb{Z}\}$ and we are interested in the unknown finite sample distribution of T_n as introduced in (1.1) for some functions w and g fulfilling assumption (A); a generalization to the multivariate case is straightforward and is omitted here for simplicity.

The normal approximation to the finite sample distribution based on the central limit theorem in (1.2) might be poor or difficult to achieve. These problems usually occur for small sample sizes and if the limiting variance contains expressions that are not feasible analytically or difficult to estimate. For these reasons, bootstrap methods are an alternative way to estimate the unknown finite sample distribution of $\sqrt{n}\{T_n - w(\vartheta)\}$.

In this paper, we investigate the capability of three major bootstrap techniques to yield valid bootstrap approximations to the distribution of $\sqrt{n}\{T_n - w(\vartheta)\}$ for statistics that belong to the class of functions of generalized means in (1.1), and discuss under which structural assumptions on the process this is the case. First, we consider the linear process bootstrap (LPB) introduced by McMurry and Politis (2010), and further studied in Jentsch and Politis (2012), which is particularly designed for bootstrapping stationary processes with an abruptly dying-out autocovariance structure as e.g. moving-average processes. Then we study the autoregressive sieve bootstrap (AR sieve), proposed by Kreiss (1988, 1992) and Bühlmann (1997) whose validity extends beyond the class of autoregressive processes as discovered recently in Kreiss et al. (2011).

In Section 3, we show that contrary to well-known block bootstrap techniques—such as the moving block bootstrap (MBB) of Künsch (1989) or the stationary bootstrap of Politis and Romano (1994)—the LPB and AR sieve bootstrap fail in general for higher order statistics and without further assumptions; cf. Kreiss et al. (2011) and Kreiss and Paparoditis (2011) for the AR sieve.

Inspired by the block-of-blocks bootstrap introduced by Politis and Romano (1992) and to circumvent the general non-validity of the LPB and AR sieve for higher order statistics, we adapt the preliminary blocking idea and suggest to apply both methods not directly to given observations, but to form *blocks of data* corresponding to the function g in (1.1) first, and then to apply these methods appropriately to the transformed data. We call the resulting procedures LPB-of-blocks and AR sieve-of-blocks bootstrap, respectively.

The paper is organized as follows. In Section 2, we describe the common bootstrap procedures applied directly to the observations as usual. We illustrate the non-validity of the LPB and AR sieve for higher order statistics if applied directly to the data in Section 3, and we show that this issue still remains even under linearity if it does not come along with invertibility and causality with respect to an i.i.d. noise. Section 4 deals with modifications of LPB, AR sieve and MBB that apply to blocked data. Finally, a simulation study that reinforces our findings is presented in Section 5.

2. THE BOOTSTRAP ALGORITHMS

In this section, we summarize the bootstrap schemes of the LPB, AR sieve and MBB. Although we consider only univariate processes $\{X_t, t \in \mathbb{Z}\}$, we phrase all algorithms for d -dimensional time series data $\underline{X}_1, \dots, \underline{X}_n$ with $\underline{X}_t = (X_{1,t}, \dots, X_{d,t})'$. In Section 4, this allows for a convenient formulation of the modified bootstrap procedures that apply to blocked data.

Suppose we observe d -variate time series data $\underline{X}_1, \dots, \underline{X}_n$ from a stationary process $\{\underline{X}_t, t \in \mathbb{Z}\}$ with mean vector $E(\underline{X}_t) = \underline{\mu}$ and autocovariance matrix $\mathbf{C}(h) = E\{(\underline{X}_{t+h} - \underline{\mu})(\underline{X}_t - \underline{\mu})'\}$.

2.1. Multivariate LPB.

Step 1. Let \mathbf{X} be the $(d \times n)$ -data matrix consisting of columns $\underline{X}_1, \dots, \underline{X}_n$. Compute the centered observations $\underline{Y}_t = \underline{X}_t - \overline{\mathbf{X}}$, where $\overline{\mathbf{X}} = n^{-1} \sum_{t=1}^n \underline{X}_t$, let \mathbf{Y} be the corresponding $(d \times n)$ -matrix of centered observations and define $\underline{\mathbf{Y}} = \text{vec}(\mathbf{Y})$ to be the dn -dimensional vectorized version of \mathbf{Y} .

Step 2. Compute $\underline{\mathbf{W}} = (\widehat{\Gamma}_{\kappa, l}^\epsilon)^{-1/2} \underline{\mathbf{Y}}$, where $(\widehat{\Gamma}_{\kappa, l}^\epsilon)^{1/2}$ denotes the lower left triangular matrix \mathbf{L} of the Cholesky decomposition $\widehat{\Gamma}_{\kappa, l}^\epsilon = \mathbf{L}\mathbf{L}'$. Here, $\widehat{\Gamma}_{\kappa, l}^\epsilon$ is a positive definite version of the tapered covariance matrix estimator

$$\widehat{\Gamma}_{\kappa, l}^\epsilon = \begin{pmatrix} \kappa_l(i-j)\widehat{\mathbf{C}}(i-j) \\ i, j = 1, \dots, n \end{pmatrix} = \mathbf{S}\mathbf{D}\mathbf{S}',$$

where the diagonal matrix of eigenvalues $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_{dn})$ is replaced by $\mathbf{D}^\epsilon = \text{diag}(\lambda_1^\epsilon, \dots, \lambda_{dn}^\epsilon)$ with $\lambda_i^\epsilon = \max(\lambda_i, \epsilon \max_i \widehat{C}_{ii}(0)n^{-\beta})$ for some constants $\beta > 1/2$ and $\epsilon > 0$, i.e. $\widehat{\Gamma}_{\kappa, l}^\epsilon = \mathbf{S}\mathbf{D}^\epsilon\mathbf{S}'$. Here κ is a flat-top taper with $\kappa_l(x) = \kappa(x/l)$, where l denotes a banding parameter; see Jentsch and Politis (2012) for details.

Step 3. Let $\underline{\mathbf{Z}}$ be the standardized version of $\underline{\mathbf{W}}$, that is, $Z_i = (W_i - \overline{W})/\widehat{\sigma}_W$, $i = 1, \dots, dn$, where $\overline{W} = (dn)^{-1} \sum_{t=1}^{dn} W_t$ and $\widehat{\sigma}_W^2 = (dn)^{-1} \sum_{t=1}^{dn} (W_t - \overline{W})^2$.

Step 4. Generate $\underline{\mathbf{Z}}^* = (Z_1^*, \dots, Z_{dn}^*)'$ by random resampling with replacement from $\{Z_1, \dots, Z_{dn}\}$.

Step 5. Compute $\underline{\mathbf{Y}}^* = (\widehat{\Gamma}_{\kappa, l}^\epsilon)^{1/2} \underline{\mathbf{Z}}^*$ and let \mathbf{Y}^* be the matrix that is obtained from $\underline{\mathbf{Y}}^*$ by putting this vector column-wise into a $(d \times n)$ -matrix and denote its columns by $\underline{\mathbf{Y}}_1^*, \dots, \underline{\mathbf{Y}}_n^*$. Define \mathbf{X}^* to be the $(d \times n)$ -matrix consisting of columns $\underline{X}_t^* = \underline{\mathbf{Y}}_t^* + \overline{\mathbf{X}}$.

2.2. Multivariate AR Sieve.

- Step 1. Compute the centered observations $\underline{Y}_t = \underline{X}_t - \overline{\underline{X}}$, where $\overline{\underline{X}} = n^{-1} \sum_{t=1}^n \underline{X}_t$.
- Step 2. Fit a vector-autoregressive process of order p to the centered data $\underline{Y}_1, \dots, \underline{Y}_n$. This leads to estimated $(d \times d)$ -coefficient matrices $\widehat{\mathbf{A}}_1(p), \dots, \widehat{\mathbf{A}}_p(p)$, which are obtained from the multivariate Yule–Walker equations.
- Step 3. Compute the estimated residuals $\widehat{\underline{u}}_t = \underline{Y}_t - \sum_{k=1}^p \widehat{\mathbf{A}}_k(p) \underline{Y}_{t-k}$, and center them to get $\widetilde{\underline{u}}_t = \widehat{\underline{u}}_t - (n-p)^{-1} \sum_{t=p+1}^n \widehat{\underline{u}}_t$, $t = p+1, \dots, n$.
- Step 4. Generate bootstrap observations $\underline{Y}_1^*, \dots, \underline{Y}_n^*$ according to

$$\underline{Y}_t^* = \sum_{k=1}^p \widehat{\mathbf{A}}_k(p) \underline{Y}_{t-k}^* + \underline{u}_t^*,$$

where (\underline{u}_t^*) are randomly drawn with replacement from $\{\widetilde{\underline{u}}_{p+1}, \dots, \widetilde{\underline{u}}_n\}$. Define \mathbf{X}^* to be the $(d \times n)$ -matrix consisting of columns $\underline{X}_t^* = \underline{Y}_t^* + \overline{\underline{X}}$.

2.3. Multivariate MBB.

- Step 1. Let $s \in \mathbb{N}$ be the block length, $N = \lceil n/s \rceil$ and denote by $B_{i,s} = (\underline{X}_i, \dots, \underline{X}_{i+s-1})$ the block of s consecutive observations starting at time index i .
- Step 2. Define discrete and independent Laplace random variables i_1, \dots, i_N taking values in $\{1, \dots, n-s+1\}$.
- Step 3. Lay the blocks $B_{i_1,s}, \dots, B_{i_N,s}$ end-to-end in the order sampled together and discard the last $Ns - n$ observations to get a bootstrap sample $\underline{X}_1^*, \dots, \underline{X}_n^*$.

The bootstrap procedures in 2.1–2.3 have in common that their major tuning parameters, i.e. banding parameter l for the LPB, autoregressive order p for the AR sieve and the block length s for the MBB need to be of an order smaller than n , but need to increase with n to yield valid bootstrap approximations; see e.g. Jentsch and Politis (2012), Kreiss et al. (2011), and Lahiri (2003) for details.

3. NON-VALIDITY OF THE LPB AND AR SIEVE FOR HIGHER ORDER STATISTICS

As discussed in Kreiss and Paparoditis (2011), the question whether a certain bootstrap procedure is valid for some statistic T_n depends essentially on its capability to mimic all characteristics of the underlying distribution of $\{\underline{X}_t, t \in \mathbb{Z}\}$ that crop up in the limiting distribution of T_n .

In the following, we illustrate that the LPB and AR sieve are not valid for higher order statistics under general conditions and that this is still the case even under linearity of the process.

3.1. Non-Validity Under General Conditions.

The LPB and the AR sieve have one crucial property in common: they are both designed to capture the entire autocovariance structure of $\{X_t, t \in \mathbb{Z}\}$ asymptotically, but they do not capture any higher order structure of the underlying process. However, by computing the estimated residuals in Step 2 of the LPB and Step 3 of the AR sieve, respectively, the data gets *decorrelated* and in the limit one obtains *uncorrelated, but not necessarily independent* white noise residuals.

The LPB and AR sieve treat estimated residuals as being independent rather than being uncorrelated. Kreiss et al. (2011) show that this does not influence the validity of the AR sieve to lead to valid bootstrap approximations of the distribution of T_n if its limiting distribution depends exclusively on autocovariances of the underlying process and not on any higher order dependence structure. Thanks to the capability of the LPB to capture all autocovariances in the limit, it is reasonable to expect the same for this bootstrap approach. Examples of statistics

with this property are the sample mean, kernel spectral density estimates as well as sample autocorrelations under linearity in the univariate case; cf. Jentsch and Kreiss (2010).

However, the limiting distributions of higher order statistics as for example those contained in the class (1.1) usually depend on higher order structure of the process and not only on autocovariances. To illustrate this issue, we consider sample autocovariances in the following. Under suitable assumptions, sample autocovariances $\widehat{C}(h)$ satisfy a joint central limit theorem, i.e. for $M \in \mathbb{N}_0$, it holds

$$\sqrt{n} \left(\widehat{C}(h) - C(h) : h = 0, \dots, M \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{V}) \quad (3.1)$$

(cf. Kreiss et al., 2011, Example 3.2), where the covariance matrix $\mathbf{V} = \{V(h, k)\}_{h, k=0, \dots, M}$ is such that

$$V(h, k) = \sum_{r=-\infty}^{\infty} \text{cum}_4(h, r+k, r) + \sum_{r=-\infty}^{\infty} \{C(r+h-k)C(r) + C(r+h)C(r-k)\}, \quad (3.2)$$

where $\text{cum}_4(h, r+k, r) = \text{cum}_4(X_h, X_0, X_{r+k}, X_r)$ denotes the fourth order joint cumulant of X_h, X_0, X_{r+k} and X_r . This result goes along with the central limit theorems in (1.2) and (1.3).

The limiting variance in (3.2) depends additionally to autocovariances also on fourth order joint cumulants of the process. In general, this causes the LPB and the AR sieve to fail for sample autocovariances in particular and generally for all statistics that depend on more than the second order structure of the underlying process.

To shed more light onto this fact, we show that this non-validity still remains even under additionally assumed linearity of the process and we illustrate this issue in the following.

3.2. Non-Validity Under Linearity.

Throughout this section, we assume that the process $\{X_t, t \in \mathbb{Z}\}$ is linear, i.e. it holds

$$X_t = \sum_{\nu=-\infty}^{\infty} c_\nu e_{t-\nu}, \quad t \in \mathbb{Z}, \quad (3.3)$$

where $\{c_j, j \in \mathbb{Z}\}$ is an absolutely summable sequence and $\{e_t, t \in \mathbb{Z}\}$ is an i.i.d. white noise process with $E(e_t^2) = \sigma_e^2 \in (0, \infty)$ and $E(e_t^4) = \eta_e \sigma_e^4 \in (0, \infty)$. The process $\{X_t, t \in \mathbb{Z}\}$ is called *causal* (with respect to $\{e_t, t \in \mathbb{Z}\}$), if $c_k = 0$ for $k < 0$ and it is called *invertible* (with respect to $\{e_t, t \in \mathbb{Z}\}$) if there exists a representation

$$X_t = \sum_{k=1}^{\infty} d_k X_{t-k} + e_t, \quad t \in \mathbb{Z}, \quad (3.4)$$

such that $\sum_{k=1}^{\infty} |d_k| < \infty$ holds.

However, (3.3) and (3.4) are not the only representations of the process $\{X_t, t \in \mathbb{Z}\}$. By the Wold decomposition, every purely nondeterministic, stationary and zero mean process $\{X_t, t \in \mathbb{Z}\}$ can be represented as

$$X_t = \sum_{k=1}^{\infty} b_k \epsilon_{t-k} + \epsilon_t, \quad t \in \mathbb{Z}, \quad (3.5)$$

where $\sum_{k=1}^{\infty} b_k^2 < \infty$ and $\{\epsilon_t, t \in \mathbb{Z}\}$ is a zero mean white noise with variance $E(\epsilon_t^2) = \sigma_\epsilon^2 \in (0, \infty)$. Moreover, for stationary zero mean processes with non-vanishing spectral density, there exist

autoregressive coefficients $\{a_k, k \in \mathbb{N}\}$ such that

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \epsilon_t, \quad t \in \mathbb{Z}, \quad (3.6)$$

with $\{a_k, k \in \mathbb{N}\}$ being absolutely summable; cf. Pourahmadi (2001) or Kreiss et al. (2011). The process (3.5) is invertible with respect to $\{\epsilon_t, t \in \mathbb{Z}\}$ only if $1 + \sum_{k=1}^{\infty} b_k z^k \neq 0$ for $|z| \leq 1$ and (3.6) is causal with respect to $\{\epsilon_t, t \in \mathbb{Z}\}$ only if $1 - \sum_{k=1}^{\infty} a_k z^k \neq 0$ for $|z| \leq 1$. The i.i.d. white noise process $\{\epsilon_t, t \in \mathbb{Z}\}$ in (3.3) is in general different to the uncorrelated, but typically *not* independent white noise process $\{\epsilon_t, t \in \mathbb{Z}\}$ in (3.5) and (3.6).

To return to the question why the LPB and AR sieve may fail even for linear processes, note that the AR sieve typically relies on the Yule–Walker estimators for AR coefficients that are assured to be causal (cf. Brockwell and Davis, 1991). Hence, faced with data from a non-causal, linear AR model, the AR sieve bootstrap may have difficulties. Similarly, the LPB relies on the Cholesky decomposition of a covariance matrix which is equivalent to the innovations algorithm that provides estimates of the MA coefficients that are invertible; cf. Rissanen and Barbosa (1969), Brockwell and Davis (1988). Hence, faced with data from a non-invertible, linear MA model, the LPB may have difficulties.

In the following examples, we illustrate how non-invertible MA(1) processes and non-causal AR(1) processes can be written as AR(∞) in (3.5) and as MA(∞) in (3.6), respectively. The first MA(1) example has been considered previously in Kreiss et al. (2011); the second example can be considered as its canonical AR(1) counterpart.

Example 3.1 (Non-invertible MA(1) processes).

Let $\{X_t, t \in \mathbb{Z}\}$ be a causal moving-average process of order one, i.e.

$$X_t = e_t + b e_{t-1}, \quad t \in \mathbb{Z}, \quad (3.7)$$

where $\{e_t, t \in \mathbb{Z}\}$ is an i.i.d. white noise process with $E(e_t^2) = \sigma_e^2 \in (0, \infty)$, $E(e_t^4) = \sigma_e^4 \eta_e \in (0, \infty)$ and $|b| > 1$ such that $\{X_t, t \in \mathbb{Z}\}$ is non-invertible and (3.7) can be written as

$$e_t = \left(\frac{1}{1 + bL} \right) X_t = - \sum_{k=1}^{\infty} \left(-\frac{1}{b} \right)^k X_{t+k}, \quad t \in \mathbb{Z}. \quad (3.8)$$

There exists another MA(1) representation with white noise $\{\epsilon_t, t \in \mathbb{Z}\}$ that is in general only uncorrelated, but not independent. More precisely, it holds

$$X_t = \left(1 + \frac{1}{b}L \right) \left(\frac{1 + bL}{1 + \frac{1}{b}L} \right) e_t = \epsilon_t + \frac{1}{b} \epsilon_{t-1}, \quad t \in \mathbb{Z}, \quad (3.9)$$

where $\epsilon_t = (1 + b^{-1}L)^{-1}(1 + bL)e_t = \sum_{k=0}^{\infty} \beta_k e_{t-k}$ with $\beta_0 = 1$ and $\beta_k = (-b)^{-k}(1 - b^2)$, $k \in \mathbb{N}$ such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = b^2 \sigma_e^2$, $E(\epsilon_t \epsilon_s) = 0$ for $s \neq t$ and

$$\begin{aligned} \eta_e - 3 &= \frac{E(\epsilon_t^4)}{\{E(\epsilon_t^2)\}^2} - 3 = \frac{3b^4 \sigma_e^4 + (\sum_{k=0}^{\infty} \beta_k^4) \{E(e_1^4) - 3\sigma_e^4\}}{b^4 \sigma_e^4} - 3 \\ &= \frac{b^2(b^4 - 3b^2 + 4)}{1 + b^2} \left[\frac{E(e_t^4)}{\{E(e_t^2)\}^2} - 3 \right] \neq \frac{E(e_t^4)}{\{E(e_t^2)\}^2} - 3 = \eta_e - 3. \end{aligned} \quad (3.10)$$

The MA representation (3.9) may be rewritten as

$$\epsilon_t = \left(\frac{1}{1 + \frac{1}{b}L} \right) X_t = \sum_{k=0}^{\infty} \left(-\frac{1}{b} \right)^k X_{t-k}, \quad t \in \mathbb{Z}. \quad (3.11)$$

Example 3.2 (Non-causal AR(1) processes).

Let $\{X_t, t \in \mathbb{Z}\}$ be an invertible autoregressive process of order one, i.e. $\{X_t, t \in \mathbb{Z}\}$ is the stationary solution of

$$X_t = aX_{t-1} + e_t, \quad t \in \mathbb{Z}, \quad (3.12)$$

where $\{e_t, t \in \mathbb{Z}\}$ is an i.i.d. white noise process with $E(e_t^2) = \sigma_e^2 \in (0, \infty)$, $E(e_t^4) = \sigma_e^4 \eta_e \in (0, \infty)$ and $|a| > 1$ such that $\{X_t, t \in \mathbb{Z}\}$ is non-causal and (3.12) can be written as

$$X_t = \left(\frac{1}{1 - aL} \right) e_t = - \sum_{k=1}^{\infty} \left(\frac{1}{a} \right)^k e_{t+k}, \quad t \in \mathbb{Z}. \quad (3.13)$$

There exists another AR(1) representation with white noise $\{\epsilon_t, t \in \mathbb{Z}\}$ that is in general only uncorrelated, but not independent. More precisely, it holds

$$e_t = (1 - aL)X_t = \left(\frac{1 - aL}{1 - \frac{1}{a}L} \right) \left(1 - \frac{1}{a}L \right) X_t, \quad t \in \mathbb{Z}$$

and, equivalently,

$$X_t = \frac{1}{a}X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \quad (3.14)$$

where $\epsilon_t = (1 - a^{-1}L)(1 - aL)^{-1}e_t = \sum_{k=0}^{\infty} \alpha_k e_{t-k}$ with $\alpha_0 = a^{-2}$ and $\alpha_k = a^{-k}(a^{-2} - 1)$, $k \in \mathbb{N}$ such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = a^{-2}\sigma_e^2$, $E(\epsilon_t \epsilon_s) = 0$ for $s \neq t$ and

$$\begin{aligned} \eta_\epsilon - 3 &= \frac{E(\epsilon_t^4)}{\{E(\epsilon_t^2)\}^2} - 3 = \frac{3a^{-4}\sigma_e^4 + (\sum_{k=0}^{\infty} \alpha_k^4) \{E(e_1^4) - 3\sigma_e^4\}}{a^{-4}\sigma_e^4} - 3 \\ &= \frac{(a^4 - 3a^2 + 4)}{a^2(1 + a^2)} \left[\frac{E(e_t^4)}{\{E(e_t^2)\}^2} - 3 \right] \neq \frac{E(e_t^4)}{\{E(e_t^2)\}^2} - 3 = \eta_e - 3. \end{aligned} \quad (3.15)$$

The AR representation (3.14) may be rewritten as

$$X_t = \left(\frac{1}{1 - \frac{1}{a}L} \right) \epsilon_t = \sum_{k=0}^{\infty} \left(\frac{1}{a} \right)^k \epsilon_{t-k}, \quad t \in \mathbb{Z}. \quad (3.16)$$

Due to their properties discussed above, Yule–Walker estimators applied to data generated by (3.7) or (3.12) estimate the AR coefficients and lead to residuals in (3.11) and in (3.14) instead of the i.i.d. residuals in (3.8) and (3.12). Similarly, the Cholesky decomposition of the covariance matrix executed for the LPB leads to coefficients and residuals in (3.9) and (3.16) instead of the i.i.d. residuals in (3.7) and (3.13), respectively. Compare also Figures 5 and 6 in Section 5, where several correlograms related to these estimated residuals are shown that emphasize that they are indeed uncorrelated, but not independent in the case of non-causality or non-invertibility of the process with respect to i.i.d. noise.

Under linearity, the limiting covariances of sample autocovariances still depend on the fourth order structure of the process. More precisely, (3.2) becomes

$$V(h, k) = C(h)C(k)(\eta_e - 3) + \sum_{r=-\infty}^{\infty} \{C(r+h-k)C(r) + C(r+h)C(r-k)\} \quad (3.17)$$

and to provide asymptotically valid bootstrap approximations for sample autocovariances, the LPB and AR sieve need to mimic not only autocovariances, but also the standardized fourth order cumulant $(\eta_e - 3)$ of $\{e_t, t \in \mathbb{Z}\}$ correctly.

If the underlying process is not causal and invertible with respect to the same independent white noise process $\{e_t, t \in \mathbb{Z}\}$, the LPB and AR sieve methods will lead to uncorrelated but not

independent residuals. As illustrated in Examples 3.1–3.2 above, by applying random resampling from these residuals, the higher order dependence structure beyond autocovariances gets distorted and the cumulant term $(\eta_e - 3)$ is no longer mimicked correctly. The LPB and AR sieve fail in this case unless the process is Gaussian in which case uncorrelated and independent coincide, $\eta_e - 3 = 0$, and eqs. (3.10) and (3.15) hold true.

To conclude this section, recall that the limiting covariances of sample autocorrelations $\widehat{R}(h) = \widehat{C}(h)/\widehat{C}(0)$ depend in the univariate case only on autocovariances of the process if the process is linear. For this reason, the LPB and AR sieve are valid for sample autocorrelations under the assumption of linearity (in the univariate case) contrary to sample autocovariances.

4. RESAMPLING OF BLOCKED DATA

Suppose now we observe univariate data X_1, \dots, X_n from a stationary process $\{X_t, t \in \mathbb{Z}\}$ with zero mean and autocovariance function $C(h) = E(X_{t+h}X_t)$. We want to estimate the unknown finite sample distribution of T_n defined in (1.1) for some functions w and g fulfilling assumption (A) and some fixed $m \in \mathbb{N}$.

Inspired by the block-of-blocks bootstrap proposed by Politis and Romano (1992) and displayed below, the validity of LPB and AR sieve for the sample mean can be used to correct the non-validity for higher order statistics as e.g. sample autocovariances in the general case.

In the following, we introduce modifications of the LPB and AR sieve that apply to blocked data rather than directly to the observations.

4.1. LPB-of-Blocks Bootstrap.

- Step 1. Define $\widetilde{X}_t := g(X_t, \dots, X_{t+m-1})$ and let $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ be the set of d -dimensional data transformed by function g .
- Step 2. Apply the multivariate LPB scheme to the d -dimensional data $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ to get bootstrap observations $\widetilde{X}_1^*, \dots, \widetilde{X}_{n-m+1}^*$.
- Step 3. Compute $T_n^* = w\{(n-m+1)^{-1} \sum_{t=1}^{n-m+1} \widetilde{X}_t^*\}$.
- Step 4. Repeat Steps 2 and 3 B -times, where B is large, and approximate the unknown distribution of $\sqrt{n}\{T_n - w(\theta)\}$ by the empirical distribution of $\sqrt{n}\{T_{n,1}^* - T_n\}, \dots, \sqrt{n}\{T_{n,B}^* - T_n\}$.

4.2. AR Sieve-of-Blocks Bootstrap.

- Step 1. Define $\widetilde{X}_t := g(X_t, \dots, X_{t+m-1})$ and let $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ be the set of d -dimensional data transformed by function g .
- Step 2. Apply the multivariate AR sieve scheme to the d -dimensional data $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ to get bootstrap observations $\widetilde{X}_1^*, \dots, \widetilde{X}_{n-m+1}^*$.
- Step 3. Compute $T_n^* = w\{(n-m+1)^{-1} \sum_{t=1}^{n-m+1} \widetilde{X}_t^*\}$.
- Step 4. Repeat Steps 2 and 3 B -times, where B is large, and approximate the unknown distribution of $\sqrt{n}\{T_n - w(\theta)\}$ by the empirical distribution of $\sqrt{n}\{T_{n,1}^* - T_n\}, \dots, \sqrt{n}\{T_{n,B}^* - T_n\}$.

4.3. Block-of-Blocks Bootstrap.

- Step 1. Define $\widetilde{X}_t := g(X_t, \dots, X_{t+m-1})$ and let $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ be the set of d -dimensional data transformed by function g .
- Step 2. Apply the multivariate MBB scheme to the d -dimensional data $\widetilde{X}_1, \dots, \widetilde{X}_{n-m+1}$ to get bootstrap observations $\widetilde{X}_1^*, \dots, \widetilde{X}_{n-m+1}^*$.

		Model I & II			Model III & IV		
	n	50	250	500	50	250	500
LPB	l	1	1	1	4	8	12
AR sieve	p	2	4	6	1	1	1
Block	s	5	9	11	5	9	11
LPB-of-blocks	l	2	2	2	5	9	13
AR sieve-of-blocks	p	3	5	7	2	2	2
Block-of-blocks	s	5	9	11	5	9	11

TABLE 1. Choice of tuning parameters.

Step 3. Compute $T_n^* = w\{(n - m + 1)^{-1} \sum_{t=1}^{n-m+1} \tilde{X}_t^*\}$.

Step 4. Repeat Steps 2 and 3 B -times, where B is large, and approximate the unknown distribution of $\sqrt{n}\{T_n - w(\theta)\}$ by the empirical distribution of $\sqrt{n}\{T_{n,1}^* - T_n\}, \dots, \sqrt{n}\{T_{n,B}^* - T_n\}$.

Remark 4.1.

The validity of the LPB, AR sieve, and MBB when applied to suitably blocked data, can be verified by checking the conditions of the corresponding theorems for the sample mean of the new process $\{\tilde{X}_t, t \in \mathbb{Z}\}$, namely:

- (i) McMurry and Politis (2010) and Jentsch and Politis (2012) for the LPB-of-blocks bootstrap;
- (ii) Bühlmann (1997), Kreiss et al. (2011) as well as Kreiss and Meyer (2012) for the AR sieve-of-blocks bootstrap;
- (iii) Künsch (1989) and Politis and Romano (1992) for the block-of-blocks bootstrap.

5. SIMULATION STUDY

We compare the LPB-of-blocks bootstrap, AR sieve-of-blocks bootstrap and MBB block-of-blocks bootstrap to the LPB, AR sieve and MBB when applied directly to the observations. In particular, we compare the performance of the aforementioned bootstrap methods in estimating the variances of sample autocovariances at lags zero and one, and of the sample autocorrelation at lag one, i.e. $\text{Var}\{\hat{C}(0)\}$, $\text{Var}\{\hat{C}(1)\}$ and $\text{Var}\{\hat{R}(1)\}$.

We consider realizations of sample size $n \in \{50, 250, 500\}$ from two moving-average models and from two autoregressive models. More precisely, we consider data from the causal and invertible MA(1) model

$$\text{Model I: } X_t = e_t + \frac{3}{10}e_{t-1}, \quad e_t \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}), \quad t \in \mathbb{Z}$$

and from the causal and non-invertible MA(1) model

$$\text{Model II: } X_t = e_t + \frac{10}{3}e_{t-1}, \quad e_t \sim \mathcal{U}\left(-\frac{\sqrt{3}}{10}, \frac{\sqrt{3}}{10}\right), \quad t \in \mathbb{Z},$$

where $\mathcal{U}(a, b)$ denotes the uniform distribution on the interval (a, b) , which has excess kurtosis $\eta_e - 3 = -6/5$ for $a < b$. Both MA(1) processes above possess exactly the same autocovariance structure and Model II is covered by Example 3.1. Furthermore, we consider data from the stationary, invertible and causal solution of the AR(1) model equation

$$\text{Model III: } X_t = \frac{3}{10}X_{t-1} + e_t, \quad e_t \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}), \quad t \in \mathbb{Z}$$

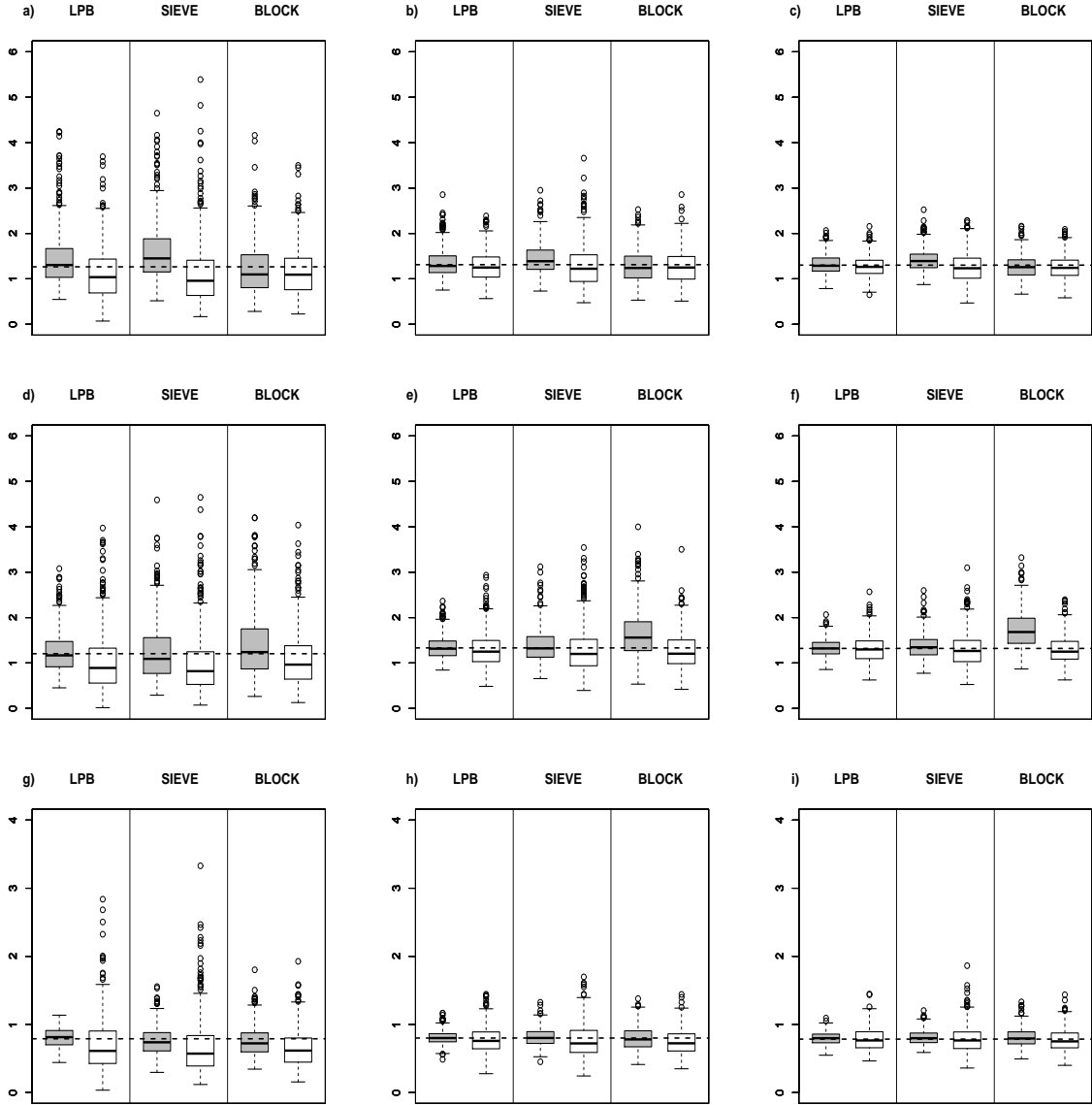


FIGURE 1. Bootstrap estimates of $\text{Var}\{\widehat{C}(0)\}$ (first row), $\text{Var}\{\widehat{C}(1)\}$ (second row), and $\text{Var}\{\widehat{R}(1)\}$ (third row) for Model I based on LPB, SIEVE, and BLOCK bootstrap applied directly (grey boxes) and to blocked data (white boxes). From left to right sample sizes $n = 50, 250$ and 500 have been used.

and from the stationary, invertible and non-causal solution of the AR(1) model equation

$$\text{Model IV: } X_t = \frac{10}{3}X_{t-1} + e_t, \quad e_t \sim \mathcal{U}\left(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}}\right), \quad t \in \mathbb{Z}.$$

Both AR(1) models above have exactly the same autocovariance structure and Model IV is covered by Example 3.2. We have chosen uniformly distributed i.i.d. white noise instead of normal distributed white noise because in the Gaussian case Models I and II as well as III and IV coincide, respectively.

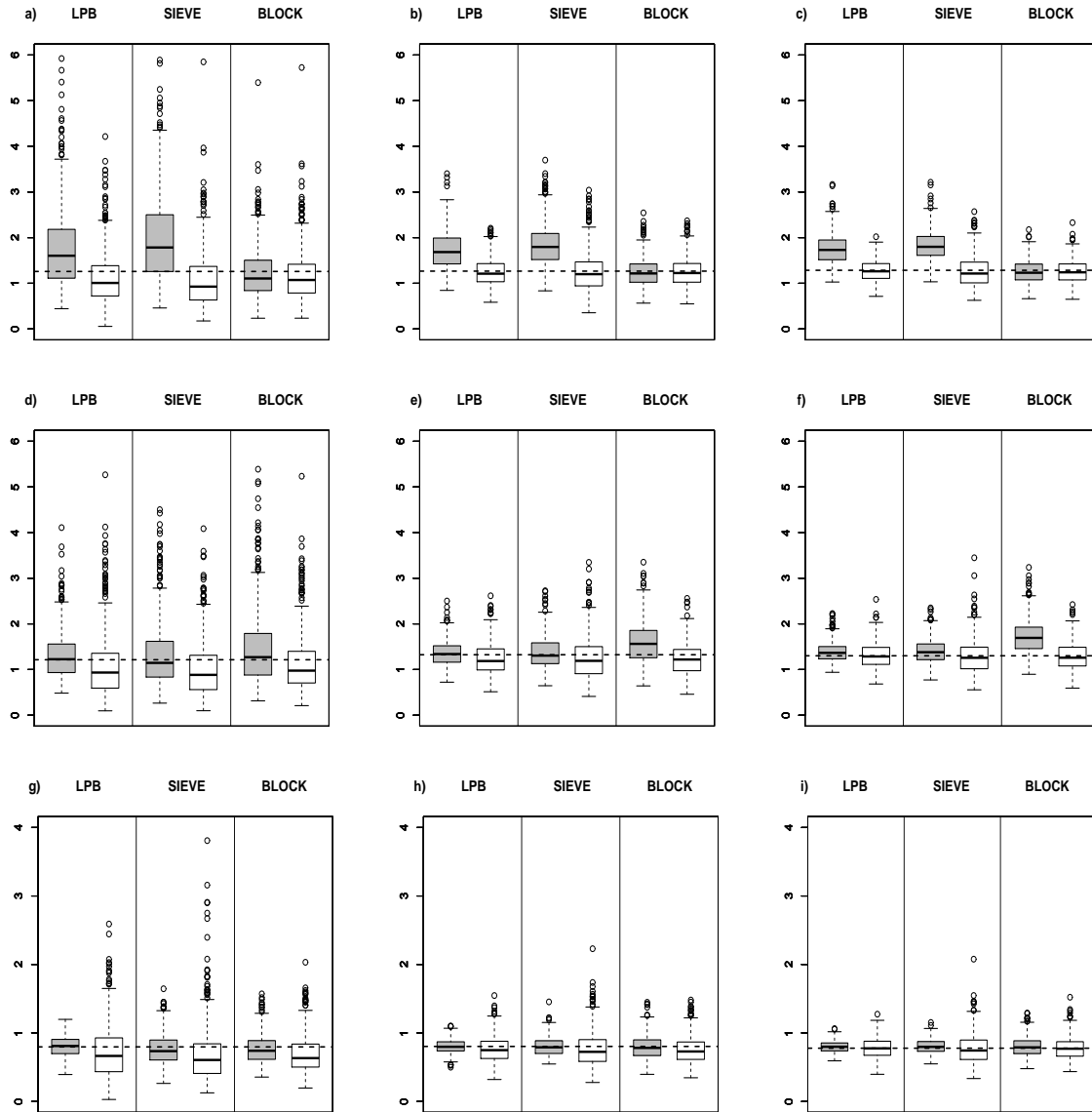


FIGURE 2. Bootstrap estimates of $\text{Var}\{\widehat{C}(0)\}$ (first row), $\text{Var}\{\widehat{C}(1)\}$ (second row), and $\text{Var}\{\widehat{R}(1)\}$ (third row) for Model II based on LPB, SIEVE, and BLOCK bootstrap applied directly (grey boxes) and to blocked data (white boxes). From left to right sample sizes $n = 50, 250$ and 500 have been used.

To estimate the exact variances, 20,000 Monte Carlo replications have been used while bootstrap approximations are based on $B = 250$ bootstrap replications, and we have simulated $T = 500$ data sets for each model. For the LPB we have used a rectangular weight function $\kappa(x) = 1$ if $|x| \leq 1$ and $\kappa(x) = 0$ otherwise to band the estimated covariance matrix. The tuning parameters ϵ and β that assure positive definiteness are set equal to one. The choice of tuning parameters according to model and sample size are summarized in Table 1.

In Figures 1–4, we show boxplots for the bootstrap estimates of $\text{Var}\{\widehat{C}(0)\}$, $\text{Var}\{\widehat{C}(1)\}$ and $\text{Var}\{\widehat{R}(1)\}$ for Models I–IV; grey boxes correspond to bootstrap procedures applied directly to

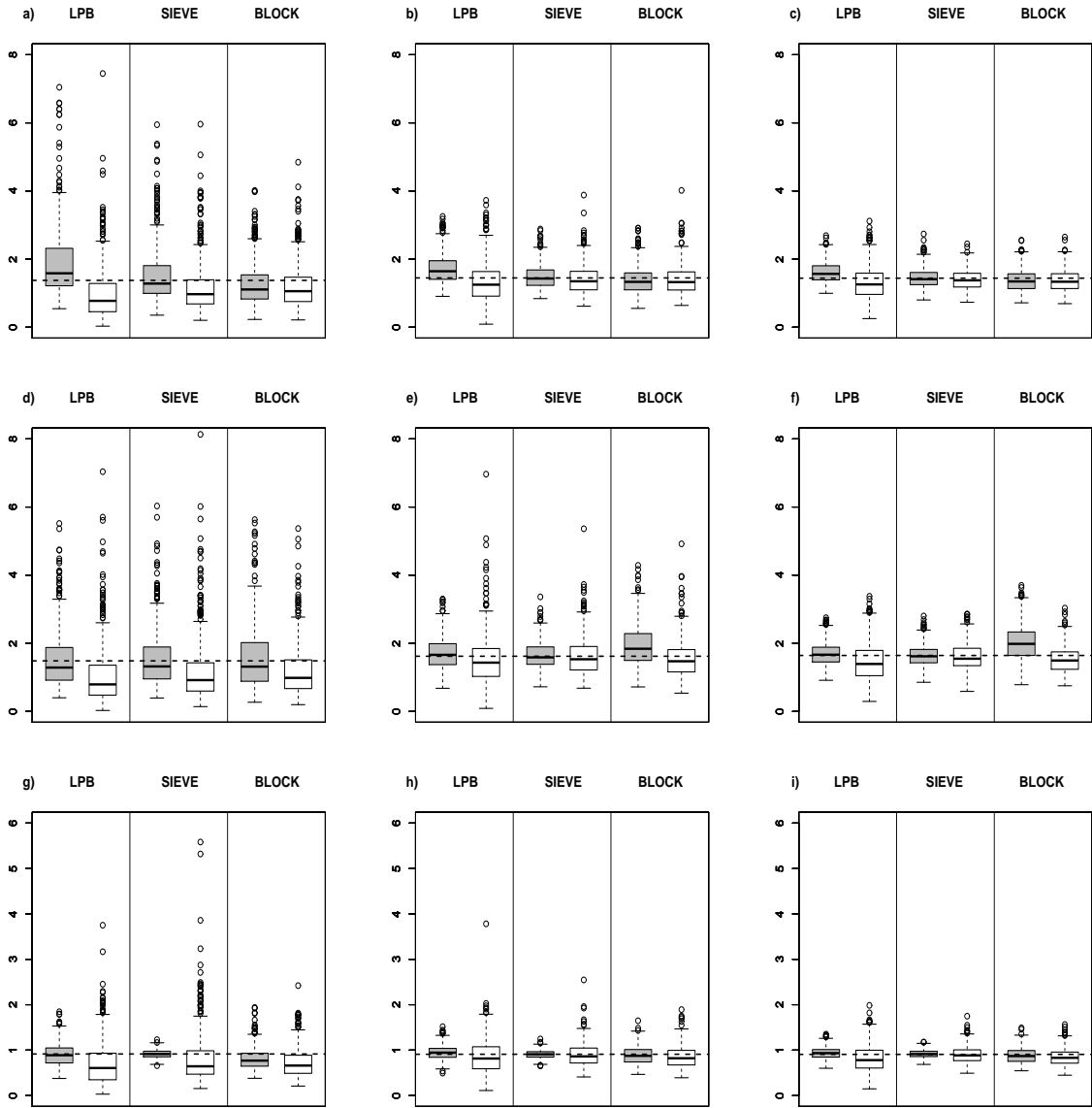


FIGURE 3. Bootstrap estimates of $\text{Var}\{\widehat{C}(0)\}$ (first row), $\text{Var}\{\widehat{C}(1)\}$ (second row), and $\text{Var}\{\widehat{R}(1)\}$ (third row) for Model III based on LPB, SIEVE, and BLOCK bootstrap applied directly (grey boxes) and to blocked data (white boxes). From left to right sample sizes $n = 50, 250$ and 500 have been used.

the data as described in Section 2 while white boxes correspond to bootstrap procedures applied to blocked data as in Section 4.

To illustrate that the residuals obtained by LPB for Model II and by AR sieve for Model IV are uncorrelated, but not independent, we show correlograms of these estimated residuals, their absolute values and their squares for sample size $n = 1000$ in Figures 5 and 6, respectively.

5.1. Discussion of the Results.

In Figures 1 and 3, the simulation results for the well-behaved (causal and invertible) Models I

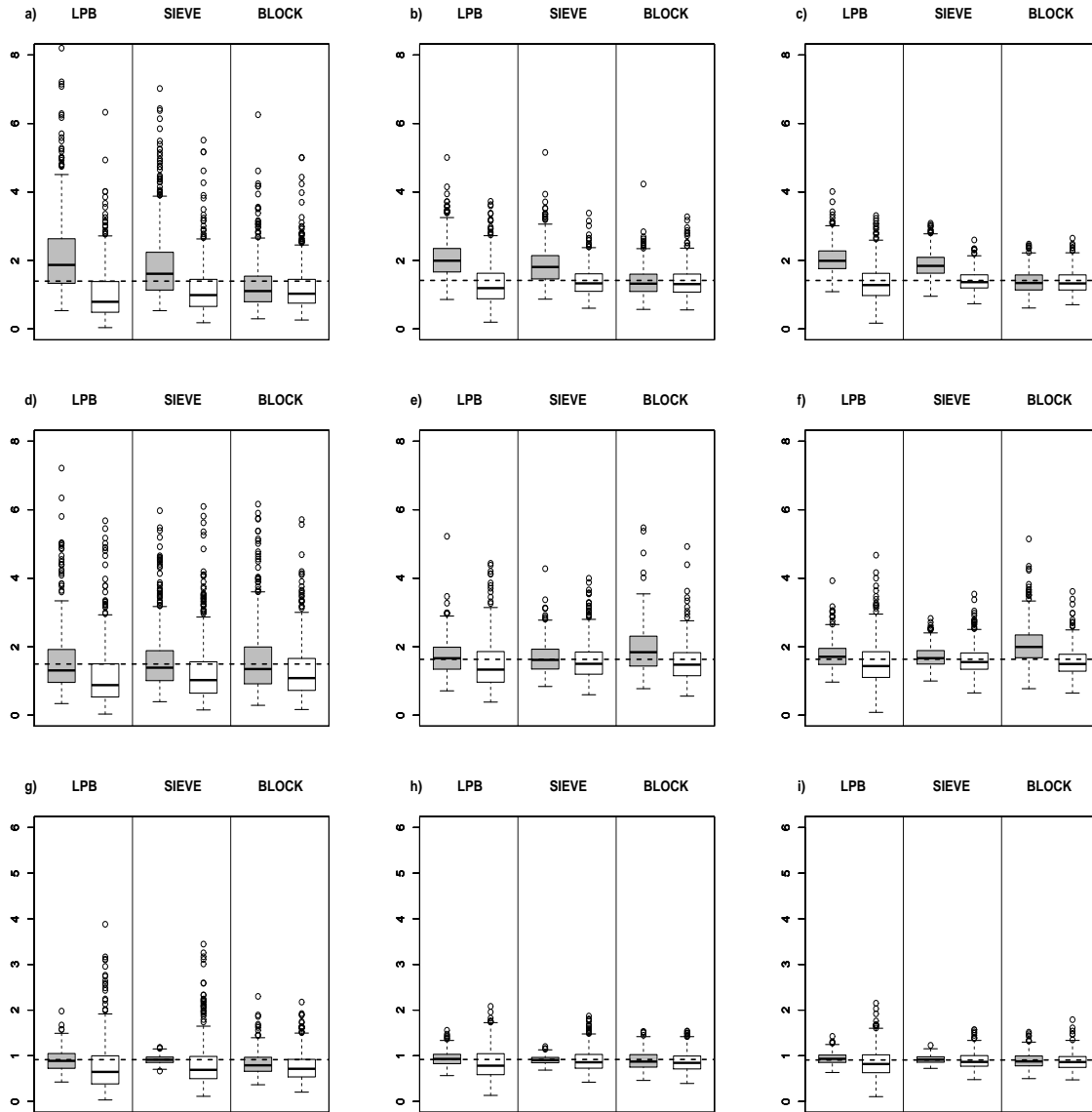


FIGURE 4. Bootstrap estimates of $\text{Var}\{\widehat{C}(0)\}$ (first row), $\text{Var}\{\widehat{C}(1)\}$ (second row), and $\text{Var}\{\widehat{R}(1)\}$ (third row) for Model IV based on LPB, SIEVE, and BLOCK bootstrap applied directly (grey boxes) and to blocked data (white boxes). From left to right sample sizes $n = 50, 250$ and 500 have been used.

and III are displayed. Except for the directly applied block bootstrap for estimating the variance of $\widehat{C}(1)$ in the second rows of both figures, all bootstrap procedures tend to estimate the targets consistently.

Although the autocovariances of Models II and IV are exactly the same as in Model I and III, respectively, in comparison to Figures 1 and 3, things become different in Figure 2 and 4 for Models II and IV. Considering the first row of panels in both Figures 2 and 4, it becomes clear that the LPB and AR sieve applied directly to the data fail systematically to estimate the true variance of $\widehat{C}(0)$ consistently, while both techniques applied to blocked data lead to success. In

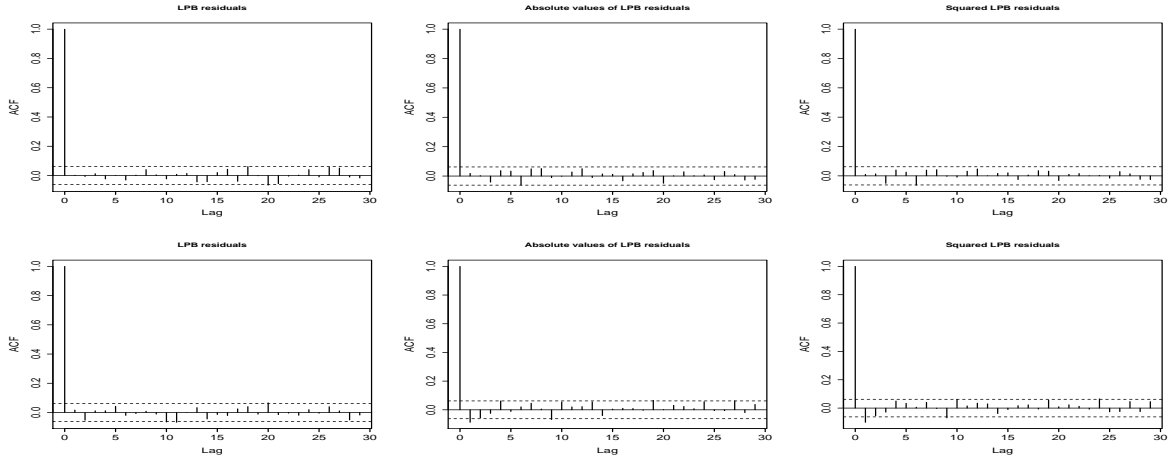


FIGURE 5. Correlograms of residuals (left panels), their absolute values (center panels) and their squares (right panels) computed by LPB based on data from Model I (first row) and from Model II (second row).

view of the discussion in Section 3, this is exactly what we expected and can be explained by noticing that $\eta_e - 3 \neq \eta_\epsilon - 3$. This inconsistency for the LPB and AR sieve is also present in the second rows of panels in both Figures 2 and 4, but it is unrecognizably suppressed here due to multiplication with $C^2(1)$ in the first term of (3.17). However, all bootstrap approaches lead to valid approximations for the variance of $\hat{R}(1)$ in the third rows of Figures 2 and 4 as expected. Again the block bootstrap applied directly to the data performs poorly, which we expect to be caused by an improper choice of the block length.

In summary, the LPB and AR sieve fail to provide valid bootstrap approximations in cases where the underlying model is either non-invertible or non-causal with respect to i.i.d. noise. This inconsistency occurs whatever the choice of the tuning parameters is, and both methods lead to success when they are not applied directly to the observations, but to blocked data.

In general, all bootstrap procedures applied directly to the data perform better if advisable. Particularly for small sample sizes, the bootstrap procedures designed for the data generating process under consideration tend to perform better than all others. For instance, the LPB performs better if the true model is a moving-average model and the AR sieve outperforms the other methods if the underlying process is an autoregression.

Considering the Figures 5 and 6, it becomes obvious that all sets of residuals are uncorrelated, but that contrary to the residuals from Model I and III (first rows) those from Models II and IV (second rows) are not independent. This is confirmed by the correlograms computed from their absolute values and their squares that show significant correlations in the first lag. This explains once again numerically the inconsistency for the LPB and the AR sieve for sample autocovariances in Model II and IV.

Acknowledgments

We are grateful to an Associate Editor and two anonymous referees for their valuable comments and suggestions on an earlier version of this paper. The research of the first author was supported by a fellowship within the PhD-Programme of the German Academic Exchange Service (DAAD)

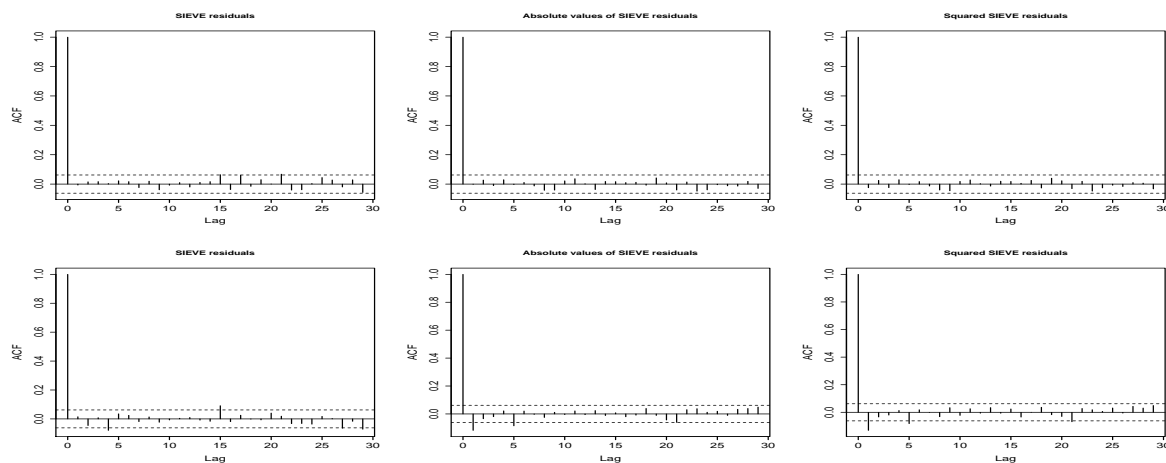


FIGURE 6. Correlograms of residuals (left panels), their absolute values (center panels) and their squares (right panels) computed by SIEVE based on data from Model III (first row) and from Model IV (second row).

when Carsten Jentsch was visiting the University of California, San Diego, USA. The second author's work was partially supported by NSF Grant DMS-10-07513.

REFERENCES

- Brockwell, P. J., Davis, R. A. (1988). Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes Appl.* 22:47–59.
- Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd edn. New York: Springer-Verlag.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3:123–148.
- Jentsch, C., Kreiss, J.-P. (2010). The multiple hybrid bootstrap—Resampling multivariate linear processes. *J. Multivariate Anal.* 101:2320–2345.
- Jentsch, C., Politis, D. N. (2012). The multivariate linear process bootstrap for stationary time series of possibly increasing dimension. Preprint.
- Kreiss, J.-P. (1988). *Asymptotic Statistical Inference for a Class of Stochastic Processes*. Habilitationsschrift, Universität Hamburg.
- Kreiss, J.-P. (1992). Bootstrap Procedures for $AR(\infty)$ Processes. In: Bootstrapping and Related Techniques. Jöckel, K. H., Rothe, G., Sender, W., eds. *Lecture Notes in Economics and Mathematical Systems 376*, Heidelberg: Springer-Verlag, pp. 107–113.
- Kreiss, J.-P., Paparoditis, E. (2011). Bootstrap for dependent data: a review. *J. Korean Statist. Soc.* 40:357–378.
- Kreiss, J.-P., Paparoditis, E., Politis, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* 39:2103–2130.
- Kreiss, J.-P., Meyer, M. (2012). On the range of the validity of the vector autoregressive sieve bootstrap. Preprint.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* 17:1217–1241
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag.
- McMurry, T. L., Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Ser. Anal.* 31:471–482.
- Politis, D. N., Romano, J. P. (1992). A general resampling scheme for triangular arrays of α -mixing random variables with applications to the problem of spectral density estimation. *Ann. Statist.* 20:1985–2007.
- Politis, D. N., Romano, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* 89:1303–1313.
- Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. New York: Wiley.
- Rissanen, J., Barbosa, L. (1969). Properties of infinite covariance matrices and stability of optimum predictors. *Information Sci.* 1:221–236.