

COVARIANCE MATRIX ESTIMATION AND LINEAR PROCESS BOOTSTRAP FOR MULTIVARIATE TIME SERIES OF POSSIBLY INCREASING DIMENSION

BY CARSTEN JENTSCH AND DIMITRIS N. POLITIS

University of Mannheim and University of California at San Diego

Multivariate time series present many challenges, especially when they are high dimensional. The paper's focus is twofold. First, we address the subject of consistently estimating the autocovariance sequence; this is a sequence of matrices that we conveniently stack into one huge matrix. We are then able to show consistency of an estimator based on the so-called *flat-top tapers*; most importantly, the consistency holds true even when the time series dimension is allowed to increase with the sample size. Secondly, we revisit the linear process bootstrap (LPB) procedure proposed by McMurry and Politis (Journal of Time Series Analysis, 2010) for univariate time series. Based on the aforementioned stacked autocovariance matrix estimator, we are able to define a version of the LPB valid for multivariate time series. Under rather general assumptions, we show that our multivariate linear process bootstrap (MLPB) has asymptotic validity for the sample mean in two important cases: (a) when the time series dimension is fixed, and (b) when it is allowed to increase with sample size. As an aside, in case (a) we show that the MLPB works also for spectral density estimators which is a novel result even in the univariate case. We conclude with a simulation study that demonstrates the superiority of the MLPB in some important cases.

1. Introduction. Resampling methods for dependent data such as time series have been studied extensively over the last decades. For an overview of existing bootstrap methods see the monograph of Lahiri (2003), and the review papers by Bühlmann (2002), Paparoditis (2002), Härdle, Horowitz and Kreiss (2003), Politis (2003a) or the recent review paper by Kreiss and Paparoditis (2011). Among the most popular bootstrap procedures in time series analysis we mention the autoregressive (AR) sieve bootstrap [cf. Kreiss (1992, 1999), Bühlmann (1997), Kreiss, Paparoditis and Politis (2011)], and block bootstrap and its variations [cf. Künsch (1989), Liu and Singh (1992), Politis and Romano (1992,1994), etc.]. A recent addition to the available time series bootstrap methods was the linear process bootstrap (LPB) in-

MSC 2010 subject classifications: Primary 62G09; secondary 62M10

Keywords and phrases: Asymptotics, bootstrap, covariance matrix, high dimensional data, multivariate time series, sample mean, spectral density

troduced by McMurry and Politis (2010) who showed its validity for the sample mean for univariate stationary processes without actually assuming linearity of the underlying process.

The main idea of the LPB is to consider the time series data of length n as one large n -dimensional vector and to estimate appropriately the entire covariance structure of this vector. This is executed by using tapered covariance matrix estimators based on flat-top kernels that were defined in Politis (2001). The resulting covariance matrix is used to whiten the data by pre-multiplying the original (centered) data with its inverse Cholesky matrix; a modification of the eigenvalues, if necessary, ensures positive definiteness. This decorrelation property is illustrated in Figures 5 and 6 in Jentsch and Politis (2013). After suitable centering and standardizing, the whitened vector is treated as having independent and identically distributed (i.i.d.) components with zero mean and unit variance. Finally, i.i.d. resampling from this vector and pre-multiplying the corresponding bootstrap vector of residuals with the Cholesky matrix itself results in a bootstrap sample that has (approximately) the same covariance structure as the original time series.

Due to the use of flat-top kernels with compact support, an abruptly dying-out autocovariance structure is induced to the bootstrap residuals. Therefore, the LPB is particularly suitable for—but not limited to—time series of moving average (MA) type. In a sense, the LPB could be considered the closest analog to an MA-sieve bootstrap which is not practically feasible due to nonlinearities in the estimation of the MA parameters. A further similarity of the LPB to MA fitting, at least in the univariate case, is the equivalence of computing the Cholesky decomposition of the covariance matrix to the innovations algorithm; cf. Rissanen and Barbosa (1969), Brockwell and Davis (1988), and Brockwell and Mitchell (1997)—the latter addressing the multivariate case.

Typically, bootstrap methods extend easily from the univariate to the multivariate case, and the same is true for time series bootstrap procedures such as the aforementioned AR-sieve bootstrap and the block bootstrap. By contrast, it has not been clear to date if/how the LPB could be successfully applied in the context of multivariate time series data; a proposal to that effect was described in Jentsch & Politis (2013)—who refer to an earlier preprint of the paper at hand—but it has been unclear to date whether the multivariate LPB is asymptotically consistent and/or it competes well with other methods. Here we attempt to fill this gap: we show how to implement the LPB in a multivariate context, and prove its validity for the sample mean and for spectral density estimators—the latter being a new result even in the univariate case. Note that the limiting distributions of the sample mean

and of kernel spectral density estimators depend only on the second-order moment structure. Hence, it is intuitive that the LPB would be well-suited for such statistics since it generates a linear process in the bootstrap world that mimics well the second-order moment structure of the real world. Furthermore, in the spirit of the times, we consider the possibility that the time series dimension is increasing with sample size, and identify conditions under which the multivariate linear process bootstrap (MLPB) maintains its asymptotic validity even in this case. The key here is to address the subject of consistently estimating the autocovariance sequence; this is a sequence of matrices that we conveniently stack into one huge matrix. We are then able to show consistency of an estimator based on the aforementioned flat-top tapers; most importantly, the consistency holds true even when the time series dimension is allowed to increase with the sample size.

The paper is organized as follows. In Section 2, we introduce the notation of this paper, discuss tapered covariance matrix estimation for multivariate stationary time series and state assumptions used throughout the paper; we then present our results on convergence with respect to operator norm of tapered covariance matrix estimators. The MLPB bootstrap algorithm and some remarks can be found in Section 3 and results concerned with validity of the MLPB for the sample mean and kernel spectral density estimates are summarized in Section 4. Asymptotic results established for the case of increasing time series dimension are stated in Section 5, where operator norm consistency of tapered covariance matrix estimates and a validity result for the sample mean are discussed. A finite-sample simulation study is presented in Section 6. Finally, all proofs, some additional simulations and a real data example on the weighted mean of an increasing number of stock prices taken from the German stock index DAX can be found at the paper's supplementary material [Jentsch and Politis (2014)], which is available at <http://www.math.ucsd.edu/~politis/PAPER/MLPBsupplement.pdf>.

2. Preliminaries. Suppose we consider an \mathbb{R}^d -valued time series process $\{\underline{X}_t, t \in \mathbb{Z}\}$ with $\underline{X}_t = (X_{1,t}, \dots, X_{d,t})^T$ and we have data $\underline{X}_1, \dots, \underline{X}_n$ at hand. The process $\{\underline{X}_t, t \in \mathbb{Z}\}$ is assumed to be strictly stationary and its $(d \times d)$ autocovariance matrix $\mathbf{C}(h) = (C_{ij}(h))_{i,j=1,\dots,d}$ at lag $h \in \mathbb{Z}$ is

$$(2.1) \quad \mathbf{C}(h) = E((\underline{X}_{t+h} - \underline{\mu})(\underline{X}_t - \underline{\mu})^T),$$

where $\underline{\mu} = E(\underline{X}_t)$ and the sample autocovariance $\widehat{\mathbf{C}}(h) = (\widehat{C}_{ij}(h))_{i,j=1,\dots,d}$ at lag $|h| < n$ is defined by

$$(2.2) \quad \widehat{\mathbf{C}}(h) = \frac{1}{n} \sum_{t=\max(1,1-h)}^{\min(n,n-h)} (\underline{X}_{t+h} - \underline{\bar{X}})(\underline{X}_t - \underline{\bar{X}})^T,$$

where $\underline{\bar{X}} = \frac{1}{n} \sum_{t=1}^n \underline{X}_t$ is the d -variate sample mean vector. Here and throughout the paper, all matrix-valued quantities are written as bold letters, all vector-valued quantities are underlined, \mathbf{A}^T indicates the transpose of a matrix \mathbf{A} , $\overline{\mathbf{A}}$ the complex conjugate of \mathbf{A} and $\mathbf{A}^H = \overline{\mathbf{A}^T}$ denotes the transposed conjugate of \mathbf{A} . Note that it is also possible to use unbiased sample autocovariances, i.e., having $n - |h|$ instead of n in the denominator of (2.2). Usually the biased version as defined in (2.2) is preferred because it guarantees a positive semi-definite estimated autocovariance function, but our tapered covariance matrix estimator discussed in Section 2.2 is adjusted in order to become positive definite in any case.

Now, let $\underline{X} = \text{vec}(\mathbf{X}) = (X_1, \dots, X_{dn})^T$ be the dn -dimensional vectorized version of the $(d \times n)$ data matrix $\mathbf{X} = [\underline{X}_1 : \underline{X}_2 : \dots : \underline{X}_n]$ and denote the covariance matrix of \underline{X} , which is symmetric block Toeplitz, by $\mathbf{\Gamma}_{dn}$, that is,

$$(2.3) \quad \mathbf{\Gamma}_{dn} = \begin{pmatrix} \mathbf{C}(i-j) \\ i, j = 1, \dots, n \end{pmatrix} = \begin{pmatrix} \Gamma_{dn}(i, j) \\ i, j = 1, \dots, dn \end{pmatrix},$$

where $\Gamma_{dn}(i, j) = \text{Cov}(X_i, X_j)$ is the covariance between the i th and j th entry of \underline{X} . Note that the second order stationarity of $\{\underline{X}_t, t \in \mathbb{Z}\}$ does *not* imply second order stationary behavior of the vectorized dn -dimensional data sequence \underline{X} . This means that the covariances $\Gamma_{dn}(i, j)$ truly depend on both i and j and not only on the difference $i - j$. However, the following one-to-one correspondence between $\{C_{ij}(h), h \in \mathbb{Z}, i, j = 1, \dots, d\}$ and $\{\Gamma_{dn}(i, j), i, j \in \mathbb{Z}\}$ holds true. Precisely, we have

$$(2.4) \quad \begin{aligned} \Gamma_{dn}(i, j) &= \text{Cov}(X_i, X_j) \\ &= \text{Cov}(X_{m_1(i), m_2(i)}, X_{m_1(j), m_2(j)}) \\ &= C_{\underline{m}_1(i, j)}(m_2(i, j)), \end{aligned}$$

where $\underline{m}_1(i, j) = (m_1(i), m_1(j))$ and $m_2(i, j) = m_2(i) - m_2(j)$ with $m_1(k) = (k - 1) \bmod d + 1$ and $m_2(k) = \lceil k/d \rceil$ and $\lceil x \rceil$ denotes the smallest integer greater or equal to $x \in \mathbb{R}$.

If one is interested in estimating the quantity $\mathbf{\Gamma}_{dn}$, it seems natural to plug in the sample covariances $\widehat{\mathbf{C}}(i - j)$ and $\widehat{\Gamma}_{dn}(i, j) = \widehat{C}_{\underline{m}_1(i, j)}(m_2(i, j))$ in

$\mathbf{\Gamma}_{dn}$ and to use

$$\widehat{\mathbf{\Gamma}}_{dn} = \begin{pmatrix} \widehat{\mathbf{C}}(i-j) \\ i, j = 1, \dots, n \end{pmatrix} = \begin{pmatrix} \widehat{\Gamma}_{dn}(i, j) \\ i, j = 1, \dots, dn \end{pmatrix}.$$

But unfortunately this estimator is *not* a consistent estimator for $\mathbf{\Gamma}_{dn}$ in the sense that the operator norm of $\widehat{\mathbf{\Gamma}}_{dn} - \mathbf{\Gamma}_{dn}$ does not converge to zero. This was shown by Wu and Pourahmadi (2009) and to dissolve this problem in the univariate case, they proposed a banded estimator of the sample covariance matrix to achieve consistency. This has been generalized by McMurry and Politis (2010), who considered general flat-top kernels as weight functions.

In Section 2.2, we follow the paper of McMurry and Politis (2010) and propose a tapered estimator of $\mathbf{\Gamma}_{dn}$ and show its consistency in Theorem 2.1 for the case of multivariate processes. Moreover, we state a modified estimator that is guaranteed to be positive definite for any finite sample size and show its consistency in Theorem 2.2 and of related quantities in Corollary 2.1. But prior to that, we state the assumptions that are used throughout this paper in the following.

2.1. Assumptions.

- (A1) $\{\underline{X}_t, t \in \mathbb{Z}\}$ is an \mathbb{R}^d -valued strictly stationary time series process with mean $E(\underline{X}_t) = \underline{\mu}$ and autocovariances $\mathbf{C}(h)$ defined in (2.1) such that $\sum_{h=-\infty}^{\infty} |h|^g |\mathbf{C}(h)|_1 < \infty$ for some $g \geq 0$ to be further specified. Let $\|\mathbf{A}\|_p = (\sum_{i,j} |a_{ij}|^p)^{1/p}$ for some matrix $\mathbf{A} = (a_{ij})$.
- (A2) There exists a constant $M < \infty$ such that for all $n \in \mathbb{N}$, all h with $|h| < n$ and all $i, j = 1, \dots, d$, we have

$$\left\| \sum_{t=1}^n (X_{i,t+h} - \bar{X}_i)(X_{j,t} - \bar{X}_j) - nC_{ij}(h) \right\|_2 \leq M\sqrt{n}.$$

where $\|\mathbf{A}\|_p = (E(\|\mathbf{A}\|_p^p))^{1/p}$.

- (A3) There exists an $n_0 \in \mathbb{N}$ large enough such that for all $n \geq n_0$ the eigenvalues $\lambda_1, \dots, \lambda_{dn}$ of the $(dn \times dn)$ covariance matrix $\mathbf{\Gamma}_{dn}$ are bounded uniformly away from zero.
- (A4) Define the projection operator $P_k(\underline{X}) = E(\underline{X}|\mathcal{F}_k) - E(\underline{X}|\mathcal{F}_{k-1})$ for $\mathcal{F}_k = \sigma(\underline{X}_t, t \leq k)$ and suppose that for all $i = 1, \dots, d$, we have $\sum_{m=0}^{\infty} \|P_0 X_{i,m}\|_q < \infty$ and $\|\bar{X}_i - \mu_i\|_q = O(\frac{1}{\sqrt{n}})$, respectively, for some $q \geq 2$ to be further specified.
- (A5) For the sample mean, a CLT holds true. That is, we have

$$\sqrt{n}(\bar{\underline{X}} - \underline{\mu}) \xrightarrow{\mathcal{D}} \mathcal{N}(\underline{0}, \mathbf{V}),$$

where $\xrightarrow{\mathcal{D}}$ denotes weak convergence, $\mathcal{N}(\mathbf{0}, \mathbf{V})$ is a normal distribution with zero mean vector and covariance matrix $\mathbf{V} = \sum_{h=-\infty}^{\infty} \mathbf{C}(h)$ with \mathbf{V} positive definite.

(A6) For kernel spectral density estimates $\widehat{f}_{pq}(\omega)$ as defined in (4.2) in Section 4, a CLT holds true. That is, for arbitrary frequencies $0 \leq \omega_1, \dots, \omega_s \leq \pi$, we have that

$$\sqrt{nb} \left(\widehat{f}_{pq}(\omega_j) - f_{pq}(\omega_j) : p, q = 1, \dots, d; j = 1, \dots, s \right)$$

converges to an sd^2 -dimensional normal distribution for $b \rightarrow 0$ and $nb \rightarrow \infty$ such that $nb^5 = O(1)$ as $n \rightarrow \infty$, where the limiting covariance matrix is obtained from

$$\begin{aligned} nbCov \left(\widehat{f}_{pq}(\omega), \widehat{f}_{rs}(\lambda) \right) &= \left(f_{pr}(\omega) \overline{f_{qs}(\omega)} \delta_{\omega, \lambda} + f_{ps}(\omega) \overline{f_{qr}(\omega)} \tau_{0, \pi} \right) \\ &\quad \times \frac{1}{2\pi} \int K^2(u) du + o(1) \end{aligned}$$

and the limiting bias from

$$E \left(\widehat{f}_{pq}(\omega) \right) - f_{pq}(\omega) = b^2 f''_{pq}(\omega) \frac{1}{4\pi} \int K(u) u^2 du + o(b^2)$$

for all $p, q, r, s = 1, \dots, d$, where $\delta_{\omega, \lambda} = 1$ if $\omega = \lambda$ and $\tau_{0, \pi} = 1$ if $\omega = \lambda \in \{0, \pi\}$ and zero otherwise, respectively. Therefore, $\mathbf{f}(\omega)$ is assumed to be component-wise twice differentiable with Lipschitz-continuous second derivatives.

Assumption (A1) is quite standard and the uniform convergence of sample autocovariances in (A2) is satisfied under different types of conditions [cf. Remark 2.1 below] and appears to be a crucial condition here. The uniform boundedness of all eigenvalues away from zero in (A3) is implied by a non-singular spectral density matrix \mathbf{f} of $(\underline{X}_t, t \in \mathbb{Z})$. This follows with (2.3) and the inversion formula from

$$\underline{c}^T \mathbf{\Gamma}_{dn} \underline{c} = \underline{c}^T \left(\int_{-\pi}^{\pi} \mathbf{J}_{\omega}^T \mathbf{f}(\omega) \mathbf{J}_{\omega} d\omega \right) \underline{c} \geq 2\pi |\underline{c}|_2^2 \inf_{\omega} \lambda_{\min}(\mathbf{f}(\omega))$$

for all $\underline{c} \in \mathbb{R}^{dn}$, where $\mathbf{J}_{\omega} = (e^{-i1\omega}, \dots, e^{-in\omega}) \otimes \mathbf{I}_d$ and \otimes denotes the Kronecker product. The requirement of condition (A3) fits into the theory for the univariate autoregressive sieve bootstrap as obtained in Kreiss, Paparoditis and Politis (2011). Similarly, a non-singular spectral density matrix \mathbf{f} implies

positive definiteness of the long-run variance $\mathbf{V} = 2\pi\mathbf{f}(0)$ defined in (A5). Assumption (A4) is for instance fulfilled, if the underlying process is linear or α -mixing with summable mixing coefficients by Ibragimov's inequality [cf. e.g. Davidson (1994), Theorem 14.2]. To achieve validity of the MLPB for the sample mean and for kernel spectral density estimates in Section 4, we have to assume unconditional CLTs in (A5) and (A6), which are satisfied also under certain mixing conditions [cf. Doukhan (1994), Brillinger (1981)], linearity [cf. Brockwell and Davis (1991), Hannan (1970)] or weak dependence [cf. Dedecker et al. (2007)]. Note also that the condition $nb^5 = O(1)$ includes the optimal bandwidth choice $nb^5 \rightarrow C^2$, $C > 0$ for second-order kernels, which leads to a non-vanishing bias in the limiting normal distribution.

REMARK 2.1. *Assumption (A2) is implied by different types of conditions imposed on the underlying process $\{\underline{X}_t, t \in \mathbb{Z}\}$. We present sufficient conditions for (A2) under assumed linearity and under mixing- and weak dependence type conditions. More precisely, (A2) is satisfied if the process $\{\underline{X}_t, t \in \mathbb{Z}\}$ fulfills one of the following conditions:*

(i) *Linearity: Suppose the process is linear, i.e. $\underline{X}_t = \sum_{k=-\infty}^{\infty} \mathbf{B}_k \underline{e}_{t-k}$, $t \in \mathbb{Z}$, where $\{\underline{e}_t, t \in \mathbb{Z}\}$ is an i.i.d. white noise with finite fourth moments $E(e_{i,t}e_{j,t}e_{k,t}e_{l,t}) < \infty$ for all $i, j, k, l = 1, \dots, d$ and the sequence of $(d \times d)$ coefficient matrices $\{\mathbf{B}_k, k \in \mathbb{Z}\}$ is component-wise absolutely summable.*

(ii) *Mixing-type condition: Let*

$$\text{cum}_{a_1, \dots, a_k}(u_1, \dots, u_{k-1}) = \text{cum}(X_{a_1, u_1}, \dots, X_{a_{k-1}, u_{k-1}}, X_{a_k, 0})$$

denote the k th order joint cumulant of $X_{a_1, u_1}, \dots, X_{a_{k-1}, u_{k-1}}, X_{a_k, 0}$ [cf. Brillinger (1981)] and suppose $\sum_{s, h=-\infty}^{\infty} |\text{cum}_{i, j, i, j}(s+h, s, h)| < \infty$ for all $i, j = 1, \dots, d$. Note that this is satisfied if $\{\underline{X}_t, t \in \mathbb{Z}\}$ is α -mixing such that $E(|X_1|_2^{4+\delta}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$ [cf. Shao (2010), p.221].

(iii) *Weak dependence-type condition: Suppose for all $i, j = 1, \dots, d$, we have*

$$|\text{Cov}((X_{i, t+h} - \mu_i)(X_{j, t} - \mu_j), (X_{i, t+s+h} - \mu_i)(X_{j, t+s} - \mu_j))| \leq C \cdot \nu_{s, h},$$

where $C < \infty$ and $(\nu_{s, h}, s, h \in \mathbb{Z})$ is absolutely summable, that is, $\sum_{s, h=-\infty}^{\infty} |\nu_{s, h}| < \infty$ [cf. Dedecker et al. (2007)].

2.2. *Tapered covariance matrix estimation of multiple time series data.*

To adopt the technique of McMurry and Politis (2010), let

$$(2.5) \quad \kappa(x) = \begin{cases} 1, & |x| \leq 1 \\ 0, & |x| > c_\kappa \\ g(|x|), & \text{otherwise} \end{cases}$$

be a so-called *flat-top taper* [cf. Politis (2001)], where $|g(x)| < 1$ and $c_\kappa \geq 1$. The l -scaled version of $\kappa(\cdot)$ is defined by $\kappa_l(x) = \kappa\left(\frac{x}{l}\right)$ for some $l > 0$. As Politis (2011) argues, it is advantageous to having a smooth taper $\kappa(x)$, so the truncated kernel that corresponds to $g(x) = 0$ for all x is not recommended. The simplest example of a continuous taper function κ with $c_\kappa > 1$ is the trapezoid

$$(2.6) \quad \kappa(x) = \begin{cases} 1, & |x| \leq 1 \\ 2 - |x|, & 1 < |x| \leq 2 \\ 0, & |x| > 2 \end{cases}$$

which is used in Section 6 for the simulation study; the trapezoidal taper was first proposed by Politis and Romano (1995) in a spectral estimation setup. Observe also that the banding parameter $l > 0$ does not need to be an integer. The tapered estimator $\widehat{\mathbf{\Gamma}}_{\kappa,l}$ of $\mathbf{\Gamma}_{dn}$ is given by

$$(2.7) \quad \widehat{\mathbf{\Gamma}}_{\kappa,l} = \begin{pmatrix} \kappa_l(i-j)\widehat{\mathbf{C}}(i-j) \\ i, j = 1, \dots, n \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{\Gamma}}_{\kappa,l}(i, j) \\ i, j = 1, \dots, dn \end{pmatrix},$$

where $\widehat{\mathbf{\Gamma}}_{\kappa,l}(i, j) = \widehat{\mathbf{C}}_{m_1(i,j)}^{\kappa,l}(m_2(i, j))$ and $\widehat{\mathbf{C}}_{i,j}^{\kappa,l}(h) = \kappa_l(h)\widehat{\mathbf{C}}_{i,j}(h)$.

The following Theorem 2.1 deals with consistency of the tapered estimator $\widehat{\mathbf{\Gamma}}_{\kappa,l}$ with respect to operator norm convergence. It extends Theorem 1 in McMurry and Politis (2010) to the multivariate case and does not rely on the concept of physical dependence only. The operator norm of a complex-valued $(d \times d)$ matrix \mathbf{A} is defined by

$$\rho(\mathbf{A}) = \max_{\underline{x} \in \mathbb{C}^d: \|\underline{x}\|_2=1} \|\mathbf{A}\underline{x}\|_2,$$

and it is well known that $\rho^2(\mathbf{A}) = \lambda_{\max}(\mathbf{A}^H \mathbf{A}) = \lambda_{\max}(\mathbf{A} \mathbf{A}^H)$, where $\lambda_{\max}(\mathbf{B})$ denotes the largest eigenvalue of a matrix \mathbf{B} [cf. Horn and Johnson (1990), p.296].

THEOREM 2.1. *Suppose that assumptions (A1) with $g = 0$ and (A2) are satisfied. Then, it holds*

$$(2.8) \quad \begin{aligned} \|\rho(\widehat{\Gamma}_{\kappa,l} - \Gamma_{dn})\|_2 &\leq \frac{4Md^2(\lfloor c_{\kappa}l \rfloor + 1)}{\sqrt{n}} + 2 \sum_{h=0}^{\lfloor c_{\kappa}l \rfloor} \frac{|h|}{n} |\mathbf{C}(h)|_1 \\ &+ 2 \sum_{h=l+1}^{n-1} |\mathbf{C}(h)|_1. \end{aligned}$$

The second term on the right-hand side of (2.8) can be represented as $2 \frac{\lfloor c_{\kappa}l \rfloor}{n} \sum_{h=0}^{\lfloor c_{\kappa}l \rfloor} \frac{|h|}{\lfloor c_{\kappa}l \rfloor} |\mathbf{C}(h)|_1$ and vanishes asymptotically due to the Kronecker-Lemma for any choice of l and is of order $o(l/n)$. The third one converges to zero for $l = l(n) \rightarrow \infty$ as $n \rightarrow \infty$ and the leading first term for $l = o(\sqrt{n})$. Hence, for $1/l + l/\sqrt{n} = o(1)$ the right-hand side of (2.8) vanishes asymptotically. However, if $|\mathbf{C}(h)|_1 = 0$ for $|h| > h_0$ for some $h_0 \in \mathbb{N}$, setting $l = h_0$ fixed suffices. In this case, the expression on the right-hand side of (2.8) is of faster order $O(n^{-1/2})$.

As already pointed out by McMurry and Politis (2010), the tapered estimator $\widehat{\Gamma}_{\kappa,l}$ is not guaranteed to be positive semi-definite or even to be positive definite for finite sample sizes. However, $\widehat{\Gamma}_{\kappa,l}$ is at least ‘‘asymptotically positive definite’’ under assumption (A3) and due to (2.8) if $1/l + l/\sqrt{n} = o(1)$ holds. In the following, we require a consistent estimator for Γ_{dn} which is positive definite for all finite sample sizes to be able to compute its Cholesky decomposition for the linear process bootstrap scheme that will be introduced in Section 3 below.

To obtain an estimator of Γ_{dn} related to $\widehat{\Gamma}_{\kappa,l}$ that is assured to be positive definite for all sample sizes, we construct a modified estimator $\widehat{\Gamma}_{\kappa,l}^{\epsilon}$ in the following. Let $\widehat{\mathbf{V}} = \text{diag}(\widehat{\Gamma}_{dn})$ be the diagonal matrix of sample variances and define $\widehat{\mathbf{R}}_{\kappa,l} = \widehat{\mathbf{V}}^{-1/2} \widehat{\Gamma}_{\kappa,l} \widehat{\mathbf{V}}^{-1/2}$. Now, we consider the spectral factorization $\widehat{\mathbf{R}}_{\kappa,l} = \mathbf{S} \mathbf{D} \mathbf{S}^T$, where \mathbf{S} is an $(dn \times dn)$ orthogonal matrix and $\mathbf{D} = \text{diag}(r_1, \dots, r_{dn})$ is the diagonal matrix containing the eigenvalues of $\widehat{\mathbf{R}}_{\kappa,l}$ such that $r_1 \geq r_2 \geq \dots \geq r_{dn}$. It is worth noting that this factorization always exists due to symmetry of $\widehat{\mathbf{R}}_{\kappa,l}$, but that the eigenvalues can be positive, zero or even negative. Now, define

$$(2.9) \quad \widehat{\Gamma}_{\kappa,l}^{\epsilon} = \widehat{\mathbf{V}}^{1/2} \widehat{\mathbf{R}}_{\kappa,l}^{\epsilon} \widehat{\mathbf{V}}^{1/2} = \widehat{\mathbf{V}}^{1/2} \mathbf{S} \mathbf{D}^{\epsilon} \mathbf{S}^T \widehat{\mathbf{V}}^{1/2},$$

where $\mathbf{D}^{\epsilon} = \text{diag}(r_1^{\epsilon}, \dots, r_{dn}^{\epsilon})$ and $r_i^{\epsilon} = \max(r_i, \epsilon n^{-\beta})$. Here, $\beta > 1/2$ and $\epsilon > 0$ are user defined constants that ensure the positive definiteness of $\widehat{\Gamma}_{\kappa,l}^{\epsilon}$. Contrary to the univariate case discussed in McMurry and Politis (2010),

we propose to adjust the eigenvalues of the (equivariant) correlation matrix $\widehat{\mathbf{R}}_{\kappa,l}$ instead of $\widehat{\mathbf{\Gamma}}_{\kappa,l}$, which then comes along without a scaling factor in the definition of r_i^ϵ . Further, note that setting $\epsilon = 0$ leads to a positive semi-definite estimate if $\widehat{\mathbf{\Gamma}}_{\kappa,l}$ is indefinite, which does not suffice for computing the Cholesky decomposition, and also that $\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon$ generally loses the banded shape of $\widehat{\mathbf{\Gamma}}_{\kappa,l}$. Theorem 2.2 below which extends Theorem 3 in McMurry and Politis (2010), shows that the modification of the eigenvalues does affect the convergence results obtained in Theorem 2.1 just slightly.

THEOREM 2.2. *Under the assumptions of Theorem 2.1, it holds*

$$(2.10) \quad \begin{aligned} \|\rho(\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon - \mathbf{\Gamma}_{dn})\|_2 &\leq \frac{8Md^2(\lfloor c_\kappa l \rfloor + 1)}{\sqrt{n}} + 4 \sum_{h=0}^{\lfloor c_\kappa l \rfloor} \frac{|h|}{n} |\mathbf{C}(h)|_1 + 4 \sum_{h=l+1}^{n-1} |\mathbf{C}(h)|_1, \\ &+ \epsilon \max_i C_{ii}(0)n^{-\beta} + O\left(\frac{1}{n^{1/2+\beta}}\right). \end{aligned}$$

In comparison to the upper bound established in Theorem 2.1, two more terms appear on the right-hand side of (2.10) which do converge as well to zero as n tends to infinity. Note that the first three summands that the right-hand sides of (2.8) and (2.10) have in common, remain the leading terms if $\beta > \frac{1}{2}$.

We also need convergence and boundedness in operator norm of quantities related to $\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon$. The required results are summarized in the following corollary.

COROLLARY 2.1. *Under assumptions (A1) with $g = 0$, (A2) and (A3), we have*

(i) $\rho(\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon - \mathbf{\Gamma}_{dn})$ and $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{-1} - \mathbf{\Gamma}_{dn}^{-1})$ are terms of order $O_P(r_{l,n})$, where

$$(2.11) \quad r_{l,n} = \frac{l}{\sqrt{n}} + \sum_{h=l+1}^{\infty} |\mathbf{C}(h)|_1,$$

and $r_{l,n} = o(1)$ if $1/l + l/\sqrt{n} = o(1)$.

(ii) $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{1/2} - \mathbf{\Gamma}_{dn}^{1/2})$ and $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{-1/2} - \mathbf{\Gamma}_{dn}^{-1/2})$ are of order $O_P(\log^2(n)r_{l,n})$ and $\log^2(n)r_{l,n} = o(1)$ if $1/l + \log^2(n)l/\sqrt{n} = o(1)$ and (A1) holds for some $g > 0$.

(iii) $\rho(\mathbf{\Gamma}_{dn})$, $\rho(\mathbf{\Gamma}_{dn}^{-1})$, $\rho(\mathbf{\Gamma}_{dn}^{-1/2})$, $\rho(\mathbf{\Gamma}_{dn}^{1/2})$ are bounded from above and below. $\rho(\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)$, $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{-1})$ and $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{-1/2})$, $\rho((\widehat{\mathbf{\Gamma}}_{\kappa,l}^\epsilon)^{1/2})$ are bounded from above and below (in probability) if $r_{l,n} = o(1)$ and $\log^2(n)r_{l,n} = o(1)$, respectively.

REMARK 2.2. In Section 2.2, we propose to use a global banding parameter l that down-weights the autocovariance matrices for increasing lag, i.e. the entire matrix $\mathbf{C}(h)$ is multiplied with the same $\kappa_l(h)$ in (2.7). However, it is possible to use individual banding parameters l_{pq} for each sequence of entries $\{C_{pq}(h), h \in \mathbb{Z}\}$, $p, q = 1, \dots, d$ as proposed in Politis (2011).

2.3. Selection of tuning parameters.

To get a tapered estimate $\widehat{\Gamma}_{\kappa, l}$ of the covariance matrix Γ_{dn} some parameters have to be chosen by the practitioner. These are the flat-top taper κ and the banding parameter l , which are both responsible for the down-weighting of the empirical autocovariances $\widehat{\mathbf{C}}(h)$ with increasing lag h .

To select a suitable taper κ from the class of functions (2.5), we have to select $c_\kappa \geq 1$ and the function g which determine the range of the decay of κ to zero for $|x| > 1$ and its form over this range, respectively. For some examples of flat-top tapers, compare Politis (2003b, 2011). However, the selection of the banding parameter l appears to be more crucial than choosing the tapering function κ among the family of well-behaved flat-top kernels as discussed in Politis (2011). This is comparable to nonparametric kernel estimation where usually the bandwidth plays a more important role than the shape of the kernel.

We focus on providing an empirical rule for banding parameter selection that has already been used in McMurry and Politis (2010) for the univariate LPB. They make use of an approach primarily proposed in Politis (2003b) to estimate the bandwidth in spectral density estimation which has been generalized to the multivariate case in Politis (2011). In the following, we adopt this technique based on the correlogram/cross-correlogram [cf. Politis (2011, Section 6)] for our purposes. Let

$$(2.12) \quad \widehat{R}_{jk}(h) = \frac{\widehat{C}_{jk}(h)}{\sqrt{\widehat{C}_{jj}(0)\widehat{C}_{kk}(0)}}, \quad j, k = 1, \dots, d$$

be the sample (cross-)correlation between the two univariate time series $(X_{j,t}, t \in \mathbb{Z})$ and $(X_{k,t}, t \in \mathbb{Z})$ at lag $h \in \mathbb{Z}$. Now, define \widehat{q}_{jk} as the smallest nonnegative integer such that

$$|\widehat{R}_{jk}(\widehat{q}_{jk} + h)| < M_0 \sqrt{\log_{10}(n)/n}$$

for $h = 1, \dots, K_n$, where $M_0 > 0$ is a fixed constant, and K_n is a positive, nondecreasing integer-valued function of n such that $K_n = o(\log(n))$. Note that the constant M_0 and the form of K_n are the practitioner's choice. As a

rule of thumb, we refer to Politis (2003b, 2011) who makes the concrete recommendation $M_0 \simeq 2$ and $K_n = \max(5, \sqrt{\log_{10}(n)})$. After having computed \hat{q}_{jk} for all $j, k = 1, \dots, d$, we take

$$(2.13) \quad \hat{l} = \max_{j,k=1,\dots,d} \hat{q}_{jk}$$

as a data-driven *global* choice of the banding parameter l . By setting $\hat{l}_{jk} = \hat{q}_{jk}$, we get data-driven *individual* banding parameter choices as discussed in Remark 2.2. For theoretical justification of this empirical selection of a global cut-off point as the maximum over individual choices and assumptions that lead to successful adaptation, we refer to Theorem 6.1 in Politis (2011).

Note also that for positive definite covariance matrix estimation, i.e. for computing $\hat{\Gamma}_{\kappa,l}^\epsilon$, one has to select two more parameters ϵ and β , which have to be nonnegative and might be set equal to one as suggested in McMurry and Politis (2010).

3. The multivariate linear process bootstrap procedure. In this section, we describe the multivariate linear process bootstrap (MLPB) in detail, discuss some modifications and comment on the special case where the tapered covariance estimator becomes diagonal.

- Step 1. Let \mathbf{X} be the $(d \times n)$ data matrix consisting of \mathbb{R}^d -valued time series data $\underline{X}_1, \dots, \underline{X}_n$ of sample size n . Compute the centered observations $\underline{Y}_t = \underline{X}_t - \underline{\bar{X}}$, where $\underline{\bar{X}} = \frac{1}{n} \sum_{t=1}^n \underline{X}_t$, let \mathbf{Y} be the corresponding $(d \times n)$ matrix of centered observations and define $\underline{Y} = \text{vec}(\mathbf{Y})$ to be the dn -dimensional vectorized version of \mathbf{Y} .
- Step 2. Compute $\underline{W} = (\hat{\Gamma}_{\kappa,l}^\epsilon)^{-1/2} \underline{Y}$, where $(\hat{\Gamma}_{\kappa,l}^\epsilon)^{1/2}$ denotes the lower left triangular matrix \mathbf{L} of the Cholesky decomposition $\hat{\Gamma}_{\kappa,l}^\epsilon = \mathbf{L}\mathbf{L}^T$.
- Step 3. Let \underline{Z} be the standardized version of \underline{W} , that is, $Z_i = \frac{W_i - \bar{W}}{\hat{\sigma}_W}$, $i = 1, \dots, dn$, where $\bar{W} = \frac{1}{dn} \sum_{t=1}^{dn} W_t$ and $\hat{\sigma}_W^2 = \frac{1}{dn} \sum_{t=1}^{dn} (W_t - \bar{W})^2$.
- Step 4. Generate $\underline{Z}^* = (Z_1^*, \dots, Z_{dn}^*)^T$ by i.i.d. resampling from $\{Z_1, \dots, Z_{dn}\}$.
- Step 5. Compute $\underline{Y}^* = (\hat{\Gamma}_{\kappa,l}^\epsilon)^{1/2} \underline{Z}^*$ and let \mathbf{Y}^* be the matrix that is obtained from \underline{Y}^* by putting this vector column-wise into an $(d \times n)$ matrix and denote its columns by $\underline{Y}_1^*, \dots, \underline{Y}_n^*$.

Regarding the Steps 3 and 4 above and due to the multivariate nature of the data, it appears to be even more natural to split the dn -dimensional vector \underline{Z} in Step 3 above in n sub-vectors, to center and standardize them and to apply i.i.d. resampling to these vectors to get \underline{Z}^* . More precisely, Steps 3 and 4 can be replaced by

Step 3'. Let $\underline{Z} = (\underline{Z}_1^T, \dots, \underline{Z}_n^T)^T$ be the standardized version of \underline{W} , that is, $\underline{Z}_i = \widehat{\Sigma}_W^{-1/2}(\underline{W}_i - \overline{W})$, where $\underline{W} = (\underline{W}_1^T, \dots, \underline{W}_n^T)^T$, $\overline{W} = \frac{1}{n} \sum_{t=1}^n \underline{W}_t$ and $\widehat{\Sigma}_W = \frac{1}{n} \sum_{t=1}^n (\underline{W}_t - \overline{W})(\underline{W}_t - \overline{W})^T$.

Step 4'. Generate $\underline{Z}^* = (\underline{Z}_1^*, \dots, \underline{Z}_n^*)^T$ by i.i.d. resampling from $\{\underline{Z}_1, \dots, \underline{Z}_n\}$.

This might preserve more higher order features of the data that are not captured by $\widehat{\Gamma}_{\kappa, l}^\epsilon$. However, comparative simulations (not reported in the paper) indicate that the finite sample performance is only slightly affected by this sub-vector resampling.

REMARK 3.1. *If $0 < l < \frac{1}{c_\kappa}$, the banded covariance matrix estimator $\widehat{\Gamma}_{\kappa, l}^\epsilon$ (and $\widehat{\Gamma}_{\kappa, l}^\epsilon$ as well) becomes diagonal. In this case and if Steps 3' and Steps 4' are used, the LPB as described above is equivalent to the classical i.i.d. bootstrap. Here, note the similarity to the autoregressive sieve bootstrap which boils down to an i.i.d. bootstrap if the autoregressive order is $p = 0$.*

4. Bootstrap consistency for fixed time series dimension.

4.1. Sample mean.

In this section, we establish validity of the MLPB for the sample mean. The following theorem generalizes Theorem 5 of McMurry and Politis (2010) to the multivariate case under somewhat more general conditions.

THEOREM 4.1. *Under assumptions (A1) for some $g > 0$, (A2), (A3), (A4) for $q = 4$, (A5) and $1/l + \log^2(n)l/\sqrt{n} = o(1)$, the MLPB is asymptotically valid for the sample mean \overline{X} , that is,*

$$\sup_{\underline{x} \in \mathbb{R}^d} \left| P \left\{ \sqrt{n} (\overline{X} - \underline{\mu}) \leq \underline{x} \right\} - P^* \left\{ \sqrt{n} \overline{Y}^* \leq \underline{x} \right\} \right| = o_P(1)$$

and $\text{Var}^* \left(\sqrt{n} \overline{Y}^* \right) = \sum_{h=-\infty}^{\infty} \mathbf{C}(h) + o_P(1)$, where $\overline{Y}^* = \frac{1}{n} \sum_{t=1}^n \underline{Y}_t^*$. The short-hand $\underline{x} \leq \underline{y}$ for $\underline{x}, \underline{y} \in \mathbb{R}^d$ is used to denote $x_i \leq y_i$ for all $i = 1, \dots, d$.

4.2. Kernel spectral density estimates.

Here we prove consistency of the MLPB for kernel spectral density matrix estimators; this result is novel even in the univariate case. Let $\mathbf{I}_n(\omega) = \underline{J}_n(\omega) \underline{J}_n^H(\omega)$ the periodogram matrix, where

$$(4.1) \quad \underline{J}_n(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \underline{Y}_t e^{-it\omega}$$

is the discrete Fourier transform (DFT) of $\underline{Y}_1, \dots, \underline{Y}_n$, $\underline{Y}_t = \underline{X}_t - \overline{\underline{X}}$. We define the estimator

$$(4.2) \quad \widehat{\mathbf{f}}(\omega) = \frac{1}{n} \sum_{k=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} K_b(\omega - \omega_k) \mathbf{I}_n(\omega_k)$$

for the spectral density matrix $\mathbf{f}(\omega)$, where $\lfloor x \rfloor$ is the integer part of $x \in \mathbb{R}$, $\omega_k = 2\pi \frac{k}{n}$, $k = -\lfloor \frac{n-1}{2} \rfloor, \dots, \lfloor \frac{n}{2} \rfloor$ are the Fourier frequencies, b is the bandwidth and K is a symmetric and square integrable kernel function $K(\cdot)$ that satisfies $\int K(x) dx = 2\pi$ and $\int K(u) u^2 du < \infty$ and we set $K_b(\cdot) = \frac{1}{b} K(\frac{\cdot}{b})$. Let $\mathbf{I}_n^*(\omega)$ be the bootstrap analogue of $\mathbf{I}_n(\omega)$ based on $\underline{Y}_1^*, \dots, \underline{Y}_n^*$ generated from the MLPB scheme and let $\widehat{\mathbf{f}}^*(\omega)$ be the bootstrap analogue of $\widehat{\mathbf{f}}(\omega)$.

THEOREM 4.2. *Suppose assumptions (A1) with $g \geq 0$ specified below, (A2), (A3), (A4) for $q = 8$ and (A6) are satisfied. If $b \rightarrow 0$ and $nb \rightarrow \infty$ such that $nb^5 = O(1)$ as well as $1/l + \sqrt{bl} \log^2(n) + \sqrt{nb} \log^2(n)/l^g$ and $1/k + bk^4 + \sqrt{nb} \log^2(n)/k^g$ for some sequence $k = k(n)$, the MLPB is asymptotically valid for kernel spectral density estimates $\widehat{\mathbf{f}}(\omega)$. That is, for all $s \in \mathbb{N}$ and arbitrary frequencies $0 \leq \omega_1, \dots, \omega_s \leq \pi$ (not necessarily Fourier frequencies), it holds*

$$\begin{aligned} & \sup_{\underline{x} \in \mathbb{R}^{d^2 s}} \left| P \left\{ \left(\sqrt{nb} (\widehat{f}_{pq}(\omega_j) - f_{pq}(\omega_j)) : p, q = 1, \dots, d; j = 1, \dots, s \right) \leq \underline{x} \right\} \right. \\ & \quad \left. - P^* \left\{ \left(\sqrt{nb} (\widehat{f}_{pq}^*(\omega_j) - \check{f}_{pq}(\omega_j)) : p, q = 1, \dots, d; j = 1, \dots, s \right) \leq \underline{x} \right\} \right| \\ & = o_P(1), \end{aligned}$$

where $\check{f}_{pq}(\omega) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \kappa_l(h) \widehat{C}_{pq}(h) e^{-ih\omega}$ and, in particular,

$$\begin{aligned} nb \text{Cov}^* \left(\widehat{f}_{pq}^*(\omega), \widehat{f}_{rs}^*(\lambda) \right) & = \left(f_{pr}(\omega) \overline{f_{qs}(\omega)} \delta_{\omega, \lambda} + f_{ps}(\omega) \overline{f_{qr}(\omega)} \tau_{0, \pi} \right) \\ & \quad \times \frac{1}{2\pi} \int K^2(u) du + o_P(1), \end{aligned}$$

and $E^*(\widehat{f}_{pq}^*(\omega)) - \check{f}_{pq}(\omega) = b^2 f_{pq}''(\omega) \frac{1}{4\pi} \int K(u) u^2 du + o_P(b^2)$, for all $p, q, r, s = 1, \dots, d$ and all $\omega, \lambda \in [0, \pi]$, respectively.

4.3. Other statistics and LPB-of-blocks bootstrap.

For statistics T_n contained in the broad class of functions of generalized means, Jentsch and Politis (2013) discussed how by using a preliminary

blocking scheme tailor-made for a specific statistic of interest, the MLPB can be shown to be consistent. This class of statistics contains estimates T_n of $w(\vartheta)$ with $\vartheta = E(g(\underline{X}_t, \dots, \underline{X}_{t+m-1}))$ such that

$$T_n = w \left\{ \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(\underline{X}_t, \dots, \underline{X}_{t+m-1}) \right\},$$

for some sufficiently smooth functions $g : \mathbb{R}^{d \times m} \rightarrow \mathbb{R}^k$, $w : \mathbb{R}^k \rightarrow \mathbb{R}$ and fixed $m \in \mathbb{N}$. They propose to block the data first according to the known function g and to apply then the (M)LPB to the blocked data. More precisely, the multivariate LPB-of-blocks bootstrap is as follows:

- Step 1. Define $\tilde{X}_t := g(\underline{X}_t, \dots, \underline{X}_{t+m-1})$ and let $\tilde{X}_1, \dots, \tilde{X}_{n-m+1}$ be the set of blocked data.
- Step 2. Apply the MLPB scheme of Section 3 to the k -dimensional blocked data $\tilde{X}_1, \dots, \tilde{X}_{n-m+1}$ to get bootstrap observations $\tilde{X}_1^*, \dots, \tilde{X}_{n-m+1}^*$.
- 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 $\sqrt{n}\{T_n - w(\vartheta)\}$ by the empirical distribution of $\sqrt{n}\{T_{n,1}^* - T_n\}, \dots, \sqrt{n}\{T_{n,B}^* - T_n\}$.

The validity of the multivariate LPB-of-blocks bootstrap for some statistic T_n can be verified by checking the assumptions of Theorem 4.1 for the sample mean of the new process $\{\tilde{X}_t, t \in \mathbb{Z}\}$.

5. Asymptotic results for increasing time series dimension.

In this section, we consider the case when the time series dimension d is allowed to increase with the sample size n , i.e. $d = d(n) \rightarrow \infty$ as $n \rightarrow \infty$. In particular, we show consistency of tapered covariance matrix estimates and derive rates that allow for an asymptotic validity result of the MLPB for the sample mean in this case.

The recent paper by Cai, Ren and Zhou (2013) gives a thorough discussion of the estimation of Toeplitz covariance matrices for univariate time series. In their setup, that covers also the possibility of having multiple datasets from the same data generating process, Cai, Ren and Zhou (2013) establish the optimal rates of convergence using the two simple flat-top kernels discussed in Section 2.2, namely the truncated (i.e., case of pure banding—no tapering), and the trapezoid taper. When the strength of dependence is quantified via a smoothness condition on the spectral density, they show that the trapezoid is superior to the truncated taper, thus confirming the intuitive recommendations of Politis (2011). The asymptotic theory of Cai,

Ren and Zhou (2013) allows for increasing number of time series and increasing sample size, but their framework does not contain the multivariate time series case neither for fixed nor for increasing time series dimension, which will be discussed in this section.

Note that Theorem 1 in McMurry and Politis (2010) for the univariate case, as well as our Theorem 2.1 for the multivariate case of fixed time series dimension, give upper bounds that are quite sharp, coming within a log-term to the (Gaussian) optimal rate found in Theorem 2 of Cai, Ren and Zhou (2013).

Instead of assumptions (A1)–(A5) that have been introduced in Section 2.1 and used in Theorem 4.1 to obtain bootstrap consistency for the sample mean for fixed dimension d , we impose the following conditions on the sequence of time series process $(\{\underline{X}_t^{(n)}, t \in \mathbb{Z}\})_{n \in \mathbb{N}}$ of now increasing dimension.

5.1. Assumptions.

(A1') $(\{\underline{X}_t = (X_{1,t}, \dots, X_{d(n),t})^T, t \in \mathbb{Z}\})_{n \in \mathbb{N}}$ is a sequence of $\mathbb{R}^{d(n)}$ -valued strictly stationary time series processes with mean vectors $E(\underline{X}_t) = \underline{\mu} = (\mu_1, \dots, \mu_{d(n)})$ and autocovariances $\mathbf{C}(h) = (C_{ij}(h))_{i,j=1,\dots,d(n)}$ defined as in (2.1). Here, $(d(n))_{n \in \mathbb{N}}$ is a non-decreasing sequence of positive integers such that $d(n) \rightarrow \infty$ as $n \rightarrow \infty$ and, further, suppose

$$\sum_{h=-\infty}^{\infty} \left\{ \sup_{n \in \mathbb{N}} \sup_{i,j=1,\dots,d(n)} |h|^g |C_{ij}(h)| \right\} < \infty$$

for some $g \geq 0$ to be further specified.

(A2') There exists a constant $M' < \infty$ such that for all $n \in \mathbb{N}$ and all h with $|h| < n$, we have

$$\sup_{i,j=1,\dots,d(n)} \left\| \sum_{t=1}^n (X_{i,t+h} - \bar{X}_i)(X_{j,t} - \bar{X}_j) - nC_{ij}(h) \right\|_2 \leq M' \sqrt{n}.$$

(A3') There exists an $n_0 \in \mathbb{N}$ large enough such that for all $n \geq n_0$ and all $d \geq d_0 = d(n_0)$ the eigenvalues $\lambda_1, \dots, \lambda_{dn}$ of the $(dn \times dn)$ covariance matrix $\mathbf{\Gamma}_{dn}$ are bounded uniformly away from zero and from above.

(A4') Define the sequence of projection operators $P_k^{(n)}(\underline{X}) = E(\underline{X} | \mathcal{F}_k^{(n)}) - E(\underline{X} | \mathcal{F}_{k-1}^{(n)})$ for $\mathcal{F}_k^{(n)} = \sigma(\underline{X}_t, t \leq k)$ and suppose

$$\sum_{m=0}^{\infty} \left\{ \sup_{n \in \mathbb{N}} \sup_{i=1,\dots,d(n)} \|P_0^{(n)} X_{i,m}\|_4 \right\} < \infty \quad \text{and} \quad \sup_{n \in \mathbb{N}} \sup_{i=1,\dots,d(n)} \|\bar{X}_i - \mu_i\|_4 = O(n^{-1/2}).$$

(A5') For the sample mean, a Cramér-Wold-type CLT holds true. That is, for any real-valued sequence $\underline{b} = \underline{b}(d(n))$ of $d(n)$ -dimensional vectors with $0 < M_1 \leq |\underline{b}(d(n))|_2^2 \leq M_2 < \infty$ for all $n \in \mathbb{N}$ and $v^2 = v_{d(n)}^2 = \text{Var}(\sqrt{n}(\underline{b}^T(\underline{X} - \underline{\mu})))$, we have

$$\sqrt{n}(\underline{b}^T(\underline{X} - \underline{\mu})) / v \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

The assumptions (A1')–(A4') are uniform analogues of (A1)–(A4), which are required here to tackle the increasing time series dimension d . In particular, (A1') implies

$$(5.1) \quad \sum_{h=-\infty}^{\infty} |\mathbf{C}(h)|_1 = O(d^2).$$

Observe also that the autocovariances $C_{ij}(h)$ are assumed to decay with increasing lag h , i.e. in time direction, but they are not assumed to decay with increasing $|i - j|$, i.e., with respect to increasing time series dimension. Therefore, we have to make use of square summable sequences in (A5') to get a CLT result. This techniques has been used e.g. by Lewis and Reinsel (1985) and Goncalves and Kilian (2007) to establish central limit results for the estimation of an increasing number of autoregressive coefficients. A simple sufficient condition for (A5') is e.g. the case of $(\{\underline{X}_t = (X_{1,t}, \dots, X_{d(n),t})^T, t \in \mathbb{Z}\})_{n \in \mathbb{N}}$ being a sequence of i.i.d. Gaussian processes with eigenvalues of $E(\underline{X}_t \underline{X}_t^T)$ bounded uniformly from above and away from zero.

5.2. Operator norm convergence for increasing time series dimension.

The following theorem generalizes the results of Theorems 2.1 and 2.2 and of Corollary 2.1 to the case where $d = d(n)$ is allowed to increase with the sample size. In contrast to the case of a stationary spatial process on the plane \mathbb{Z}^2 (where a data matrix is observed that grows in both directions asymptotically as in our setting), we do not assume that the autocovariance matrix decays in all directions. Therefore, to be able to establish a meaningful theory, we have to replace (A1)–(A5) by the uniform analogues (A1')–(A5') and due to (5.1), an additional factor d^2 turns up in the convergence rate and has to be taken into account.

THEOREM 5.1. *Under assumptions (A1') with $g \geq 0$ specified below, (A2') and (A3'), we have*

- (i) $\rho(\widehat{\Gamma}_{\kappa,l}^\epsilon - \Gamma_{dn})$ and $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{-1} - \Gamma_{dn}^{-1})$ are terms of order $O_P(d^2\tilde{r}_{l,n})$, where

$$(5.2) \quad \tilde{r}_{l,n} = \frac{l}{\sqrt{n}} + \sum_{h=l+1}^{\infty} \left\{ \sup_{n \in \mathbb{N}} \sup_{i,j=1,\dots,d(n)} |C_{ij}(h)| \right\},$$

and $d^2\tilde{r}_{l,n} = o(1)$ if $1/l + d^2l/\sqrt{n} + d^2/l^g = o(1)$.

- (ii) $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{1/2} - \Gamma_{dn}^{1/2})$ and $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{-1/2} - \Gamma_{dn}^{-1/2})$ are both terms of order $O_P(\log^2(dn)d^2\tilde{r}_{l,n})$ and $\log^2(dn)d^2\tilde{r}_{l,n} = o(1)$ if $1/l + \log^2(dn)d^2l/\sqrt{n} + \log^2(dn)d^2/l^g = o(1)$.
- (iii) $\rho(\Gamma_{dn})$, $\rho(\Gamma_{dn}^{-1})$, $\rho(\Gamma_{dn}^{-1/2})$ and $\rho(\Gamma_{dn}^{1/2})$ are bounded from above and below. $\rho(\widehat{\Gamma}_{\kappa,l}^\epsilon)$ and $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{-1})$ as well as $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{-1/2})$ and $\rho((\widehat{\Gamma}_{\kappa,l}^\epsilon)^{1/2})$ are bounded from above and below in probability if $d^2\tilde{r}_{l,n} = o(1)$ and $\log^2(dn)d^2\tilde{r}_{l,n} = o(1)$, respectively.

The required rates for the banding parameter l and the time series dimension d to get operator norm consistency $\rho(\widehat{\Gamma}_{\kappa,l}^\epsilon - \Gamma_{dn}) = o_P(1)$ can be interpreted nicely. If g is chosen to be large enough, d^2l/\sqrt{n} becomes the leading term and there is a trade-off between capturing more dependence of the time series in time direction (large l) and growing dimension of the time series in cross-sectional direction (large d).

5.3. Bootstrap validity for increasing time series dimension.

The subsequent theorem is a Cramér-Wold-type generalization of Theorem 4.1 to the case where $d = d(n)$ is allowed to grow at an appropriate rate with the sample size. To tackle the increasing time series dimension and to prove such a CLT result, we have to make use of appropriate sequences of square summable vectors $\underline{b} = \underline{b}(d(n))$ as described in (A5') above.

THEOREM 5.2. *Under assumptions (A1') with $g \geq 0$ specified below, (A2'), (A3'), (A4') for $q = 4$, (A5') as well as $1/l + \log^2(dn)d^2l/\sqrt{n} + \log^2(dn)d^2/l^g = o(1)$ and $1/k + d^5k^4/n + \log^2(dn)d^2/k^g = o(1)$ for some sequence $k = k(n)$, the MLPB is asymptotically valid for the sample mean \overline{X} . That is, for any real-valued sequence $\underline{b} = \underline{b}(d(n))$ of $d(n)$ -dimensional vectors with $0 < M_1 \leq \|\underline{b}(d(n))\|_2^2 \leq M_2 < \infty$ for all $n \in \mathbb{N}$ and $\hat{v}^2 = \hat{v}_{d(n)}^2 = \text{Var}^*(\sqrt{n}(\underline{b}^T \overline{Y}^*))$, we have*

$$\sup_{x \in \mathbb{R}} \left| P \left\{ \sqrt{n} \left(\underline{b}^T (\overline{X} - \underline{\mu}) \right) / v \leq x \right\} - P^* \left\{ \sqrt{n} \left(\underline{b}^T \overline{Y}^* \right) / \hat{v} \leq x \right\} \right| = o_P(1)$$

and $|v^2 - \hat{v}^2| = o_P(1)$.

5.4. Reduction of computation time.

In practice, the computational requirements can become very demanding for large d and n . In this case, we suggest to split the data vector \underline{X} in few subsamples $\underline{X}^{(1)}, \dots, \underline{X}^{(S)}$, say, and to apply the MLPB scheme to each subsample separately. This operation can be justified by the fact that dependence structure is distorted only few times. Precisely, we suggest the following procedure:

- Step 1. For small $S \in \mathbb{N}$, define $n_{sub} = \lceil n/S \rceil$ and $N_{sub} = dn_{sub}$ such that $SN_{sub} \geq N$ and let $\underline{X}^{(i)} = (\underline{X}_{(i-1)n_{sub}+1}^T, \dots, \underline{X}_{in_{sub}}^T)^T$, $i = 1, \dots, S$, where $\underline{X}^{(S)}$ is filled up with zeros if $SN_{sub} > N$.
- Step 2. Apply the MLPB bootstrap scheme as described in Section 3 separately to the subsamples $\underline{X}^{(1)}, \dots, \underline{X}^{(S)}$ to get $\underline{Y}^{(1)*}, \dots, \underline{Y}^{(S)*}$.
- Step 3. Put $\underline{X}^{(1)*}, \dots, \underline{X}^{(S)*}$ end-to-end together and discard the last $SN_{sub} - N$ values to get \underline{Y}^* and \mathbf{Y}^* .

Here, computationally demanding operations as eigenvalue decomposition, Cholesky decomposition and matrix inversion have to be executed only for lower-dimensional matrices, such that the algorithm above is capable to reduce the computation time considerably. Further, to regain efficiency, we propose to use the pooled sample mean $\bar{\underline{X}}$ for centering and $\hat{\Gamma}_{\kappa, l, N_{sub}}^\epsilon$ for whitening and re-introducing correlation structure for *all* subsamples in Step 2. Here, $\hat{\Gamma}_{\kappa, l, N_{sub}}^\epsilon$ is obtained analogously to (2.9), but based on the upper-left $(N_{sub} \times N_{sub})$ sub-matrix of $\hat{\Gamma}_{\kappa, l}$.

6. Simulations. In this section we compare systematically the performance of the multivariate linear process bootstrap (MLPB) to that of the vector-autoregressive sieve bootstrap (ARSieve), the moving block bootstrap (MBB) and the tapered block bootstrap (TBB) by means of simulation. In order to make such a comparison, we have chosen a statistic for which all methods lead to asymptotically correct approximations. Being interested in the distribution of the sample mean, we compare the aforementioned bootstrap methods by plotting

- a) root mean squared errors (RMSE) for estimating the variances of $\sqrt{n}(\bar{\underline{X}} - \underline{\mu})$
- b) coverage rates (CR) of 95% bootstrap confidence intervals for the components of $\underline{\mu}$

for two data generating processes (DGPs) and three sample sizes in two different setups. First, in Section 6.1, we compare the performance of all aforementioned bootstraps with respect to (wrt) tuning parameter choice.

These are the banding parameter l (MLPB), the autoregressive order p (AR-sieve) and the block length s (MBB, TBB). Furthermore, we report RMSE and CR for data-adaptively chosen tuning parameters to investigate how accurate automatic selection procedures can work in practice. Second, in Section 6.2, we investigate the effect of the time series dimension d on the performance of the different bootstrap approaches.

For each case, we have generated $T = 500$ time series and $B = 500$ bootstrap replications have been used in each step. For a), the exact covariance matrix of $\sqrt{n}(\bar{X} - \mu)$ is estimated by 20,000 Monte Carlo replications. Further, we use the trapezoidal kernel defined in (2.6) to taper the sample covariance matrix for the MLPB and the blocks for the TBB. To correct the covariance matrix estimator $\hat{\Gamma}_{\kappa,l}$ to be positive definite, if necessary, we set $\epsilon = 1$ and $\beta = 1$ to get $\hat{\Gamma}_{\kappa,l}^\epsilon$. This choice has already been used by McMurry and Politis (2010) and simulation results (not reported in this paper) indicate that the performance of the MLPB reacts only slightly to this choice. We have used the sub-vector resampling scheme, i.e. Steps 3' and 4' described in Section 3.

Some additional simulation results and a real data application of the MLPB to the weighted mean of an increasing number of German stock prices taken from the DAX index can be found in a supplementary material to this paper [Jentsch and Politis (2014)]. The R code is available at http://www.math.ucsd.edu/~politis/SOFT/function_MLPB.R.

6.1. *Bootstrap performance: the effect of tuning parameter choice.* We consider realizations $\underline{X}_1, \dots, \underline{X}_n$ of length $n = 100, 200, 500$ from two bivariate ($d = 2$) DGPs. Precisely, we study a first order vector moving average process

$$\text{VMA(1) model } \underline{X}_t = \mathbf{A}\underline{e}_{t-1} + \underline{e}_t$$

and a first order vector autoregressive process

$$\text{VAR(1) model } \underline{X}_t = \mathbf{A}\underline{X}_{t-1} + \underline{e}_t,$$

where $\underline{e}_t \sim \mathcal{N}(0, \Sigma)$ is a normally distributed i.i.d. white noise process and

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} 0.9 & -0.4 \\ 0 & 0.5 \end{pmatrix}$$

have been used in all cases. It is worth noting, that (asymptotically) all bootstrap procedures under consideration yield valid approximations for both models above. For the VMA(1) model, MLPB is valid for all (sufficiently small) choices of banding parameters $l \geq 1$, but ARsieve is valid

only asymptotically for $p = p(n)$ tending to infinity at an appropriate rate with increasing sample size n . This relationship of MLPB and ARsieve is reversed for the VAR(1) model. For the MBB and the TBB, the block length has to increase with the sample size for both DGPs.

Additionally to the results for tuning parameters $l, p, s \in \{1, \dots, 20\}$, we show also RMSE and CR for tuning parameters chosen by automatic selection procedures in Figures 1 and 2. For the MLPB, we report results for data-adaptively chosen global and individual banding parameters as discussed in Section 2.3. For the ARsieve, the order of the VAR model fitted to the data has been chosen by using the **R** routine $VAR()$ contained in the package **vars** with $lag.max = \sqrt{n/\log(n)}$. The block length is chosen by using the **R** routine $b.star()$ contained in the package **np**. In Figure 1 and 2, we report only the results corresponding to the first component of the sample mean as those for the second component lead qualitatively to the same results. We show them in the supplementary material, that contains also corresponding simulation results for a normal white noise DGP.

For data generated by the VMA(1) model, Figure 1 shows that the MLPB outperforms AR sieve, MBB and TBB for adequate tuning parameter choice, that is, $l \approx 1$. In this case, the MLPB behaves generally superior wrt RMSE and CR to the other bootstrap methods for all tuning parameter choices of p and s . This was not unexpected since, by design, the MLPB can approximate very efficiently the covariance structure of moving average processes. Nevertheless, due to the fact that all proposed bootstrap schemes are valid at least asymptotically, ARsieve gets rid of its bias with increasing order p , but at the expense of increasing variability and consequently also increasing RMSE. MLPB with data-adaptively chosen banding parameter performs quite well, where the individual choice tends to perform superior to the global choice in most cases. In comparison, MBB and TBB seem to perform quite well for adequate block length, but they lose in terms of RMSE as well as CR performance if the block length is chosen automatically.

The data from the VAR(1) model is highly persistent due to the coefficient $A_{11} = 0.9$ near to unity. This leads to autocovariances that are rather slowly decreasing with increasing lag and, consequently, to large variances of $\sqrt{n}(\bar{X} - \mu)$. Figure 2 shows that ARsieve outperforms MLPB, MBB and TBB wrt to CR for small AR orders $p \approx 1$. This was to be expected since the underlying VAR(1) model is captured well by ARsieve even with finite sample size. But the picture appears to be different wrt RMSE. Here, MLPB may perform superior for adequate tuning parameter choice, but this effect can be explained by the very small variance that compensates its large bias in comparison to the AR sieve [not reported here] leading to a

smaller RMSE. This phenomenon is also illustrated by the poor performance of MLPB wrt to CR for small choices of l . However, more surprising is the rather good performance of the MLPB if the banding parameter is chosen data-adaptively, where the MLPB appears to be comparable to the ARsieve in terms of RMSE and at least close wrt CR. Further, as observed already for the VMA(1) model in Figure 1, the individual banding parameter choice generally tends to outperform the global choice here again. Similarly, it can be seen here that the performance of ARsieve worsens with increasing p at the expense of increasing variability. The block bootstraps MBB and TBB appear to be clearly inferior to MLPB and AR sieve particularly wrt CR, but also wrt to RMSE if tuning parameters are chosen automatically.

6.2. *Bootstrap performance: the effect of larger time series dimension.*

We consider d -dimensional realizations $\underline{X}_1, \dots, \underline{X}_n$ with $n = 100, 200, 500$ from two DGPs of several dimensions. Precisely, we study first order vector moving average processes

$$\text{VMA}_d(1) \text{ model } \underline{X}_t = \mathbf{A}\underline{e}_{t-1} + \underline{e}_t$$

and first order vector autoregressive processes

$$\text{VAR}_d(1) \text{ model } \underline{X}_t = \mathbf{A}\underline{X}_{t-1} + \underline{e}_t$$

of dimension $d \in \{2, \dots, 10\}$, where $\underline{e}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_d)$ is a d -dimensional normally distributed i.i.d. white noise process and $\boldsymbol{\Sigma} = (\Sigma_{ij})$ and $\mathbf{A} = (A_{ij})$ are such that

$$\boldsymbol{\Sigma}_{ij} = \begin{cases} 1, & i = j \\ 0.5, & |i - j| = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbf{A}_{ij} = \begin{cases} 0.9, & i = j, (i + 1)/2 \in \mathbb{N} \\ 0.5, & i = j, i/2 \in \mathbb{N} \\ -0.4, & i + 1 = j \\ 0, & \text{otherwise} \end{cases}.$$

Observe that the VMA(1) and VAR(1) models considered in Section 6.1 are included in this setup for $d = 2$.

In Figures 3 and 4, we compare the performance of MLPB, ARsieve, MBB and TBB for the DGPs above using RMSE and CR averaged over all d time series coordinates. Precisely, we compute RMSE individually for the estimates of $\text{Var}(\sqrt{n}(\bar{X}_i - \mu_i))$, $i = 1, \dots, d$ and plot the averages in the upper half of Figures 3 and 4. Similarly, we plot averages of individually calculated CR of bootstrap confidence intervals for μ_i , $i = 1, \dots, d$ in the lower halves. All tuning parameters are chosen data-based and optimal as

described in Section 6.1 and to reduce computation time, the less demanding algorithm as described in Section 5.4 with $S = dn/500$ is used.

For the VMA(1) DGPs in Figure 3, the MLPB with individual banding parameter choice outperforms the other approaches essentially for all time series dimension under consideration wrt to averaged RMSE and CR. In particular, larger time series dimensions do not seem to have a large effect on the performance of all bootstraps for the VMA(1) DGPs with the only exception of the MLPB with global banding parameter choice. In particular, the latter is clearly inferior in comparison to the MLPB with individually chosen banding parameter, which might be explained by sparsity of the covariance matrix Γ_{dn} .

In Figure 4, for the VAR(1) DGPs, the picture is different to the VMA(1) case above. The influence of larger time series dimension on RMSE (and less pronounced for CR) performance is much more pronounced and clearly visible. In particular, the RMSE blows up with increasing dimension d for all four bootstrap methods, which is due to the also increasing variance of the process. Note that the zig-zag shape of the RMSE curves is due to the back and forth switching from 0.9 to 0.5 on the diagonal of A . As already observed for the VMA(1) DGPs, the MLPB with individual banding parameter choice again performs best over essentially all time series dimensions wrt to average RMSE and average CR. In particular, MLPB with individual choice is superior to the global choice. Here, the good performance of the MLPB is somewhat surprising as the VAR(1) DGPs have rather slowly decreasing autocovariance structure, where we expected an ARsieve to be more suitable.

ACKNOWLEDGEMENTS

The research of the first author was supported by a fellowship within the PhD-Program of the German Academic Exchange Service (DAAD) when Carsten Jentsch was visiting the University of California, San Diego, USA. The research of Dimitris Politis was partially supported by NSF grants DMS 13-08319 and DMS 12-23137. The authors thank Timothy McMurry for his helpful advice on the univariate case and three anonymous referees and the editor who helped to significantly improve the presentation of the paper.

SUPPLEMENTARY MATERIAL

Additional proofs, simulations and a real data example
(doi: [10.1214/00-AOASXXXXSUPP](https://doi.org/10.1214/00-AOASXXXXSUPP); .pdf). In the supplementary material we provide proofs, additional supporting simulations and an application of the MLPB to German stock index data.

The supplementary material to this paper is also available online at <http://www.math.ucsd.edu/~politiss/PAPER/MLPBsupplement.pdf>

REFERENCES

- [1] BRILLINGER, D. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.
- [2] BROCKWELL, P.J. and DAVIS, R.A. (1988). Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes and Their Applications* **22**, 47–59.
- [3] BROCKWELL, P.J. and DAVIS, R.A. (1991). *Time series: theory and methods*. 2nd ed. Springer-Verlag, Berlin.
- [4] BROCKWELL, P.J. and MITCHELL, H. (1997). Estimation of the coefficients of a multivariate linear filter using the innovations algorithm. *J. Time Ser. Anal.* **18** 157–179.
- [5] BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.
- [6] BÜHLMANN, P. (2002). Bootstraps for Time Series. *Statist. Sci.* **17** 52–72.
- [7] CAI, T., REN, Z., and ZHOU, H. (2013). Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices. *Probab. Theory Relat. Fields* **156** 101–143.
- [8] DAVIDSON, J. (1994). *Stochastic Limit Theory*. Oxford University Press, New York.
- [9] DEDECKER, J., DOUKHAN, P., LANG, G., LEON, J.R.R., LOUHICHI, S., and PRIEUR, C. (2007). *Weak dependence. With examples and applications. Lecture Notes in Statistics 190*. Springer, New York.
- [10] DOUKHAN, P. (1994). *Mixing: Properties and examples. Lecture Notes in Statistics 85*. Springer, New York.
- [11] GONCALVES, S. and KILIAN, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews* **26** 609–641.
- [12] HANNAN, E.J. (1970). *Multiple Time Series*. Wiley, New York.
- [13] HÄRDLE, W., HOROWITZ, J., and KREISS, J.-P. (2003). Bootstrap Methods for Time Series. *Int. Statist. Rev.* **71** 435–459.
- [14] HORN, R.A. and JOHNSON, C.R. (1990). *Matrix analysis*. Cambridge University Press, New York.
- [15] JENTSCH, C. and POLITIS, D.N. (2013). Valid resampling of higher order statistics using linear process bootstrap and autoregressive sieve bootstrap. *Commun. Stat., Theory Methods* **42** 1277–1293.
- [16] JENTSCH, C. and POLITIS, D.N. (2014). Supplement to ‘Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension’.
- [17] KREISS, J.-P. (1992). Bootstrap Procedures for AR(∞) 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.
- [18] KREISS, J.-P. (1999). *Residual and Wild Bootstrap for Infinite Order Autoregression*. Unpublished manuscript.
- [19] KREISS, J.-P. and PAPARODITIS, E. (2011). Bootstrap for dependent data: a review. *J. Korean Statist. Soc.* **40** 357–378.
- [20] KREISS, J.-P., PAPARODITIS, E., and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130.
- [21] KÜNSCH, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Statist.* **17** 1217–1241.
- [22] LAHIRI, S.N. (2003). *Resampling methods for dependent data*. Springer, New York.

- [23] LEWIS, R. and REINSEL, G. (1985). Prediction of multivariate time series of autoregressive model fitting. *Journal of Multivariate Analysis*. **16** 393–411
- [24] LIU, R.Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In: *Exploring the Limits of Bootstrap* (Eds.: LePage, R. and Billard, L.), pp. 225–248. Wiley, New York.
- [25] MCMURRY, T.L. and POLITIS, D.N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Ser. Anal.* **31** 471–482. CORRIGENDUM: *J. TIME SER. ANAL.* **33**, 2012.
- [26] PAPANODITIS, E. (2002). Frequency Domain Bootstrap for Time Series. In *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sorensen, eds.) 365–381. Birkhäuser, Boston.
- [27] POLITIS, D.N. (2001). On nonparametric function estimation with infinite-order flat-top kernels, in *Probability and Statistical Models with applications*, Ch. Charalambides et al. (Eds.), pp. 469–483. Chapman and Hall/CRC, Boca Raton.
- [28] POLITIS, D.N. (2003a). The impact of bootstrap methods on time series analysis. *Statistical Science* **18** 219–230.
- [29] POLITIS, D.N. (2003b). Adaptive bandwidth choice. *J. Nonparametric Statist.* **15** 517–533.
- [30] POLITIS, D.N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744.
- [31] POLITIS, D.N. and ROMANO, J.P. (1992). A General Resampling Scheme for Triangular Arrays of alpha-mixing Random Variables with Application to the Problem of Spectral Density Estimation. *Ann. Statist.* **20** 1985–2007.
- [32] POLITIS, D.N. and ROMANO, J.P. (1994). Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Applications to the Stationary Bootstrap. *Statistica Sinica* **4** 461–476.
- [33] POLITIS, D.N. and ROMANO, J.P. (1995). Bias-Corrected Nonparametric Spectral Estimation. *J. Time Ser. Anal.* **16** 67–104.
- [34] RISSANEN, J. and BARBOSA, L. (1969). Properties of infinite covariance matrices and stability of optimum predictors. *Information Sci.* **1** 221–236.
- [35] SHAO, X. (2010). The dependent wild bootstrap. *J. Am. Stat. Assoc.* **105** 218–235.
- [36] WU, W.B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *STATISTICA SINICA* **19** 1755–1768.

C. JENTSCH
 DEPARTMENT OF ECONOMICS
 UNIVERSITY OF MANNHEIM
 L7, 3-5
 68131 MANNHEIM
 GERMANY
 E-MAIL: cjentsch@mail.uni-mannheim.de

D. N. POLITIS
 DEPARTMENT OF MATHEMATICS
 UNIVERSITY OF CALIFORNIA, SAN DIEGO
 LA JOLLA, CA 92093-0112
 USA
 E-MAIL: politis@math.ucsd.edu

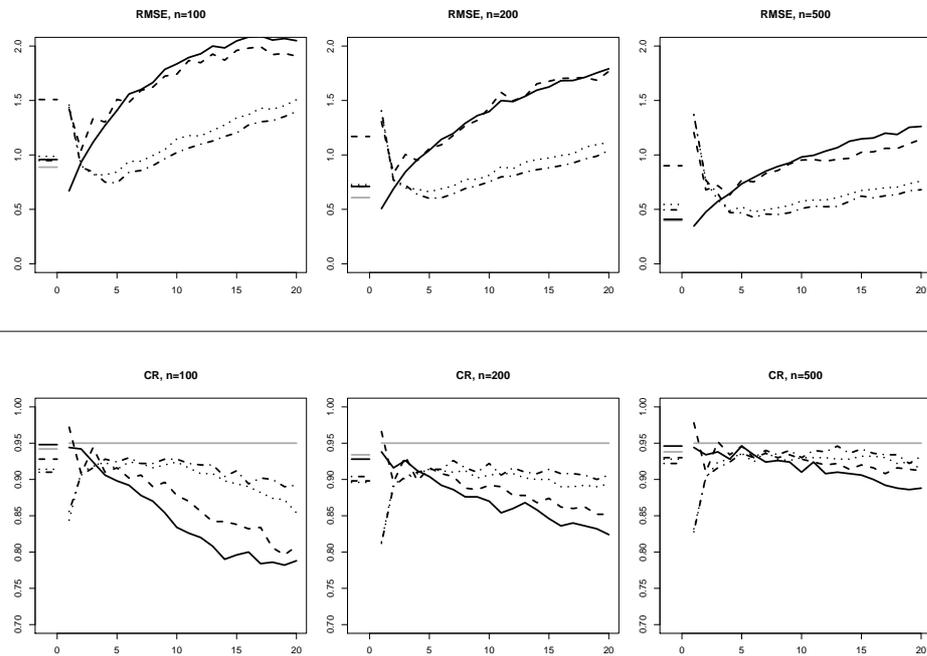


FIG 1. $RMSE$ for estimating $Var(\sqrt{n}(\bar{X}_1 - \mu_1))$ and CR of bootstrap confidence intervals for μ_1 by MLPB (solid), ARsieve (dashed), MBB (dotted) and TBB (dash-dotted) are reported vs. the respective tuning parameters $l, p, s \in \{1, \dots, 20\}$ for the VMA(1) model with sample size $n \in \{100, 200, 500\}$. Line segments indicate results for data-adaptively chosen tuning parameters. MLPB with individual (grey) and global (black) banding parameter choice are reported.

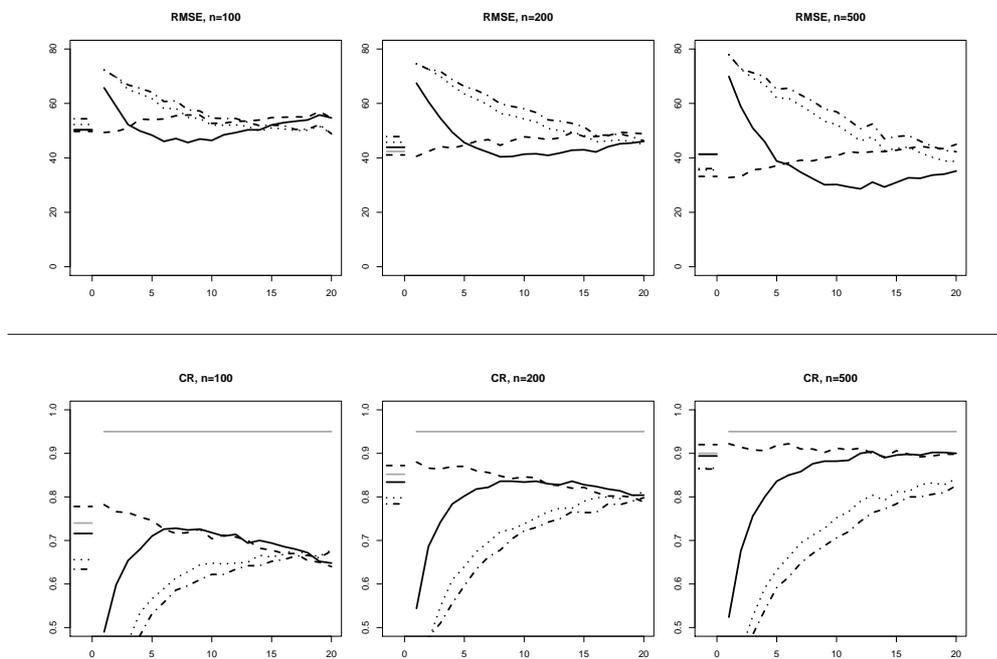


FIG 2. As in Figure 1, but with VAR(1) model.

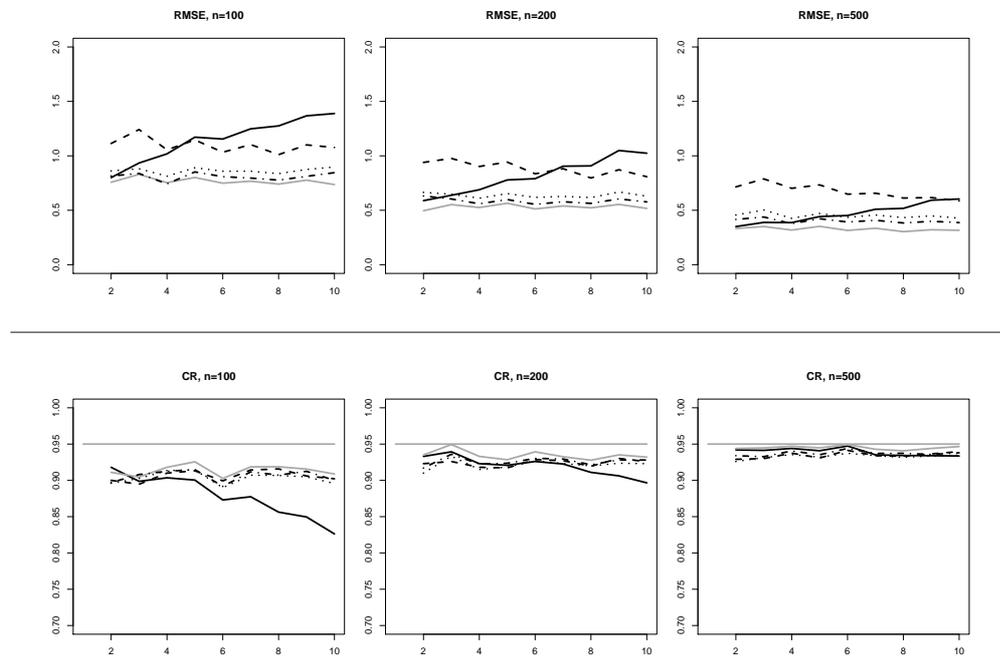


FIG 3. Average RMSE for estimating $\text{Var}(\sqrt{n}(\bar{X}_i - \mu_i))$, $i = 1, \dots, d$ and average CR of bootstrap confidence intervals for μ_i , $i = 1, \dots, d$, by MLPB (solid), ARsieve (dashed), MBB (dotted) and TBB (dash-dotted) with data-based optimal tuning parameter choices are reported vs. the dimension $d \in \{2, \dots, 10\}$ for the $\text{VMA}_d(1)$ model with sample size $n \in \{100, 200, 500\}$.

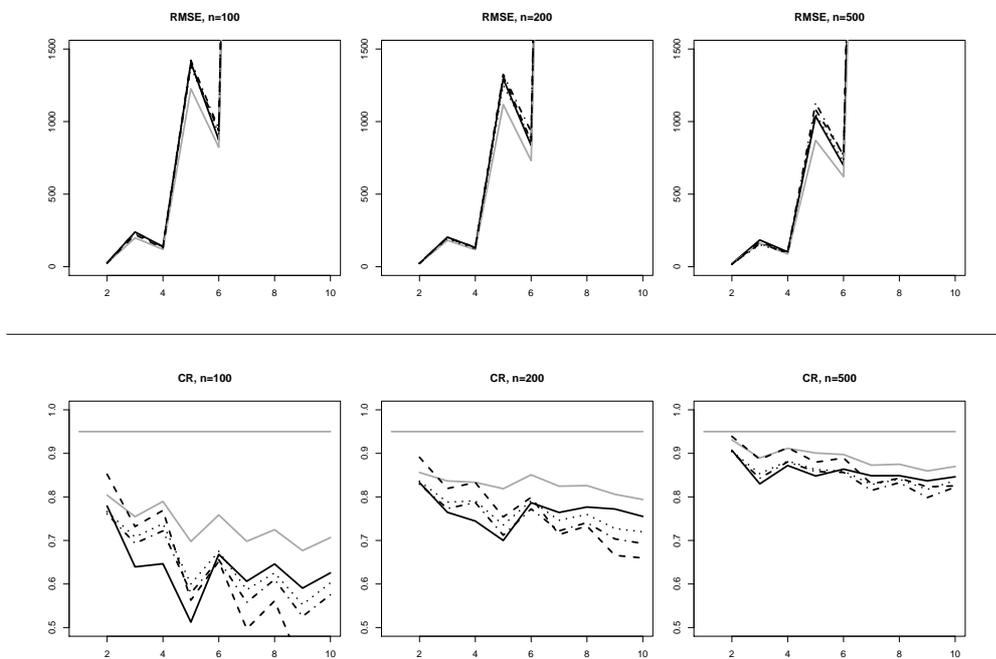


FIG 4. As in Figure 3, but with $VAR_d(1)$ model.