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

CARSTEN JENTSCH AND DIMITRIS N. POLITIS

ABSTRACT. 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 esti-
mating 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 autore-
(1992), Politis and Romano (1992,1994), etc.]. A recent addition to the available time series
bootstrap methods was the linear process bootstrap (LPB) introduced 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

\textit{Date:} February 23, 2014.

\textit{2000 Mathematics Subject Classification.} 62G09; 62M10.

\textit{Key words and phrases.} Asymptotics, bootstrap, covariance matrix, high dimensional data, multivariate time
series, sample mean, spectral density.
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 the LPB could be applied in the context of multivariate time series data. 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. 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 are deferred to Section 7.

2. Preliminaries

Suppose we consider an \( \mathbb{R}^d \)-valued time series process \( \{ X_t, t \in \mathbb{Z} \} \) with \( X_t = (X_{1,t}, \ldots, X_{d,t})^T \) and we have data \( X_1, \ldots, X_n \) at hand. The process \( \{ X_t, t \in \mathbb{Z} \} \) is assumed to be strictly stationary and its \((d \times d)\) autocovariance matrix \( C(h) = (C_{ij}(h))_{i,j=1,\ldots,d} \) at lag \( h \in \mathbb{Z} \) is

\[
C(h) = E \left( (X_{t+h} - \mu)(X_t - \mu)^T \right),
\]

where \( \mu = E(X_t) \) and the sample autocovariance \( \hat{C}(h) = (\hat{C}_{ij}(h))_{i,j=1,\ldots,d} \) at lag \( |h| < n \) is defined by

\[
\hat{C}(h) = \frac{1}{n} \sum_{t=\max(1,1-h)}^{\min(n,n-h)} (X_{t+h} - \bar{X})(X_t - \bar{X})^T,
\]
where $\overline{X} = \frac{1}{n} \sum_{t=1}^{n} X_t$ 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, $A^T$ indicates the transpose of a matrix $A$, $\overline{A}$ the complex conjugate of $A$ and $A^H = \overline{A}^T$ denotes the transposed conjugate of $A$. 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 $\underline{X} = vec(X) = (X_1, \ldots, X_{dn})^T$ be the $dn$-dimensional vectorized version of the $(d \times n)$ data matrix $X = [X_1 : X_2 : \cdots : X_n]$ and denote the covariance matrix of $\underline{X}$, which is symmetric block Toeplitz, by $\Gamma_{dn}$, that is,

$$\Gamma_{dn} = \begin{pmatrix} C(0) & C(-1) & \cdots & C(-(n-1)) \\ C(1) & C(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & C(-1) \\ C(n-1) & C(1) & \cdots & C(0) \end{pmatrix} = \begin{pmatrix} \Gamma_{dn}(i,j) \\ i,j = 1, \ldots, dn \end{pmatrix},$$

(2.3)

where $\Gamma_{dn}(i,j) = Cov(X_i, X_j)$ is the covariance between the $i$th and $j$th entry of $\underline{X}$. Note that the second order stationarity of $\{X_t, t \in \mathbb{Z}\}$ 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 $\{C_{ij}(h), h \in \mathbb{Z}, i,j = 1, \ldots, d\}$ and $\{\Gamma_{dn}(i,j), i,j \in \mathbb{Z}\}$ holds true. Precisely, we have

$$\Gamma_{dn}(i,j) = Cov(X_i, X_j) = Cov(X_{m_1(i),m_2(i)}, X_{m_1(j),m_2(j)}) = C_{m_1(i,j)}(m_2(i,j)),$$

(2.4)

where $m_1(i,j) = (m_1(i), m_1(j))$ and $m_2(i,j) = m_2(i) - m_2(j)$ with $m_1(k) = (k-1)\text{mod } d + 1$ and $m_2(k) = [k/d]$ and $[x]$ denotes the smallest integer greater or equal to $x \in \mathbb{R}$.

If one is interested in estimating the quantity $\Gamma_{dn}$, it seems natural to plug in the sample covariances $\hat{C}(i-j)$ and $\hat{\Gamma}_{dn}(i,j) = \hat{C}_{m_1(i,j)}(m_2(i,j))$ in $\Gamma_{dn}$ and to use

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

But unfortunately this estimator is not a consistent estimator for $\Gamma_{dn}$ in the sense that the operator norm of $\hat{\Gamma}_{dn} - \Gamma_{dn}$ 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, which 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 $\Gamma_{dn}$ 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) $\{X_t, t \in \mathbb{Z}\}$ is an $\mathbb{R}^d$-valued strictly stationary time series process with mean $E(X_t) = \mu$ and autocovariances $C(h)$ defined in (2.1) such that $\sum_{h=-\infty}^{\infty} |h|^g |C(h)| < \infty$ for some $g \geq 0$ to be further specified, where $||A||_p = (\sum_{i,j} |a_{ij}|^p)^{1/p}$ for some matrix $A = (a_{ij})$. 

(A2) There exists a constant $M < \infty$ such that for all $n \in \mathbb{N}$, all $h$ with $|h| < n$ and all $i, j = 1, \ldots, d$, 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 $\|A\|_p = (E(|A|^p))^1/p$.

(A3) There exists an $n_0 \in \mathbb{N}$ large enough such that for all $n \geq n_0$ the eigenvalues $\lambda_1, \ldots, \lambda_{dn}$ of the $(dn \times dn)$ covariance matrix $\Gamma_{dn}$ are bounded uniformly away from zero.

(A4) Define the projection operator $P_k(\bar{X}) = E(\bar{X}|F_k) = E(\bar{X}|F_{k-1})$ for $F_k = \sigma(X_{i,t}, t \leq k)$ and suppose that for all $i = 1, \ldots, d$, we have $\sum_{m=0}^{\infty} \|P_0X_{i,m}\|_q < \infty$ 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
$$
\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} \mathcal{N}(0, \mathbf{V}),
$$
where $\xrightarrow{D}$ denotes convergence in distribution, $\mathcal{N}(0, \mathbf{V})$ 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 $\hat{f}_{jk}(\omega)$ as defined in (4.2) in Section 4, a CLT holds true. That is, for arbitrary frequencies $0 \leq \omega_1, \ldots, \omega_s \leq \pi$, we have that
$$
\sqrt{nb} \left( \hat{f}_{pq}(\omega_1) - f_{pq}(\omega_1) : p, q = 1, \ldots, d; l = 1, \ldots, s \right)
$$
converges to an $sd^2$-dimensional normal distribution for $b \to 0$ and $nb \to \infty$ such that $nb^5 = O(1)$ as $n \to \infty$, where the limiting covariance matrix is obtained from
$$
nbCov \left( \hat{f}_{pq}(\omega), \hat{f}_{rs}(\lambda) \right) = \left( f_{pr}(\omega)f_{qs}(\omega)\delta_{\omega,\lambda} + f_{ps}(\omega)f_{qr}(\omega)\tau_{0,\pi} \right) \frac{1}{2\pi} \int K^2(u)du + o(1)
$$
and the limiting bias from
$$
E \left( \hat{f}_{pq}(\omega) \right) - f_{pq}(\omega) = b^2 f''_{pq}(\omega) \frac{1}{4\pi} \int K(u)u^2du + o(b^2)
$$
for all $p, q, r, s = 1, \ldots, d$, where $\delta_{\omega,\lambda} = 1$ if $\omega = \lambda$ and $\tau_{0,\pi} = 1$ if $\omega = \lambda \in \{0, \pi\}$ and zero otherwise, respectively. Therefore, $f(\omega)$ 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 $(X_{i,t}, t \in \mathbb{Z})$. This follows with (2.3) and the inversion formula from
$$
\mathbf{c}^T \Gamma_{dn} \mathbf{c} = \mathbf{c}^T \left( \int_{-\pi}^{\pi} \mathbf{J}_\omega^T \mathbf{f}(\omega) \mathbf{J}_\omega d\omega \right) \mathbf{c} \geq 2\pi |\mathbf{c}|^2 \inf_{\omega} \lambda_{\text{min}}(\mathbf{f}(\omega))
$$
for all $\mathbf{c} \in \mathbb{R}^{dn}$, where $\mathbf{J}_\omega = (e^{-i\omega}, \ldots, e^{-in\omega}) \otimes \mathbf{I}_d$ 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 $\alpha$-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...
Let the underlying process be linear, i.e. $X_t = \sum_{k=-\infty}^{\infty} B_k \xi_{t-k}$, $t \in \mathbb{Z}$, where $\{\xi_t, t \in \mathbb{Z}\}$ 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 $i, j, k, l = 1, \ldots, d$ and the sequence of $(d \times d)$ coefficient matrices $\{B_k, k \in \mathbb{Z}\}$ is component-wise absolutely summable.

### Linear-type condition

(i) **Suppose the underlying process is linear, i.e.** $X_t = \sum_{k=-\infty}^{\infty} B_k \xi_{t-k}$, $t \in \mathbb{Z}$, **where** $\{\xi_t, t \in \mathbb{Z}\}$ **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** $i, j, k, l = 1, \ldots, d$ **and the sequence of** $(d \times d)$ **coefficient matrices** $\{B_k, k \in \mathbb{Z}\}$ **is component-wise absolutely summable.**

(ii) **Mixing-type condition:** Let $\text{cum}_{a_1, \ldots, a_d}(u_1, \ldots, u_k-1) = \text{cum}(X_{a_1 u_1}, \ldots, X_{a_k u_k-1}, X_{a_0})$ **denote the** $k$th **order joint cumulant of** $X_{a_1 u_1}, \ldots, X_{a_k u_k-1}, X_{a_0}$ [cf. Brillinger (1981)] **and suppose** $\sum_{s,h=0}^{\infty} s, h \rightarrow (\text{cum}_{i,j,i,j}(s + h, s, h)) < \infty$ **for all** $i, j = 1, \ldots, d$. **Note that this is satisfied if** $\{X_i, t \in \mathbb{Z}\}$ **is an i.i.d.** **random sequence with finite fourth moments** and $\sum_{j=1}^{\infty} j^4 \sigma(j)^2 < \infty$ **for some** $\sigma > 0$ [cf. Shao (2010), p.221].

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

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

where $(\nu_{s,h})$ is an absolutely summable sequence in both arguments, that is, $\sum_{s,h=0}^{\infty} |\nu_{s,h}| < \infty$ [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

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

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 $\kappa_l(x) = \kappa(\frac{x}{l})$. 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 $\kappa$ with $c_\kappa > 1$ is the trapezoid

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

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 $\hat{\Gamma}_{n,l}$ of $\Gamma_{dn}$ is given by

$$
\hat{\Gamma}_{n,l} = \left( \kappa_l(i-j) \hat{C}(i-j) \right)_{i,j=1, \ldots, n} = \left( \Gamma_{n,l}(i,j) \right)_{i,j=1, \ldots, n},
$$

where $\hat{\Gamma}_{n,l}(i,j) = \hat{C}_{\kappa_l}^{\gamma_{n,l}}(m_2(i,j))$ and $\hat{C}_{\kappa_l}(h) = \kappa_l(h) \hat{C}_{\kappa_l}(h)$.

The following Theorem 2.1 deals with consistency of the tapered estimator $\hat{\Gamma}_{n,l}$ 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

[cf. Brockwell and Davis (1989), Hannan (1970)] or weak dependence [cf. Dedecker et al. (2007)]. Note also that the condition $nb^2 = O(1)$ includes the optimal bandwidth choice $nb^2 \rightarrow C^2$, $C > 0$ for second-order kernels, which leads to a non-vanishing bias in the limiting normal distribution.
norm of a complex-valued \((d \times d)\) matrix \(A\) is defined by

\[
\rho(A) = \max_{\xi \in \mathbb{C}^d, \xi_2 = 1} |A \xi|_2,
\]

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

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

\[
\|\rho(\hat{\Gamma} - \Gamma_{dn})\|_2 \leq \frac{4MD^2(|c_o|l + 1)}{\sqrt{n}} + 2 \sum_{h=0}^{[c_o]} \frac{|h|}{n} |C(h)|_1 + 2 \sum_{l=1}^{n-1} |C(h)|_1,
\]

\[(2.8)\]

The second term on the right-hand side of (2.8) can be written as

\[
2 \sum_{h=0}^{[c_o]} \frac{|h|}{n} |C(h)|_1 \leq 2 \sum_{h=0}^{[c_o]} \frac{|h|}{n} |C(h)|_1 + \frac{2}{l} |C(h)|_1 + \frac{2}{l+1} |C(h)|_1.
\]

Theorem 2.2 below which extends Theorem 2.1 just slightly.

As already pointed out by McMurry and Politis (2010), the tapered estimator \(\hat{\Gamma}_{n,l}\) is not guaranteed to be positive semi-definite or even to be positive definite for finite sample sizes. However, \(\hat{\Gamma}_{n,l}\) is at least “asymptotically positive definite” under assumption (A3) and due to (2.8) if \(1/l + 1/\sqrt{n} = o(1)\) holds. In the following, we require a consistent estimator for \(\Gamma_{dn}\) 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 \(\Gamma_{dn}\) related to \(\hat{\Gamma}_{n,l}\) that is assured to be positive definite for all sample sizes, we construct a modified estimator \(\hat{\Gamma}_{n,l}^{\epsilon}\) in the following. Let \(\hat{V} = \text{diag}(\hat{\Gamma}_{dn})\) be the diagonal matrix of sample variances and define \(\hat{\Gamma}_{n,l} = \hat{V}^{-1/2}\hat{\Gamma}_{n,l}\hat{V}^{-1/2}\). Now, we consider the spectral factorization

\[
\hat{\Gamma}_{n,l} = \hat{S}D\hat{S}^T,
\]

where \(\hat{S}\) is an \((dn \times dn)\) orthogonal matrix and \(D = \text{diag}(r_1, \ldots, r_{dn})\) is the diagonal matrix containing the eigenvalues of \(\hat{\Gamma}_{n,l}\) such that \(r_1 \geq r_2 \geq \cdots \geq r_{dn}\). It is worth noting that this factorization always exists due to symmetry of \(\hat{\Gamma}_{n,l}\), but that the eigenvalues can be positive, zero or even negative. Now, define

\[
\hat{\Gamma}_{n,l}^{\epsilon} = \hat{V}^{1/2}\hat{\Gamma}_{n,l}^{\epsilon}\hat{V}^{1/2} = \hat{V}^{1/2}SD^\epsilon S^T\hat{V}^{1/2},
\]

\[(2.9)\]

where \(D^\epsilon = \text{diag}(r_1^\epsilon, \ldots, r_{dn}^\epsilon)\) and \(r_i^\epsilon = \max(r_i, en^{-\beta})\). Here, \(\beta > 1/2\) and \(\epsilon > 0\) are user defined constants that ensure the positive definiteness of \(\hat{\Gamma}_{n,l}^{\epsilon}\). Contrary to the univariate case discussed in McMurry and Politis (2010), we propose to adjust the eigenvalues of the (equivariant) correlation matrix \(\hat{R}_{n,l}\) instead of \(\hat{\Gamma}_{n,l}\), which then comes along without a scaling factor in the definition of \(r_i^\epsilon\). Further, note that setting \(\epsilon = 0\) leads to a positive semi-definite estimate if \(\hat{\Gamma}_{n,l}\) is indefinite, which does not suffice for computing the Cholesky decomposition, and also that \(\hat{\Gamma}_{n,l}^{\epsilon}\) generally loses the banded shape of \(\hat{\Gamma}_{n,l}\). 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

\[
\|\rho(\hat{\Gamma}_{n,l}^{\epsilon} - \Gamma_{dn})\|_2 \leq \frac{8MD^2(|c_o|l + 1)}{\sqrt{n}} + 4 \sum_{h=0}^{[c_o]} \frac{|h|}{n} |C(h)|_1 + 4 \sum_{l=1}^{n-1} |C(h)|_1 + \epsilon \max_i C_i(0)n^{-\beta} + O\left(\frac{1}{n^{1/2+\beta}}\right).
\]

\[(2.10)\]
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 \( \hat{\Gamma}_{n,t} \). The required results are summarized in the following corollary.

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

(i) \( \rho(\hat{\Gamma}_{n,t} - \Gamma_{dn}) \) and \( \rho((\hat{\Gamma}_{n,t})^{-1} - \Gamma_{dn}^{-1}) \) are terms of order \( O_P(r_{t,n}), \) where

\[
r_{t,n} = \frac{l}{\sqrt{n}} + \sum_{h=1}^{\infty} |C(h)|_1,
\]

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

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

(iii) \( \rho(\Gamma_{dn}), \rho(\Gamma_{dn}^{-1}), \rho((\Gamma_{n,t})^{1/2}), \rho((\Gamma_{n,t}^{-1})^{-1/2}) \) and \( \rho((\Gamma_{n,t})^{-1/2}), \rho((\Gamma_{n,t}^{1/2})^{-1}) \) are bounded from above and below. \( \rho(\Gamma_{n,t}), \rho((\Gamma_{n,t})^{-1}) \) and \( \rho((\Gamma_{n,t}^{-1})^{-1/2}), \rho((\Gamma_{n,t}^{1/2})^{-1}) \) are bounded from above and below (in probability) if \( r_{t,n} = o(1) \) and \( \log^2(n)r_{t,n} = o(1) \), 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 \( C(h) \) is multiplied with the same \( r_{t}(h) \) in (2.7). However, it is possible to use individual banding parameters \( l_{pq} \) for each sequence of entries \( \{C_{pq}(h), h \in \mathbb{Z}\} \), \( p, q = 1, \ldots, d \) as proposed in Politis (2011).

### 2.3. Selection of tuning parameters.

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

To select a suitable taper \( \kappa \) 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 \( \kappa \) 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 \( \kappa \) 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 and 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

\[
\hat{R}_{jk}(h) = \frac{\hat{C}_{jk}(h)}{\sqrt{\hat{C}_{jj}(0)\hat{C}_{kk}(0)}}, \quad j, k = 1, \ldots, d
\]

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 \( q_{jk} \) as the smallest nonnegative integer such that

\[
|\hat{R}_{jk}(q_{jk} + h)| < M_0\sqrt{\log(n)/n}
\]

for \( h = 1, \ldots, K_n \), 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 $j, k = 1, \ldots, d$, we take

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

as a data-driven global choice of the banding parameter $l$. By setting $\hat{I}_{jk} = \hat{q}_{jk}$, 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 $\hat{\Gamma}_{\kappa,l}^c$, one has to select two more parameters $\epsilon$ and $\beta$, 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 discussing some modifications and comment on the special case where the tapered covariance estimator becomes diagonal.

Step 1. Let $X$ be the $(d \times n)$ data matrix consisting of $\mathbb{R}^d$-valued time series data $X_1, \ldots, X_n$ of sample size $n$. Compute the centered observations $\bar{X}_t = X_t - \bar{X}$, where $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$, let $\bar{Y}$ be the corresponding $(d \times n)$ matrix of centered observations and define $\bar{Y} = \text{vec}(\bar{Y})$ to be the $dn$-dimensional vectorized version of $\bar{Y}$.

Step 2. Compute $W = (\hat{\Gamma}_{\kappa,l}^c)^{-1/2} \bar{Y}$, where $(\hat{\Gamma}_{\kappa,l}^c)^{-1/2}$ denotes the lower left triangular matrix $L$ of the Cholesky decomposition $\hat{\Gamma}_{\kappa,l}^c = LL^T$.

Step 3. Let $\hat{Z}$ be the standardized version of $W$, that is, $Z_i = \frac{W_i - \bar{W}}{\sigma_W}$, $i = 1, \ldots, dn$, where $\bar{W} = \frac{1}{dn} \sum_{t=1}^{dn} W_t$ and $\sigma_W^2 = \frac{1}{dn} \sum_{t=1}^{dn} (W_t - \bar{W})^2$.

Step 4. Generate $\hat{Z}_i^* = (Z_1^*, \ldots, Z_{dn}^*)^T$ by i.i.d. resampling from $\{Z_1, \ldots, Z_{dn}\}$.

Step 5. Compute $Y_i^* = (\hat{\Gamma}_{\kappa,l}^c)^{1/2} Z_i^*$ and let $Y_i^*$ be the matrix that is obtained from $Y_i^*$ by putting this vector column-wise into an $(d \times n)$ matrix and denote its columns by $Y_1^*, \ldots, Y_n^*$.

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 $Z$ 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 $Z_i^*$. More precisely, Steps 3 and 4 can be replaced by

Step 3′. Let $\hat{Z} = (\hat{Z}_1^T, \ldots, \hat{Z}_n^T)^T$ be the standardized version of $\hat{W} = (\hat{W}_1^T, \ldots, \hat{W}_n^T)^T$, that is, $\hat{Z} = \hat{Z}_i^* = \frac{\hat{Z}_i^* - \bar{W}}{\hat{\sigma}_W}$, where $\bar{W} = \frac{1}{n} \sum_{t=1}^n \hat{W}_t$, and $\hat{\sigma}_W = \frac{1}{n} \sum_{t=1}^n (\hat{W}_t - \bar{W})^2$.

Step 4′. Generate $\hat{Z}_i^* = (\hat{Z}_1^*, \ldots, \hat{Z}_n^*)^T$ by i.i.d. resampling from $\{\hat{Z}_1, \ldots, \hat{Z}_n\}$.

This might preserve more higher order features of the data that are not captured by $\hat{\Gamma}_{\kappa,l}^c$. 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 $0 < l < \frac{1}{t_0}$, the banded covariance matrix estimator $\hat{\Gamma}_{\kappa,l}$ (and $\hat{\Gamma}_{\kappa,l}^c$ 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 $\bar{X}$, that is,

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

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

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 $I_n(\omega) = \mathcal{J}_n(\omega)I^H_n(\omega)$ the periodogram matrix, where

$$\mathcal{J}_n(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Y_t e^{-i\omega t}$$

is the discrete Fourier transform (DFT) of $Y_1, \ldots, Y_n, Y_t = \bar{X}_t - \bar{X}$. We define the estimator

$$\hat{f}(\omega) = \frac{1}{n} \sum_{k=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n-1}{2} \rfloor} K_b(\omega - \omega_k)I_n(\omega_k)$$

for the spectral density matrix $f(\omega)$, where $|x|$ is the integer part of $x \in \mathbb{R}$, $\omega_k = 2\pi \frac{k}{n}, k = -\lfloor \frac{n-1}{2} \rfloor, \ldots, \lfloor \frac{n-1}{2} \rfloor$ 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 $\int K(u)u^2du < \infty$ and we set $K_b(\cdot) = \frac{1}{b}K(b\cdot)$. Let $\mathcal{I}_n(\omega)$ be the bootstrap analogue of $I_n(\omega)$ based on $Y_n^*, \ldots, Y_n$ generated from the MLPB scheme and let $\hat{f}^*(\omega)$ be the bootstrap analogue of $\hat{f}(\omega)$.

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

$$\sup_{x \in \mathbb{R}^d} \left| P \left\{ \left( \sqrt{\text{nb}}(\hat{f}_{pq}(\omega) - \hat{f}_{pq}(\omega_1)) : p, q = 1, \ldots, d; l = 1, \ldots, s \right) \leq x \right\} \right| = o_P(1),$$

where $\hat{f}_{pq}(\omega) = \frac{1}{2\pi} \sum_{h = -\lfloor n/2 \rfloor}^{\lfloor n/2 \rfloor} K_d(h)\hat{C}_{pq}(h)e^{-ih\omega}$ and, in particular,

$$\text{nbCov}^* \left( \hat{f}_{pq}^*(\omega), \hat{f}_{pq}^*(\lambda) \right) = \left( \int f_{pq}(\omega)f_{pq}(\omega)\text{d}\omega, \lambda + f_{pq}(\omega)f_{qr}(\omega)\tau_{0,\omega,\lambda} \right) = \frac{1}{2\pi} \int K^2(u)du + o_P(1),$$

and $E^* \left( \hat{f}_{pq}(\omega) \right) - \hat{f}_{pq}(\omega) = b^2 f_{pq}(\omega)\frac{1}{2\pi} \int K(u)u^2du + p_P(b^2)$, for all $p, q, r, s = 1, \ldots, d$ 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 $\vartheta = E(g(X_1, \ldots, X_{m+1}))$ such that

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

for some sufficiently smooth functions $g : \mathbb{R}^{d \times m} \to \mathbb{R}^k$, $w : \mathbb{R}^k \to \mathbb{R}$ 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 $\tilde{X}_t := g(X_t, \ldots, X_{t+m-1})$ and let $\tilde{X}_1, \ldots, \tilde{X}_{n-m+1}$ be the set of blocked data.
Step 2. Apply the MLPB scheme of Section 3 to the $k$-dimensional blocked data $\tilde{X}_1, \ldots, \tilde{X}_{n-m+1}$ to get bootstrap observations $\tilde{X}_{1}^*, \ldots, \tilde{X}_{n-m+1}^*$.
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 $\sqrt{n}(T_n^* - T_n), \ldots, \sqrt{n}(T_n^{*K} - T_n)$.

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. $d = d(n) \to \infty$ as $n \to \infty$. 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 McMurtry 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 the assumption (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 $\{\{X_t^{(n)}, t \in \mathbb{Z}\}\}_{n \in \mathbb{N}}$ of now increasing dimension.

5.1. Assumptions.

(A1') $\{\{X_t = (X_{1,t}, \ldots, X_{d(n),t})^T, t \in \mathbb{Z}\}\}_{n \in \mathbb{N}}$ is a sequence of $\mathbb{R}^{d(n)}$-valued strictly stationary time series processes with mean vectors $E(X_t) = \mu = (\mu_1, \ldots, \mu_{d(n)})$ and autocovariances $C(h) = (C_{ij}(h))_{i,j=1,\ldots,d(n)}$ defined as in (2.1). Here, $(d(n))_{n \in \mathbb{N}}$ is a non-decreasing
Therefore, to be able to establish a meaningful theory, we have to 
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

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

(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 $\lambda_1, \ldots, \lambda_{dn}$ of the $(dn \times dn)$ covariance matrix $\Gamma_{dn}$ are bounded uniformly away from zero and from above.

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

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

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

\[
\sqrt{n} (\bar{X}^T (\bar{X} - \mu)) / \nu \xrightarrow{D} \mathcal{N}(0, 1).
\]

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

\[
\sum_{h=-\infty}^{\infty} |C(h)|_1 = O(d^2).
\]

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.

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
To tackle the increasing time series dimension and to prove such a CLT result, we have to make use of appropriate operator norm consistency. In particular, we are interested in the unknown distribution of the block bootstrap (MBB) and the tapered block bootstrap (TBB) by means of simulation. In order to 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. In particular, we are interested in the unknown distribution of the sample mean of bivariate time series data. We compare the aforementioned bootstrap methods by plotting

a) root mean squared errors (RMSE) for estimating the covariance matrix of \( \sqrt{n}(\bar{X} - \mu) \)

b) coverage rates (CR) of bootstrap confidence intervals for both components of \( \mu \)

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 \( \hat{b} = \hat{b}(d(n)) \) as described in (A5′) above.

\[ \rho(\hat{\Gamma}_{n,l}^{(c)} - \Gamma_{dn}) \text{ and } \rho((\hat{\Gamma}_{n,l}^{(c)})^{-1} - \Gamma_{dn}^{-1}) \text{ are terms of order } O_P(d^2\tilde{r}_{l,n}), \]

where

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

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

(ii) \( \rho((\hat{\Gamma}_{n,l}^{(c)})^{1/2} - \Gamma_{dn}^{1/2}) \text{ and } \rho((\hat{\Gamma}_{n,l}^{(c)})^{-1/2} - \Gamma_{dn}^{-1/2}) \) are both 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 \( 1/l + \log^2(dn)d^2 l/\sqrt{n} + \log^2(dn)d^2/l^2 = o(1) \).

(iii) \( \rho(\Gamma_{dn}), \rho(\Gamma_{dn}^{1/2}) \text{ and } \rho(\Gamma_{dn}^{-1/2}) \) are bounded from above and below. \( \rho(\hat{\Gamma}_{n,l}^{(c)}) \) and \( \rho((\hat{\Gamma}_{n,l}^{(c)})^{-1}) \) as well as \( \rho((\hat{\Gamma}_{n,l}^{(c)})^{1/2}) \) and \( \rho((\hat{\Gamma}_{n,l}^{(c)})^{-1/2}) \) 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 \( \rho(\hat{\Gamma}_{n,l}^{(c)} - \Gamma_{dn}) = o_P(1) \) can be interpreted nicely. If \( g \) is chosen to be large enough, \( d^2 l/\sqrt{n} \) 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 \( d \)).


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. In particular, we are interested in the unknown distribution of the sample mean of bivariate time series data. We compare the aforementioned bootstrap methods by plotting

a) root mean squared errors (RMSE) for estimating the covariance matrix of \( \sqrt{n}(\bar{X} - \mu) \)

b) coverage rates (CR) of bootstrap confidence intervals for both components of \( \mu \)
for three models and three sample sizes for different choices of corresponding tuning parameters. These are the banding parameter $l$ (MLPB), the autoregressive order $p$ (ARsieve) and the block length $s$ (MBB, TBB). For the MLPB, we report also RMSE and CR for data-adaptively chosen global and individual banding parameters as discussed in Section 2.3. The R code is available at http://www.math.ucsd.edu/~politis/SDFT/function_MLPB.R.

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 $\sqrt{n}(X - \mu)$ 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’.

6.1. The data generating processes (DGP).
We consider realizations $X_1, \ldots, X_n$ of length $n = 100, 200, 500$ from three bivariate models. Precisely, we study an i.i.d. white noise process

$$WN \quad X_t = \varepsilon_t,$$

a first order vector moving average process

$$VMA(1) \quad X_t = A X_{t-1} + \varepsilon_t,$$

and a first order vector autoregressive process

$$VAR(1) \quad X_t = A X_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, I_2)$ is a normally distributed i.i.d. white noise process and

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

have been used in all cases. It is worth noting, that (asymptotically) all bootstrap procedures under consideration yield valid approximations for all three models above. In contrast to the i.i.d. case, where all proposed techniques remain valid for all choices of sufficiently small tuning parameters, this is not true for the other DGPs. 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 the VMA(1) and VAR(1) models.

6.2. Simulation results.
In Figures 1–3, by plotting RMSE and CR for different choices of tuning parameters, the MLPB is compared to ARsieve, MBB and TBB for the three models under investigation, respectively. In each figure, the upper half of panels corresponds to the estimation of $Var(\sqrt{n}(X_1 - \mu_1))$ (first row, RMSE) and $\mu_1$ (second row, CR) and the second half of panels to $Var(\sqrt{n}(X_2 - \mu_2))$ (third row, RMSE) and $\mu_2$ (fourth row, CR).

6.2.1. WN model.
In Figure 1, the data is i.i.d. and in fact a bootstrap for dependent data is redundant to capture any dependence structure. However, we compare MLPB, ARsieve, MBB and TBB for $l, p, s = 1, 2, \ldots, 20$, i.e. the data generating process is over-fitted and hence slightly misspecified by MLPB and ARsieve. Observe that $\hat{\Gamma}_{\kappa,l}^{\epsilon}$ computed for the MLPB becomes block-diagonal for $l < 0.5$ and the scheme degenerates to become an i.i.d. bootstrap as is true for the ARsieve if $p = 0$ [compare Remark 3.1], which would be of course most appropriate in this case, but are not
reported here. This is in contrast to MBB and TBB, which simplify to become Efron’s bootstrap for \( s = 1 \). However, the MLPB with data-adaptively selected banding parameters performs very well. In Figure 1, it can be seen that wrt to RMSE and CR, all bootstrap procedures lose in terms of efficiency with redundantly increasing tuning parameters.

6.2.2. VMA(1) model.

For data generated by the VMA(1) model, Figure 2 shows that the MLPB outperforms ARsieve, 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. However, the performance wrt to CR is comparable to MLPB already for small AR orders \( p \). The MLPB with data-adaptively chosen \( l \) performs quite well, where the individual choice tends to performs superior to the global choice in most cases. In comparison, MBB and TBB perform quite well for adequate block length and, actually, appear to be more robust to this choice. In particular, MBB and TBB are superior to MLPB and ARsieve for a large range of tuning parameters wrt RMSE and CR.

6.2.3. VAR(1) model.

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 3 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 seems to 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. However, this is also illustrated by the poor performance of MLPB wrt to CR for small choices of \( l \). Nevertheless, the MLPB performs comparable to the ARsieve wrt to CR for a whole range of banding parameters. Furthermore, the plots indicate that the MLPB with data-adaptively chosen \( l \) can achieve good CR, where the individual choice generally outperforms the global choice particularly wrt to RMSE. Similar to the considerations of MLPB for the VMA(1) model in Figure 2, it can be seen that the performance of AR sieve 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, but for the upper half of Figure 3, this might be explained by the too small block lengths reported here. However, for the lower half of plots, MBB and TBB are not capable to achieve CR as good as MLPB and AR sieve, particularly, for small sample size.

Our simulation experience may be summarized as follows:

- In case of an i.i.d. process, e.g. the WN model, MLPB and SIEVE perform comparably well [cf. Figure 1], but are both outperformed by the two block bootstrap methods. Wrt to RMSE the tapered version of the MBB performs tends to be slightly better to the MBB, but the picture is not so clear wrt CR, where both are superior in some cases.
- If the underlying process is a moving average process, e.g. VMA(1) model, MLPB is comparable to the ARsieve as shown in Figure 2, but tends to be slightly superior.
- In case of an autoregressive process, e.g. VAR(1) model, ARsieve performs better than MLPB with respect CR, but not with respect to RMSE; see Figure 3. This is due to the higher variability of ARsieve estimates in comparison to MLPB estimates.
• For MLPB and ARsieve, all simulations show that larger tuning parameters $l$ and $p$ introduce more and more variability into the bootstrap estimation. This is due to involving sample autocovariances of lags up to $p$ and $l$, respectively; each of those involved sample autocovariances carries with it its own standard error that is of order $1/\sqrt{n}$. Consequently, both bootstrap procedures perform best in situations where $l$ and $p$ can be chosen small, respectively.

• MLPB with data-adaptively chosen banding parameters performs well, where the individual choice tends to be superior to the global choice in many cases.

7. Proofs

Proof of Theorem 2.1.
Symmetry of $\hat{\Gamma}_{\kappa,l} - \Gamma_{dn}$ together with problem 21, p. 313 in Horn and Johnson (1990) and plugging-in for $\hat{\Gamma}_{\kappa,l}(i,j)$ and $\Gamma_{dn}(i,j)$ yields

$$
\rho(\hat{\Gamma}_{\kappa,l} - \Gamma_{dn}) \leq \max_{1 \leq j \leq dn} \sum_{i=1}^{dn} |\kappa_l(m_2(i,j))\hat{C}_{ij}(i,j)(m_2(i,j)) - C_{ij}(m_2(i,j))|
$$

$$
\leq 2 \sum_{i,j=1}^{d} \sum_{h=0}^{n-1} |\kappa_l(h)\hat{C}_{ij}(h) - C_{ij}(h)|,
$$

where the second inequality is implied by the definitions of $m_4(\cdot, \cdot)$ and $m_2(\cdot, \cdot)$ in (2.4). By splitting-up the expression on the last right-hand side above and plugging-in for $\kappa_l(h)$, we get

$$
\rho(\hat{\Gamma}_{\kappa,l} - \Gamma_{dn}) \leq 2 \sum_{i,j=1}^{d} \left( \sum_{h=0}^{l} |\hat{C}_{ij}(h) - C_{ij}(h)| + \sum_{h=l+1}^{[c_s l]} |\kappa_l(h)\hat{C}_{ij}(h) - C_{ij}(h)| + \sum_{h=[c_s l]+1}^{n-1} |C_{ij}(h)| \right)
$$

$$
=: A_1 + A_2 + A_3.
$$

Now, before considering $A_1$ and $A_2$ separately, note that

$$
\|\hat{C}_{ij}(h) - C_{ij}(h)\|_2 \leq \frac{1}{n} \sum_{t=\max(1,1-h)}^{\min(n,n-h)} (X_{i,t+h} - \bar{X}_i)(X_{j,t} - \bar{X}_j) - (n - |h|)C_{ij}(h)\|_2 + \frac{|h|}{n}||C(h)||_2
$$

$$
\leq \frac{M}{\sqrt{n}} + \frac{|h|}{n}||C_{ij}(h)||,
$$

(7.1)

where the second inequality follows from assumption (A2). Using this result, we obtain

$$
\|A_1\|_2 \leq 2 \sum_{i,j=1}^{d} \sum_{h=0}^{l} \left( \frac{M}{\sqrt{n}} + \frac{|h|}{n}||C_{ij}(h)|| \right) \leq 2Md^2(l+1)\frac{1}{\sqrt{n}} + 2 \sum_{h=0}^{l} \frac{|h|}{n}||C(h)||_1.
$$

Now, consider $A_2$. First, it holds

$$
A_2 \leq 2 \sum_{i,j=1}^{d} \sum_{h=0}^{l+1} |\kappa_l(h)||\hat{C}_{ij}(h) - C_{ij}(h)| + 2 \sum_{h=l+1}^{[c_s l]} |\kappa_l(h) - 1||C(h)||_1
$$

$$
\leq 2 \sum_{i,j=1}^{d} \sum_{h=0}^{l+1} |\hat{C}_{ij}(h) - C_{ij}(h)| + 2 \sum_{h=l+1}^{[c_s l]} ||C(h)||_1
$$

\null
and straightforward application of (7.1) results in
\[
\|A_2\|_2 \leq 2 \sum_{i,j=1}^{d} \sum_{h=l+1}^{[c_r, l]} \left( \frac{M}{\sqrt{n}} + \frac{|h|}{n} |C_{ij}(h)| \right) + 2 \sum_{h=l+1}^{[c_r, l]} |C(h)|_1
\]
\[
\leq \frac{2Md^2([c_r, l] - l)}{\sqrt{N}} + 2 \sum_{h=l+1}^{[c_r, l]} \frac{|h|}{n} |C(h)|_1 + 2 \sum_{h=l+1}^{[c_r, l]} |C(h)|_1.
\]

Proof of Theorem 2.2.
Without loss of generality, the eigenvalues of \( \tilde{R}_{n,l} \) can be ordered such that \( r_1 \geq r_2 \geq \cdots \geq r_{dn} \)
and let \( \lambda_{\text{max}}(A) \) and \( \lambda_{\text{min}}(A) \) denote the largest and the smallest eigenvalue of a matrix \( A \), respectively. Then, by Corollary 4.3.3 in Horn and Johnson (1990), it holds
\[
r_{dn} = -\lambda_{\text{min}}(\tilde{R}_{n,l}) = \lambda_{\text{max}}(-\tilde{R}_{n,l}) \leq \lambda_{\text{max}}(R_{dn} - \tilde{R}_{n,l}) \leq \rho(R_{dn} - \tilde{R}_{n,l}),
\]
where \( R_{dn} = \tilde{V}^{-1/2} \Gamma_{dn} \tilde{V}^{-1/2} \) and the last inequality follows from the definition of the operator norm and the spectral factorization of a symmetric matrix. Further, it holds
\[
\hat{\Gamma}_{n,l} - \hat{\Gamma}_{n,l} = \tilde{V}^{1/2} S(D' - D) S^T \tilde{V}^{1/2},
\]
where
\[
D' - D = \text{diag} \left( \max(r_i, \epsilon n^{-\beta}) - r_i, i = 1, \ldots, dn \right) = \text{diag} \left( \max(0, \epsilon n^{-\beta} - r_i), i = 1, \ldots, dn \right).
\]
Together with (7.2) and
\[
\rho(\hat{\Gamma}_{dn} - \hat{\Gamma}_{n,l}) \leq \rho^2(\tilde{V}^{1/2} \rho(\text{diag}(D' - D) S^T) = \max_i \hat{C}_{ii}(0) \cdot \max(0, \epsilon n^{-\beta} - r_{dn})
\]
this leads to
\[
\rho(\tilde{\Gamma}_{n,l} - \Gamma_{dn}) \leq \max(0, \epsilon \max_i \hat{C}_{ii}(0)n^{-\beta} - r_{dn}) + \rho(\hat{\Gamma}_{n,l} - \Gamma_{dn})
\]
\[
\leq \max_i \hat{C}_{ii}(0) \left( \epsilon n^{-\beta} + \frac{1}{\max_i \hat{C}_{ii}(0)} \rho(\Gamma_{dn} - \hat{\Gamma}_{n,l}) + \rho(\Gamma_{dn} - \hat{\Gamma}_{n,l}) \right)
\]
\[
= \epsilon \max_i \hat{C}_{ii}(0)n^{-\beta} + 2\rho(\hat{\Gamma}_{n,l} - \Gamma_{dn}).
\]
From \( \| \hat{C}_{ii}(0) \|_2 = C_{ii}(0) + O(1), i = 1, \ldots, d \) and Theorem 2.1, we get the desired result. □

Proof of Corollary 2.1.
By (A1), Gerschgorins Theorem and by (A3), we have that \( \rho(\Gamma_{dn}), \rho(\Gamma_{dn}^{-1}), \rho(\Gamma_{dn}^{1/2}) \) and \( \rho(\Gamma_{dn}^{-1/2}) \)
are bounded from above and from below. The claimed result for \( \rho(\hat{\Gamma}_{n,l} - \Gamma_{dn}) \) follows directly from Theorem 2.2 and the corresponding result for the inverses of \( \hat{\Gamma}_{n,l} \) and \( \Gamma_{dn} \) follows from the proof of Theorem 2 in McMurry and Politis (2010). The claimed convergence of the Cholesky matrices is established through Theorem 2.1 of Drmac, Omladic and Veselic (1994), which provides the bound
\[
\rho \left( (\hat{\Gamma}_{n,l}^{1/2} - 1) / \Gamma_{dn}^{1/2} \right) \leq \frac{2c_n \rho(\Gamma_{dn}^{1/2}) \rho(\Gamma_{dn}^{-1/2}) \rho(\hat{\Gamma}_{n,l} - \Gamma_{dn}) (\Gamma_{dn}^{-1/2} \hat{\Gamma}_{n,l} - \Gamma_{dn} / \Gamma_{dn}^{-1/2})^T}{\sqrt{1 + 4c_n^2 \rho(\Gamma_{dn}^{-1/2}) \rho(\hat{\Gamma}_{n,l} - \hat{\Gamma}_{n,l}) (\Gamma_{dn}^{-1/2} \hat{\Gamma}_{n,l} - \Gamma_{dn} / \Gamma_{dn}^{-1/2})^T}}
\]
where \( c_n = \frac{1}{2} + \lfloor \log_2(dn) \rfloor \) if the radicand in the denominator above is strictly positive. Since \( \rho(\Gamma_{dn}^{1/2}) \) and \( \rho(\Gamma_{dn}^{-1/2}) \) are bounded from above, the desired results hold if \( \log^2(n) \cdot r_{t,n} = o(1) \).
The corresponding result for their inverses follows from
\[
\Gamma_{dn}^{-1/2} - (\hat{\Gamma}_{\kappa,l})^{-1/2} = (\hat{\Gamma}_{\kappa,l})^{-1/2} \left( (\hat{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2} \right) \Gamma_{dn}^{-1/2},
\]
which also implies boundedness of \(\rho((\hat{\Gamma}_{\kappa,l})^{1/2})\) and \(\rho((\hat{\Gamma}_{\kappa,l})^{-1/2})\) from above and from below, which concludes this proof. □

**Proof of Theorem 4.1.**

Let \(Z^i\) be the bootstrap sample that is defined analogue to \(Z^i\) in Step 4 of Section 3 except that the resample is drawn from the standardized values of \(\bar{W} = \Gamma_{dn}^{-1/2} Y\) instead of \(\bar{W} = (\hat{\Gamma}_{\kappa,l})^{-1/2} Y\). Now, the proof of Theorem 4.1 proceeds through a sequence of lemmas. Lemma 7.1(i) gives the justification for using \(\sim Z^i\) instead of \(Z^i\). Furthermore, we define \(C(k)(h) = C(h)1(|h| \leq k)\) and let \(\Gamma_{dn,k} = (C(k)(i-j), i,j = 1, \ldots, n)\) be the \(k\)-banded version of \(\Gamma_{dn}\). The matrix \(\Gamma_{dn,k}\) is banded in the sense that only the \((2k+1)d\) main diagonals are not equal to zero and for all sequences \(k = k(n) \to \infty\), it holds \(\rho(\Gamma_{dn,k} - \Gamma_{dn}) \to 0\) as \(n \to \infty\) due to \(\rho(\Gamma_{dn,k} - \Gamma_{dn}) \leq 2 \sum_{k(k+1)}^{\infty} C(h)1\), which is obtained analogue to the proof of Theorem 2.1. Let \(\Gamma_{dn,k}\) be the Cholesky decomposition of \(\Gamma_{dn,k}\) which exists for sufficiently large \(k\) by (A3) and note that only its main diagonal and the \((k+1)\) secondary diagonals below the main diagonal contain non zero elements which are all bounded by \(\max_i C_i^{1/2}(0)\). The second part of Lemma 7.1 allows us also to replace \(\Gamma_{dn}^{1/2}\) in (7.4) by \(\Gamma_{dn,k}^{1/2}\) and to get asymptotically the same results. Lemma 7.2 gives the proper limiting bootstrap variance, while Lemma 7.3 proves boundedness in probability of \(E^*(Z_i^2)\) and Lemma 7.4 deals with asymptotic normality of \(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Y}_i^*\).

**Lemma 7.1.** Under the assumptions of Theorem 4.1, it holds

\[
(i) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Y}_i^* - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Y}_i^* = O_p(1) \quad \text{and} \quad (ii) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Y}_i^* - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Y}_i^* = O_p(1),
\]

where \(\tilde{Y}_i^*\) is the bootstrap sample that is defined analogue to \(Y_i^*\) in (7.4).

**Proof.** (i) Let \(J = [\mathbf{I}_d : \cdots : \mathbf{I}_d]\) be the \((d \times dn)\) matrix that consists of \(n\) \((d \times d)\) unit matrices. In the following, we show that
\[
\frac{1}{\sqrt{n}} b^T \sum_{t=1}^{n} Y_t^* = \frac{1}{\sqrt{n}} b^T J(\Gamma_{dn})^{1/2} \tilde{Z}^* + \frac{1}{\sqrt{n}} b^T \Gamma_{dn}^{1/2} (\tilde{Z}^* - \tilde{Z}^*) + \frac{1}{\sqrt{n}} b^T J(\tilde{\Gamma}_{\kappa,l})^{1/2} \tilde{Z}^* + O_P(\log^2(n) \cdot r_{t,n})
\]
for any \(\mathbb{R}^d\)-valued vector \(b\). We have
\[
\frac{1}{\sqrt{n}} b^T \sum_{t=1}^{n} Y_t^* = \frac{1}{\sqrt{n}} b^T J(\Gamma_{dn})^{1/2} \tilde{Z}^* + \frac{1}{\sqrt{n}} b^T \Gamma_{dn}^{1/2} (\tilde{Z}^* - \tilde{Z}^*) + \frac{1}{\sqrt{n}} b^T J(\tilde{\Gamma}_{\kappa,l})^{1/2} \tilde{Z}^* + O_P(\log^2(n) \cdot r_{t,n})
\]
and it remains to show that \(R_1^*\) and \(R_2^*\) are asymptotically negligible. Considering \(R_2^*\), we get

\[
E^*(R_2^*) = \frac{1}{n} b^T J((\hat{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2})(\tilde{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2} J \tilde{b}^T_2
\]
\[
\leq \frac{1}{n} 2 \lambda_{\text{max}} \left( (\hat{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2} \left( (\hat{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2} \right)^T \right)
\]
\[
= \lambda_{\text{max}}^2 \left( (\hat{\Gamma}_{\kappa,l})^{1/2} - \Gamma_{dn}^{1/2} \right)
\]
and this leads to $R_3^s = O_P^* (\log^2(n) \cdot r_{l,n})$ by Corollary 2.1. Now, we turn to $R_4^s$. First of all, observe that $Z^*$ and $\tilde{Z}^*$ may be represented as

$$Z^* = M^* \frac{1}{\sigma_W} (I_{dn} - \frac{1}{dn} 1_{dn \times dn}) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} Y,$$

$$\tilde{Z}^* = M^* \frac{1}{\sigma_W} (I_{dn} - \frac{1}{dn} 1_{dn \times dn}) \big( \Gamma_{dn} \big)^{-1/2} Y,$$

where $I_{dn}$ is the $(dn \times dn)$ unit matrix, $1_{dn \times dn}$ is the $(dn \times dn)$ matrix of ones and each row of the $(dn \times dn)$ matrix $M^*$ is independently and uniformly selected from the standard basis vectors $e_1, \ldots, e_{dn}$. This yields

$$R_1^s = \frac{1}{\sigma_W} \sqrt{n} b^T J\Gamma_{dn}^{1/2} M^* \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$+ \frac{1}{\sigma_W} \sqrt{n} b^T J\Gamma_{dn}^{1/2} M^* \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} \right) Y$$

$$= R_3^s + R_4^s$$

and for $R_3^s$, we get

$$E^* (R_3^s) = E^* \left( \frac{1}{\sigma_W} \sqrt{n} b^T J\Gamma_{dn}^{1/2} M^* \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$E^* (\tilde{V}^*V^*^T) = \frac{1}{\sigma_W} \sqrt{n} b^T J\Gamma_{dn}^{1/2} E^* \left( \tilde{V}^*V^*^T \right) \big( \Gamma_{dn} \big)^{1/2} J\Gamma_{dn}^{1/2} b,$$

where $\tilde{V}^*$ is a bootstrap vector drawn from $\left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$. Due to independent resampling, it holds $E^* \left( \tilde{V}^*V^*^T \right) = \sigma_V^2 I_{dn}$ with

$$\sigma_V^2 = \frac{1}{dn} Y^T \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$\times E^* \left( M_{1\cdot}^* M_{1\cdot}^* \right) \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$\leq \frac{1}{dn} Y^T \left( \hat{\Gamma}_{n,l}^* \right)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$\times E^* \left( M_{1\cdot}^* M_{1\cdot}^* \right) \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) \big( \hat{\Gamma}_{n,l}^* \big)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

$$\leq \frac{1}{dn} Y^T Y \lambda_{\max} \left( \left( \hat{\Gamma}_{n,l}^* \right)^{-1/2} - \left( \Gamma_{dn} \right)^{-1/2} \right) Y$$

where $M_{1\cdot}^*$ denotes the $j$th row of $M^*$ and $E^* \left( M_{1\cdot}^* M_{1\cdot}^* \right) = \frac{1}{dn} I_{dn}$. Thanks to $Y^T Y = O_P (dn)$, $\rho((\hat{\Gamma}_{n,l}^*)^{-1/2} - (\Gamma_{dn})^{-1/2})^T = O_P (\log^2(n) \cdot r_{l,n})$ and $\rho(I_{dn} - \frac{1}{dn} 1_{dn \times dn}) = O(1)$, we obtain

$$E^* (R_3^s) \leq |\sigma_W^2 \sqrt{n} b^T J\Gamma_{dn}^{1/2} \lambda_{\max} (\Gamma_{dn}) = \sigma_W^2 |b^T J\Gamma_{dn}^{1/2}$$

resulting in $R_3^s = O_P^* (\log^2(n) \cdot r_{l,n})$ because $\sigma_W^2$ is bounded away from zero and from above with probability tending to one. Finally, since the same holds true for $\sigma_V^2$, to handle $R_4^s$, it is
sufficient to show \(|\hat{\sigma}_W^2 - \hat{\sigma}_W^2| = O_P(\log^2(n) \cdot r_{t,n})\). By plugging-in, we have
\[
|\hat{\sigma}_W^2 - \hat{\sigma}_W^2| = \frac{1}{dn} \left( Y^T \left( I_{dn} - \frac{1}{dn} 1_{dn \times dn} \right) + \frac{1}{dn} \right) \left( \Gamma_{dn}^{-1/2} - \Gamma_{dn}^{-1/2} \right) Y \epsilon
\]

and by Cauchy-Schwarz inequality, the first term above is bounded by
\[
\frac{1}{dn} \left( Y^T \left( \Gamma_{dn}^{-1/2} - \Gamma_{dn}^{-1/2} \right) \right) \frac{1}{dn} \left( \Gamma_{dn}^{-1/2} - \Gamma_{dn}^{-1/2} \right) Y \epsilon
\]

Analogue computations for the second term leads to \(R_1^* = O_P(\log^2(n) \cdot r_{t,n})\), which concludes the proof of the first assertion. The claimed equality in (ii) is obtained from a similar calculation as executed for \(R_2^*\), where \(\rho(\Gamma_{dn,k} - \Gamma_{dn}^{-1/2}) = O(\log^2(n) \cdot s_k)\) is used. The last result follows as in the proof of Corollary 2.1, from (7.3) with \((\hat{\Gamma}_{dn,j})^{1/2}\) replaced by \(\Gamma_{dn,k}^{1/2}\) and \(\rho(\Gamma_{dn,k}^{1/2} - \Gamma_{dn}^{1/2}) = O(s_k)\), where \(s_k = \sum_{h=1}^{\infty} C(h)\).

**Lemma 7.2.** Under the assumptions of Theorem 4.1, \(\text{Var}^t(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Z}_t^u) = \sum_{h \in \mathbb{Z}} C(h) + o_P(1)\).

**Proof.** For any \(\mathbb{R}^d\)-valued vector \(b\), we get by standard arguments
\[
\text{Var}^t\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Z}_t^u \right) = \frac{1}{n} \text{Var}^t \left( \frac{1}{n} \sum_{i,j=1}^{n} C(i-j) \right) b b^T \left( \sum_{h=-\infty}^{\infty} C(h) \right) b + o(1).
\]

**Lemma 7.3.** Under (A1)–(A4) with some \(q \geq 2\), we have \(E^t(\tilde{Z}_1^u)^q = O_P(1)\). Under (A1')–(A4'), we have \(E^t(\tilde{Z}_1^u)^q = O_P(d^{q/2})\).

**Proof.** Due to \(E^t(\tilde{Z}_1^u)^q = \frac{1}{dn} \sum_{t=1}^{dn} |\tilde{Z}_t^u|^q\), we show \(E(\tilde{Z}_1^u)^q = \| \tilde{Z}_1^u \|_q^q = O(1)\) uniformly in \(t\). By defining (\(I_{dn} - \frac{1}{dn} 1_{dn \times dn}) \left( \Gamma_{dn}^{1/2} - \Gamma_{dn}^{1/2} \right) Y = \frac{1}{\hat{\sigma}_W^2} \left( \sum_{j=1}^{n} \hat{A}_j^T(j)(\mathbf{X}_j - \mu) - \sum_{j=1}^{n} \hat{A}_j^T(j)(\mathbf{X}_j - \mu) \right)\) and thanks to boundedness of \(\frac{1}{\hat{\sigma}_W^2}\), it suffices to consider the two terms in parentheses above in the following. For the second one, we get from triangle inequality and (A4) that
\[
\left\| \sum_{j=1}^{n} \hat{A}_j^T(j)(\mathbf{X}_j - \mu) \right\|_q^q = \left\| \sum_{j=1}^{n} \sum_{s=1}^{d} a_{t,(j-1)d+s} (\mathbf{X}_s - \mu_s) \right\|_q^q \leq \left( \sum_{j=1}^{n} \sum_{s=1}^{d} |a_{t,(j-1)d+s}| \right) O_P(n^{-q/2})
\]
holds and together with
\[ \sum_{j=1}^{\infty} \sum_{s=1}^{\infty} |a_{t,(j-1)d+s}| = \sum_{j=1}^{dn} |a_{t,i}| \leq \left( \sum_{j=1}^{dn} a_{t,i}^2 \right)^{1/2} \leq \sqrt{dn} \rho(\Gamma_{dn}^{-1/2}) = O(\sqrt{dn}) \] (7.5)
by Cauchy-Schwarz, this term is of order \( O(d^{d/2}) \) and bounded in probability for fixed \( d \). Now, we turn to the first term. We define \( P_{t-m}X_j = E(X_j - \mu|F_{j-m}) - E(X_j - \mu|F_{j-m-1}) \), where \( F_t = \sigma(X_t - \mu, s \leq t) \) and set \( M_{m,n,t} = \sum_{j=1}^{n} A_T(j) \{ E(X_j - \mu|F_{j-m}) - E(X_j - \mu|F_{j-m-1}) \} \). This yields \( \sum_{j=1}^{n} A_T(j)(X_j - \mu) = \sum_{m=0}^{\infty} M_{m,n,t} \) almost surely and now the desired result follows from \( \| \sum_{m=0}^{\infty} M_{m,n,t} \|_q \leq (\sum_{m=0}^{\infty} \| M_{m,n,t} \|_q)^{q} < \infty \) uniformly in \( t \). By Proposition 4 in Dedecker and Doukhan (2003), we get \( \| M_{m,n,t} \|_q \leq (2q \sum_{i=1}^{n} b_{t,m,n,t})^{1/2} \), where
\[ b_{t,m,n,t} = \max_{1 \leq i \leq n} \| A_T(i)P_{t-m}X_i \|_q \sum_{k=i}^{t} E(\| A_T(k)P_{t-m}X_k \|_q) \]
with \( M_i = \sigma(A_T(i)P_{t-m}X_k, 0 \leq k \leq i) \). For the conditional expectations above, we obtain
\[ E(A_T(i)P_{t-m}X_i | M_i) = A_T(i)P_{t-m}X_i \]
for \( k = i \) and
\[ E(A_T(i)P_{t-m}X_i | M_i) = A_T(i)E(\| M_i \|_q) \]
for all \( k > i \) due to \( M_i \subset \sigma(F_{k-m-1}) \). This leads to
\[ b_{t,m,n,t} = \| (A_T(i)P_{t-m}X_i)^2 \|_q / 2 \leq \| A_T(i)A_T(i)(P_{t-m}X_i)^2P_{t-m}X_i \|_q / q \]
by Cauchy-Schwarz, where the first factor equals \( \sum_{i=1}^{dn} a_{t,i}^2 \) and the second is bounded by
\[ \sum_{p=1}^{dn} (E(P_{t-m}X_{p,i})^2)^2 / 2 \leq d \max_{p=1,...,d} (E(P_{t-m}X_{p,i})^2)^2 / q = d \max_{p=1,...,d} (E(P_0X_{p,m})^2)^2 / q. \]
This results in
\[ \| M_{m,n,t} \|_q^2 \leq 2q \sum_{i=1}^{dn} a_{t,i}^2 \max_{p=1,...,d} (E(P_0X_{p,m})^2)^1 / 2 = 2qd \sum_{i=1}^{dn} a_{t,i}^2 \max_{p=1,...,d} ||P_0X_{p,m}||_q^2 \] (7.6)
and \( (\sum_{m=0}^{\infty} \| M_{m,n,t} \|_q)^q = O(d^{d/2}) \) by similar arguments used to get (7.5). In particular, the last term is bounded in probability for fixed \( d \), which concludes this proof. \( \square \)

**Lemma 7.4.** Under the assumptions of Theorem 4.1, \( \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \tilde{Y}^{x,k} \) converges in distribution to a centered normal distribution with variance obtained in Lemma 7.2.

**Proof.** By Lemma 7.1(ii), we may consider \( b_T \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \tilde{Y}^{x,k} \) for some \( \mathbb{R}^d \)-valued vector \( b \) and prove its asymptotic normality by using the Cramér-Wold device. We have
\[ b_T \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \tilde{Y}^{x,k} = \frac{1}{\sqrt{n}} b_TJ^{(1/2)} \tilde{Z}^x = \sum_{i=1}^{dn} \frac{n^{-1/2}b_TJ^{(1/2)}_{dn,k}(\bullet, i)\tilde{Z}^x_i}{ \sqrt{n}} = \sum_{i=1}^{dn} U_{i}^x \] (7.7)
with an obvious notation for \( U_{i}^x \), where \( J^{(1/2)}_{dn,k}(\bullet, i) \) denotes the \( i \)th column of \( J^{(1/2)}_{dn,k} \). Then, the desired result follows from a CLT for triangular arrays [cf. Billingsley (1995, p.362)] which follows if the Lyapunov condition (for \( \delta = 2 \)), i.e.
\[ \frac{1}{Var^*(\sum_{i=1}^{dn} U_{i}^x)^2} \sum_{i=1}^{dn} E^*(U_{i}^{x^4}) \rightarrow 0 \] (7.8)
in probability as \( n \to \infty \) is satisfied. Considering the denominator, we get that

\[
\left( \sum_{i=1}^{dn} E^*(U_i^2) \right)^2 \leq \left( \sum_{i=1}^{dn} \frac{1}{n} b'^2 \mathbf{J}^{1/2}_{dn,k}(\cdot, i) b'^T \mathbf{J}^{1/2}_{dn,k}(\cdot, i) \right)^2 \leq \left( \frac{1}{n} b'^2 \mathbf{J}^{1/2}_{dn,k} \mathbf{J}^T b' \right)^2
\]

is bounded from below and from above due to \( \rho(\mathbf{J}_{dn,k} - \mathbf{G}_{dn}) \to 0 \) for any \( k \to \infty \) and (A3). With Lemma 7.3, we get for the numerator

\[
\sum_{i=1}^{dn} E^*(U_i^4) = \sum_{i=1}^{dn} \frac{1}{n^2} \left( \frac{n}{2} \mathbf{J}^{1/2}_{dn,k}(\cdot, i) \right)^4 E^*(\tilde{Z}_i^4) \leq \sum_{i=1}^{dn} \frac{1}{n^2} \left( b'^{1/2} \mathbf{J}^{1/2}_{dn,k}(\cdot, i) \right)^2 O_P(d^2)
\]

and with \( |\mathbf{J}^{1/2}_{dn,k}(\cdot, i)|^2 = O(k^4d^2) \) uniformly in \( i \), we obtain \( \sum_{i=1}^{dn} E^*(U_i^4) = O_P(d^4k^4/n) \). Altogether, this leads to

\[
1 \left( \text{Var}^*(\sum_{j=1}^{n} U_j) \right)^2 \sum_{j=1}^{n} E^*(U_j^4) = O_P \left( \frac{k^4}{n} \right) = o_P(1)
\]

for some appropriate sequence \( k = k(n) \) that satisfies \( \log^2(n)s_k = o(1) \) and \( k^4 = o(n) \), which is assured to exist by (A1) with some \( g > 0 \). \( \square \)

**Proof of Theorem 4.2.**

Let \( \tilde{f}_n(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \tilde{Y}_t^* e^{-it\omega} \) and \( \tilde{f}_{n,k}(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \tilde{Y}_{-1}^{*k} e^{-it\omega} \) be the discrete Fourier transforms based on \( \tilde{Y}_1^*, ..., \tilde{Y}_n^* \) and \( \tilde{Y}_{-1}^{*k}, ..., \tilde{Y}_n^{*k} \) as defined in (7.4) and in Lemma 7.1, respectively. The corresponding periodograms are denoted by \( \hat{f}_n(\omega) = \tilde{f}_n(\omega) \tilde{f}_{n,H}(\omega) \) and \( \hat{f}_{n,k}(\omega) = \tilde{f}_{n,k}(\omega) \tilde{f}_{n,k,H}(\omega) \) such that

\[
\hat{f}_n(\omega) = \frac{1}{n} \sum_{j=-[\frac{n}{2}]}^{[\frac{n}{2}]} K_1(\omega - \omega_j) \hat{f}_n(\omega_j)
\]

and \( \hat{f}_{n,k}(\omega) \) analogue with \( \hat{f}_n(\omega_j) \) replaced by \( \hat{f}_{n,k}(\omega_j) \) are approximations to the bootstrap kernel spectral density estimator \( \hat{f}(\omega) \). The proof proceeds through a sequence of lemmas. Lemma 7.5 gives the justification for considering

\[
\sqrt{n}\left( \hat{f}_{np}(\omega) - \hat{f}(\omega) \right) = \sqrt{n}\left( \hat{f}_{np}(\omega) - E^*(\hat{f}_{np}(\omega)) \right) + \sqrt{n}\left( E^*(\hat{f}_{np}(\omega)) - \hat{f}(\omega) \right)
\]

in the following and to prove the CLT for these expressions, where \( \hat{f} \) as defined in Lemma 7.5 below. Lemma 7.6 gives the covariance structure of the stochastic leading term above, Lemma 7.7 deals with the asymptotics of the bias term and Lemma 7.8 provides asymptotic normality.

**Lemma 7.5.** Under the assumptions of Theorem 4.2, it holds

\[
(i) \quad \mathbf{J}_n(\omega) - \hat{\mathbf{J}}_n(\omega) = o_P(1) \quad \text{and} \quad (ii) \quad \mathbf{J}_n(\omega) - \hat{\mathbf{J}}_{n,k}(\omega) = o_P(1)
\]

uniformly in \( \omega \), respectively. Further, it holds

\[
(iii) \quad \sqrt{n}\left( \tilde{f}_n(\omega) - \hat{f}(\omega) \right) = o_P(1) \quad \text{and} \quad (iv) \quad \sqrt{n}\left( \hat{f}_n(\omega) - \hat{f}_{n,k}(\omega) \right) = o_P(1)
\]

and, for \( \tilde{f}(\omega) = \frac{1}{2\pi} \sum_{h=-n-1}^{n-1} C(h)e^{-ih\omega} \), we have \( v) \quad \sqrt{n}\left( \hat{f}(\omega) - \hat{f}(\omega) \right) = o_P(1) \) for all \( \omega \).

**Proof.** (i) Let \( \mathbf{J}_n(\omega) = (e^{-i1\omega}, ..., e^{-in\omega}) \odot \mathbf{I}_d \) and \( b \in \mathbb{C}^d \). Then, we have

\[
\frac{1}{\sqrt{2\pi n}} \mathbf{J}_n^T (\mathbf{J}_n^*) = \frac{1}{\sqrt{2\pi n}} \mathbf{J}_n^T \mathbf{J}_\omega \left( (\mathbf{I}_n)^{1/2} (\mathbf{Z}^* - \mathbf{Z}) + (\mathbf{I}_n^{1/2} - (\mathbf{I}_n)^{1/2}) (\mathbf{Z}) \right)
\]
and analogue to the proof of Lemma 7.1 we obtain the claimed result, where uniformity follows from the fact that $b_i^2 J_0 J_i^{-1} b$ is independent of $\omega$. Part (ii) follows similarly. (iii) Plugging-in for $\tilde{f}^*(\omega)$ and $\hat{f}^*(\omega)$, yields

$$
\rho(\sqrt{nb}(\tilde{f}^*(\omega) - \hat{f}^*(\omega))) \leq \sqrt{\frac{b}{n}} \sum_{j=-\left\lfloor \frac{n}{b} \right\rfloor}^{\left\lfloor \frac{n}{b} \right\rfloor} K_b(\omega - \omega_j) \rho \left(J_n^*(\omega_j) \left(J_n^*(\omega_j) - J_n^*(\omega_j) \right)^H \right)
$$

By using part (i) of this lemma, we get

$$
A_1 \leq \sqrt{\frac{b}{n}} \sum_{j=-\left\lfloor \frac{n}{b} \right\rfloor}^{\left\lfloor \frac{n}{b} \right\rfloor} K_b(\omega - \omega_j) \rho \left(J_n^*(\omega_j) \right) = O_P(\sqrt{\log^2(n) r_{1,n}})
$$

and an analogue result for $A_2$. Part (iv) follows in the same way and is omitted. (v) By plugging-in for $f(\omega)$ and $\hat{f}(\omega)$, we get

$$
|\sqrt{nb}(\tilde{f}(\omega) - \hat{f}(\omega))| \leq \frac{\sqrt{nb}}{2\pi} \sum_{h=-(n-1)}^{n-1} (1 - \kappa_i(h))|C(h)|_1 + \frac{\sqrt{nb}}{2\pi} \sum_{h=-(n-1)}^{n-1} \kappa_i(h)|C(h) - \hat{C}(h)|_1
$$

and the first summand above is of order $O_P(\sqrt{nb} \cdot s_1)$ and the second one is $O_P(\sqrt{b})$.

\textbf{Lemma 7.6.} Under the assumptions of Theorem 4.2, it holds

$$
nbCov^* \left( \tilde{f}_{pq}^*(\omega), \tilde{f}_{rs}^*(\lambda) \right) = \left( f_{pr}(\omega) f_{qs}(\omega) \delta_{\omega,\lambda} + f_{ps}(\omega) f_{qr}(\omega) \tau_{0,0} \right) \frac{1}{2\pi} \int K^2(v) \, dv + O_P(1)
$$

for all $p, q, r, s = 1, \ldots, d$ and all $\omega, \lambda \in [0, \pi]$.

\textbf{Proof.} We have

$$
nbCov^* \left( \tilde{f}_{pq}^*(\omega), \tilde{f}_{rs}^*(\lambda) \right) = \frac{b}{n} \sum_{k_1, k_2=-\left\lfloor \frac{n}{b} \right\rfloor}^{\left\lfloor \frac{n}{b} \right\rfloor} K_b(\omega - \omega_{k_1}) K_b(\lambda - \omega_{k_2}) Cov^* \left( \tilde{T}_{n,pq}(\omega_{k_1}), \tilde{T}_{n,rs}(\omega_{k_2}) \right)
$$

and the conditional covariance on the last right-hand side above becomes

$$
\frac{1}{4\pi^2 n^2} \sum_{t_1, t_2, t_3, t_4=1}^{\left\lfloor \frac{n}{b} \right\rfloor} \Gamma_{d_n}^{1/2}(\Gamma_{d_n}^{1/2}(t_1 - 1)d + p, i_1) \Gamma_{d_n}^{1/2}(t_2 - 1)d + q, i_2) \Gamma_{d_n}^{1/2}(t_3 - 1)d + r, i_3)
$$

$$
\times \Gamma_{d_n}^{1/2}(t_4 - 1)d + s, i_4) Cov^* \left( \tilde{Z}_{t_1}^*, \tilde{Z}_{t_2}^*, \tilde{Z}_{t_3}^*, \tilde{Z}_{t_4}^* \right) e^{-i(t_1 - t_2)\omega_{k_1}} e^{i(t_3 - t_4)\omega_{k_2}}
$$

(7.10)

and due to i.i.d. resampling, we have

$$
Cov^* \left( \tilde{Z}_{t_1}^*, \tilde{Z}_{t_2}^*, \tilde{Z}_{t_3}^*, \tilde{Z}_{t_4}^* \right) = \begin{cases} 1, & i_1 = i_3, i_2 = i_4 \text{ or } i_1 = i_4, i_2 = i_3 \\ E^* \left( \tilde{Z}_{i_1}^4 \right) - 3, & i_1 = i_2 = i_3 = i_4 \\ 0, & \text{otherwise} \end{cases}
$$

(7.11)
Both combinations of the first case in (7.11) together with (7.10) lead to
\[
\frac{1}{2\pi n} \sum_{t_1,t_3=1}^{n} C_{pr}(t_1 - t_3)e^{-it_1\omega_1_1}e^{it_3\omega_2} + \frac{1}{2\pi n} \sum_{t_2,t_4=1}^{n} C_{qs}(t_2 - t_4)e^{-it_2\omega_1_1}e^{-it_4\omega_2}
\]
\[
\frac{1}{2\pi n} \sum_{t_1,t_4=1}^{n} C_{ps}(t_1 - t_4)e^{-it_1\omega_1_1}e^{-it_4\omega_2} + \frac{1}{2\pi n} \sum_{t_2,t_3=1}^{n} C_{qr}(t_2 - t_3)e^{it_2\omega_1_1}e^{it_3\omega_2}
\]
\[
= f_{ps}(\omega_1_1)f_{qs}(\omega_1_1)\mathbf{1}(j_1 = j_2) + f_{ps}(\omega_1_1)f_{qr}(\omega_1_1)\mathbf{1}(j_1 = -j_2) + o(1)
\]
(7.12)
uniformly in \(\omega_1_1\) and \(\omega_2\), respectively. To handle the second case in (7.11), observe that we may replace all entries of \(\Gamma_{dn,k}^{1/2}\) by the corresponding entries of its \(k\)-banded version \(\Gamma_{dn,k}^{1/2}\) in (7.10) by Lemma 7.5. Therefore, we obtain
\[
\frac{1}{4\pi^2 n^2} \sum_{i=1}^{dn} \left( \sum_{t_1=1}^{\Gamma_{dn,k}^{1/2}((t_1 - 1)d + p,i) e^{-it_1\omega_1_1} \right) \left( \sum_{t_2=1}^{\Gamma_{dn,k}^{1/2}((t_2 - 1)d + q,i) e^{it_2\omega_1_1} \right)
\]
\[
\times \left( \sum_{t_3=1}^{\Gamma_{dn,k}^{1/2}((t_3 - 1)d + r,i) e^{it_3\omega_2} \right) \left( \sum_{t_4=1}^{\Gamma_{dn,k}^{1/2}((t_4 - 1)d + s,i) e^{-it_4\omega_2} \right) \left( E^*(\tilde{Z}_1^4) - 3 \right)
\]
\[
= O_P \left( \frac{k^4}{n} \right)
\]
uniformly in \(\omega_1_1\) and \(\omega_2\) due to the banded shape of \(\Gamma_{dn,k}^{1/2}\) and Lemma 7.3. Together with (7.9), the second case in (7.11) becomes \(O_P(bk^4) = o_P(1)\) under the assumptions. \(\square\)

**Lemma 7.7.** Under the assumptions of Theorem 4.2, it holds
\[
E^*(\tilde{f}_{pq}^*(\omega)) - \tilde{f}_{pq}(\omega) = b^2 f_{pq}''(\omega) \frac{1}{4\pi} \int_{-\pi}^{\pi} v^2 K(v)dv + o_P(b^2)
\]
for all \(p,q = 1, \ldots, d\) and all \(\omega\).

**Proof.** Straightforward calculations yield
\[
E^*(\tilde{f}^*(\omega)) - \tilde{f}(\omega) = \frac{1}{n} \sum_{j=1}^{n} K_b(\omega - \omega_j) \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \left( 1 - \frac{|h|}{n} \right) C(h) \left( e^{-ih\omega_j} - e^{-ih\omega} \right)
\]
\[
+ \left( \frac{1}{n} \sum_{j=1}^{n} K_b(\omega - \omega_j) - 1 \right) \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} C(h) e^{-ih\omega},
\]
where the second term is negligible thanks to \(\left| \frac{1}{n} \sum_{j=1}^{n} K_b(\omega - \omega_j) - 1 \right| = O(\frac{1}{nb})\). The first term can be treated as in Lemma 7.4 in Jentsch and Kreiss (2010), which concludes this proof. \(\square\)

**Lemma 7.8.** Under the assumptions of Theorem 4.2, \(\sqrt{n}\tilde{f}_{pq}^*(\omega_l) - E^*(\tilde{f}_{pq}^*(\omega_l)) : p,q = 1, \ldots, d; l = 1, \ldots, s\) converges in distribution in probability to a centered \(d^2 s\)-dimensional normal distribution with covariance matrix as obtained in Lemma 7.6.

**Proof.** We prove asymptotic normality only for \(\sqrt{n}\tilde{f}_{pq}^*(\omega_l) - E^*(\tilde{f}_{pq}^*(\omega_l))\) in the following and a more general result follows from the Cramér-Wold device. First note that the quantity of
interest may be expressed as a generalized quadratic form, i.e.

\[
\sqrt{n}b \left( \hat{f}^*_pq(\omega) - E^* \left( \hat{f}^*_{pq}(\omega) \right) \right)
= \sum_{i_1,i_2=1}^{dn} \frac{\sqrt{b}}{2\pi \sqrt{n}} \sum_{t_1,t_2=1}^{n} \left( \frac{1}{n} \sum_{j=-[n^{1/2}]}^{[n^{1/2}]} K_b(\omega - \omega_j) e^{-i(t_1-t_2)\omega_j} \right) \\
\times \Gamma^{1/2}_{dn}((t_1 - 1)d + p, i_1) \Gamma^{1/2}_{dn}((t_2 - 1)d + q, i_2) \left( \tilde{Z}^*_i \tilde{Z}^*_j - E^*(\tilde{Z}^*_i \tilde{Z}^*_j) \right)
= \sum_{i_1,i_2=1}^{dn} w_{i_1,i_2} (\tilde{Z}^*_i, \tilde{Z}^*_j)
= \sum_{1 \leq i_1 < i_2 \leq dn} w_{i_1,i_2} + \sum_{i=1}^{dn} w_{ii} (\tilde{Z}^*_i, \tilde{Z}^*_i)
\]

with an obvious notation for \(w_{i_1,i_2} (\tilde{Z}^*_i, \tilde{Z}^*_j)\) and with \(W_{i_1,i_2} := w_{i_1,i_2} (\tilde{Z}^*_i, \tilde{Z}^*_j) + w_{i_2,i_1} (\tilde{Z}^*_i, \tilde{Z}^*_i)\). Due to i.i.d. resampling, we can apply Theorem 2.1 of deJong (1987) to the quadratic form. It is clean by an easy computation and its variance is bounded, such that it remains to show that

\[
(i) \quad \max_{1 \leq i_1 \leq dn} E^* \left( \left| W_{i_1,i_2} \right|^2 \right) \to 0 \quad \text{and} \quad (ii) \quad \frac{E^* \left| \sum_{1 \leq i_1 < i_2 \leq dn} W_{i_1,i_2} \right|^2}{\left( E^* \left| \sum_{1 \leq i_1 < i_2 \leq dn} W_{i_1,i_2} \right|^2 \right)^{1/2}} \to 3
\]

hold in probability, respectively. To show (i), examplarily, we consider only the first term of \(\left| W_{i_1,i_2} \right|^2\) in more detail, that is, \(\left| w_{i_1,i_2} (\tilde{Z}^*_i, \tilde{Z}^*_j) \right|^2\) and we get

\[
E^* \left( \left| w_{i_1,i_2} (\tilde{Z}^*_i, \tilde{Z}^*_j) \right|^2 \right) \quad (7.13)
= \frac{b}{4\pi^2 n} \sum_{t_1,t_2,t_3,t_4=1}^{n} \left( \frac{1}{n} \sum_{j_1=-[n^{1/2}]}^{[n^{1/2}]} K_b(\omega - \omega_{j_1}) e^{-i(t_1-t_2)\omega_{j_1}} \right) \left( \frac{1}{n} \sum_{j_2=-[n^{1/2}]}^{[n^{1/2}]} K_b(\omega - \omega_{j_2}) e^{i(t_3-t_4)\omega_{j_2}} \right)
\times \Gamma^{1/2}_{dn}((t_1 - 1)d + p, i_1) \Gamma^{1/2}_{dn}((t_2 - 1)d + q, i_2) \Gamma^{1/2}_{dn}((t_3 - 1)d + p, i_1) \Gamma^{1/2}_{dn}((t_4 - 1)d + q, i_2) \text{Var}^* (\tilde{Z}^*_i \tilde{Z}^*_j).
\]

The conditional variance above equals \(1 + (E^*(\tilde{Z}^*_i^4) - 2)1(i_1 = i_2)\) and we may replace all entries of \(\Gamma^{1/2}_{dn}\) by the corresponding entries of its k-banded version \(\Gamma^{1/2}_{dn,k}\) by Lemma 7.5 in the following. For any fixed \(i_1\), we obtain by summing over \(i_2\) for the first case

\[
\frac{b}{4\pi^2 n} \sum_{t_1,t_2,t_3,t_4=1}^{n} \left( \frac{1}{n} \sum_{j_1=-[n^{1/2}]}^{[n^{1/2}]} K_b(\omega - \omega_{j_1}) e^{-i(t_1-t_2)\omega_{j_1}} \right) \left( \frac{1}{n} \sum_{j_2=-[n^{1/2}]}^{[n^{1/2}]} K_b(\omega - \omega_{j_2}) e^{i(t_3-t_4)\omega_{j_2}} \right)
\times \Gamma^{1/2}_{dn,k}((t_1 - 1)d + p, i_1) \Gamma_{dn,k}((t_2 - 1)d + q, (t_4 - 1)d + q) \Gamma^{1/2}_{dn,k}((t_3 - 1)d + p, i_1)
= O \left( \frac{k^2}{n} \right) + o \left( b^2 \right)
\]
and with use of Lemma 7.3, we get for the second case
\[
\frac{b}{4\pi^2 n} \sum_{t_1, t_2, t_3, t_4 = 1}^{n} \left( \frac{1}{n} \sum_{j = 1}^{\lfloor \frac{n}{M} \rfloor} K_h(\omega - \omega_{j_1}) e^{-i(t_1 - t_2)\omega_{j_1}} \right) \left( \frac{1}{n} \sum_{j = 1}^{\lfloor \frac{n}{M} \rfloor} K_h(\omega - \omega_{j_2}) e^{i(t_3 - t_4)\omega_{j_2}} \right) \times \Gamma_d^{1/2}(t_1 - 1)d + p, i_1) \Gamma_d^{1/2}(t_2 - 1)d + q, i_2) \right) = O_P\left( \frac{bk^4}{n} \right)
\]
uniformly in \(i_1\), respectively, and both terms above vanish asymptotically for some suitably chosen sequence \(k\) which is possible by the imposed conditions. Being concerned with (ii), we consider first the numerator and get
\[
E^s \left( \sum_{1 \leq i_1 < i_2 \leq dn} W_{i_1i_2}^{4} \right) = E^s \left( \sum_{i_1, i_2 = 1}^{dn} w_{i_1i_2} (\tilde{Z}_{i_1}, \tilde{Z}_{i_2})^{4} \right)
\]
and by expanding the above expression, we see that the most contributing case is where we have four twins of equal indices. All other cases are of lower order. If we take all summands above into account, only three combinations of twins do not vanish and each of them yields \(\text{const}^2 \cdot f^{4}_p(\omega)\). For the denominator, we get \((\text{const} \cdot f^{4}_p(\omega))^2\) which concludes this proof. For details compare for instance the proof of Theorem 2 in Jentsch (2012).

**Proof of Theorem 5.1.**
Under assumptions (A1') and (A2'), we get the same bounds as obtained in Theorems 2.1 and 2.2, respectively, and \(|C(h)|_1 \leq d^2 \sup_{n \in \mathbb{N}} \sup_{i,j=1,...,d(n)} |C_{ij}(h)|\) leads to the first part of (i). By similar arguments as employed in the proof of Corollary 2.1, we get also the second part under (A3') and convergence to zero in probability under the imposed conditions. analogue to the proof of Corollary 2.1, we get (ii) and (iii) by exploiting the bound in (7.3) for the Cholesky factorization.

**Proof of Theorem 5.2.**
We follow the proof of Theorem 4.1 and adopt the notation therein. First, analogue to the proof of Lemma 7.1, we get for any real-valued sequence \(b = b(d(n))\) of \(d(n)\)-dimensional vectors with
\[
0 < M_1 \leq |b(d(n))|^2 \leq M_2 < \infty \quad \text{for all} \quad n \in \mathbb{N}
\]
that the following holds
\[
(i) \quad b^T \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Y}_t^{s} \right) = b^T \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Y}_t^{s,k} \right) + O_P(\log^2(dn) d^2 \tilde{r}_{1,n}),
\]
\[
(ii) \quad b^T \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Y}_t^{s} \right) = b^T \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{Y}_t^{s,k} \right) + O_P(\log^2(dn) d^2 \tilde{s}_k),
\]
where \(\tilde{r}_{1,n}\) is defined in Theorem 5.1 and \(\tilde{s}_k = \sum_{h=k+1}^{\infty} \{ \sup_{n \in \mathbb{N}} \sup_{i,j=1,...,d(n)} |C_{ij}(h)| \}\). Both \(O_P\)-terms above vanish under the imposed conditions. Similar to the proof of Lemma 7.4, we want to show asymptotic normality for (7.7) and, therefore, we have to check the Lyapunov condition (7.8). For the denominator, we get that
\[
\left( \sum_{i=1}^{dn} E^s(U^{s,2}_i) \right)^2 = \left( \sum_{i=1}^{dn} \frac{1}{n} b^T \mathbf{J}^{1/2}_{d,n,k}(\bullet, i) b^T \mathbf{J}^{1/2}_{d,n,k}(\bullet, i) \right)^2 = \left( \frac{1}{n} b^T \mathbf{J}^{1/2}_{d,n,k} b^T \mathbf{J}^{1/2}_{d,n,k} b \right)^2
\]
is bounded from below and from above due to (7.14), (A3') and $\rho(\Gamma_{dn,k} - \Gamma_{dn}) \rightarrow 0$ for a suitable sequence $k \rightarrow \infty$ assured to exist by the imposed assumptions. For the numerator, we get

$$\sum_{i=1}^{dn} E^*(U_i^4) \leq \sum_{i=1}^{dn} \frac{1}{n^2} |B|_2^4 |J_{dn,k}(\cdot,i)|^2 O_P(d^2) = O_P(d^5 k^4/n),$$

where Lemma 7.3 for increasing time series dimension has been used. This gives

$$\frac{1}{(\text{Var}^*(\sum_{j=1}^{n} U_j^2))^2} \sum_{j=1}^{n} E^*(U_j^4) = O_P\left(\frac{d^5 k^4}{n}\right) = o_P(1)$$

for some appropriate sequence $k = k(n)$ that satisfies $\log^2(dn)d^2\varepsilon_k = o(1)$ and $d^5 k^4 = o(n)$, which is assured to exist by (A1') with some sufficiently large $g > 0$. Further, we get

$$|a^2 - \hat{a}^2| = \frac{1}{n} b^T J \left( E(XX^T) - E^*(XX^T) \right) J^T b = \frac{1}{n} b^T J \{\Gamma_{dn} - \hat{\Gamma}_{n,i}\} J^T b$$

$$\leq \left(\frac{1}{n} b^T J J^T b\right)^{1/2} \left(\frac{1}{n} b^T J \{\Gamma_{dn} - \hat{\Gamma}_{n,i}\} \{\Gamma_{dn} - \hat{\Gamma}_{n,i}\} J^T b\right)^{1/2}$$

$$\leq |b|^2 \rho(\Gamma_{dn} - \hat{\Gamma}_{n,i}) = o_P(1)$$

which concludes the proof as $\rho(\Gamma_{dn} - \hat{\Gamma}_{n,i}) = o_P(1)$ and $|b|^2 = O(1)$ by assumption. \hfill $\Box$

**Acknowledgement 1.** 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 the second author was partially supported by NSF grant DMS 13-08319. The authors thank Timothy McMurty for his helping advice on the univariate case and three anonymous referees and the editor who helped to significantly improve the presentation of the paper.

**References**


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

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, SAN DIEGO, LA JOLLA, CA 92093-0112, USA
E-mail address: politis@math.ucsd.edu
Figure 1. Comparison of bootstrap performance: RMSE for estimating the covariance matrix of $\sqrt{n}(\bar{X} - \mu)$ and CR of bootstrap confidence intervals for $\mu$ by MLPB (solid), ARsieve (dashed), MBB (dotted) and TBB (dash-dotted) are reported for the WN model with sample sizes $n \in \{100, 200, 500\}$ and tuning parameters $l, p, s \in \{1, \ldots, 20\}$. The upper half of panels corresponds to $\text{Var}(\sqrt{n}(\bar{X}_1 - \mu_1))$ and $\mu_1$ and the second one to $\text{Var}(\sqrt{n}(\bar{X}_2 - \mu_2))$ and $\mu_2$. Line segments indicate the results for MLPB with data-adaptive individual (grey) and global (black) banding parameter choices.
Figure 2. Comparison of bootstrap performance: RMSE for estimating the covariance matrix of $\sqrt{n}(X - \mu)$ and CR of bootstrap confidence intervals for $\mu$ by MLPB (solid), ARsieve (dashed), MBB (dotted) and TBB (dash-dotted) are reported for the VMA(1) model with sample sizes $n \in \{100, 200, 500\}$ and tuning parameters $l, p, s \in \{1, \ldots, 20\}$. The upper half of panels corresponds to $Var(\sqrt{n}(X_1 - \mu_1))$ and $\mu_1$ and the second one to $Var(\sqrt{n}(X_2 - \mu_2))$ and $\mu_2$. Line segments indicate the results for MLPB with data-adaptive individual (grey) and global (black) banding parameter choices.
Figure 3. Comparison of bootstrap performance: RMSE for estimating the covariance matrix of $\sqrt{n}(X - \mu)$ and CR of bootstrap confidence intervals for $\mu$ by MLPB (solid), ARsieve (dashed), MBB (dotted) and TBB (dash-dotted) are reported for the VAR(1) model with sample sizes $n \in \{100, 200, 500\}$ and tuning parameters $l, p, s \in \{1, \ldots, 20\}$. The upper half of panels corresponds to $\text{Var}(\sqrt{n}(X_1 - \mu_1))$ and $\mu_1$ and the second one to $\text{Var}(\sqrt{n}(X_2 - \mu_2))$ and $\mu_2$. Line segments indicate the results for MLPB with data-adaptive individual (grey) and global (black) banding parameter choices.