Banded and tapered estimates for autocovariance matrices and the linear process bootstrap

Timothy L. McMurry
DePaul University

Dimitris N. Politis *
University of California, San Diego

June 19, 2010

*dpolitis@ucsd.edu

Banded and tapered estimates for autocovariance matrices and the linear process bootstrap

Abstract

We address the problem of estimating the autocovariance matrix of a stationary process. Under short range dependence assumptions, convergence rates are established for a gradually tapered version of the sample autocovariance matrix and for its inverse. The proposed estimator is formed by leaving the main diagonals of the sample autocovariance matrix intact while gradually down-weighting off-diagonal entries towards zero. In addition we show the same convergence rates hold for a positive definite version of the estimator, and we introduce a new approach for selecting the banding parameter. The new matrix estimator is shown to perform well theoretically and in simulation studies. As an application we introduce a new resampling scheme for stationary processes termed the linear process bootstrap (LPB). The LPB is shown to be asymptotically valid for the sample mean and related statistics. The effectiveness of the proposed methods are demonstrated in a simulation study.

Key words: autocovariance matrix; stationary process; bootstrap; block bootstrap; sieve bootstrap

1 Introduction

Let $X_1, \ldots, X_n$ be a realization of a mean zero, stationary process $\{X_t\}_{t\in\mathbb{Z}}$, and let $\gamma_k = \text{cov}\,[X_0, X_k]$ be its autocovariance function. The goal of the present work is to estimate the $n \times n$ autocovariance matrix

$$\Sigma_n = \left[\gamma_{|i-j|}\right]_{i,j=1}^{n}.$$

The lag-k autocovariance γ_k has a natural estimate given by the sample autocovariance

$$\hat{\gamma}_k = n^{-1}\sum_{i=1}^{n-k} X_i X_{i+k}$$

However, plugging in $\hat{\gamma}_k$ instead of γ_k in Σ_n does not work because

$$\hat{\Sigma}_n = \left[\hat{\gamma}_{|i-j|}\right]_{i,j=1}^{n}$$

is not a consistent estimator of Σ_n in the sense that the operator norm of $\Sigma_n - \hat{\Sigma}_n$ does not converge to zero. To achieve consistency, Wu and Pourahmadi (2009) proposed a banded estimator of the sample covariance matrix.

In the present work, we propose a more general estimator of Σ_n which leaves the $2l+1$ main diagonals of $\hat{\Sigma}_n$ intact, and then gradually down-weighs increasingly distant off-diagonal entries instead of setting them to zero as in the banded matrix case. We establish rates of convergence and demonstrate the efficacy of the proposed method. In addition, by analogy with the related problem of spectral density estimation, we introduce a natural estimate for the banding parameter, l, that is useful even for the Wu and Pourahmadi (2009) estimator.

The remainder of the paper is structured as follows: Section 2 presents our main results; Section 3 addresses a correction to positive definiteness; Section 4 presents a method to choose the banding parameter; Section 5 introduces as an application the linear process bootstrap, a new bootstrap for stationary processes; Section 6 provides a small simulation study; and Section 7 contains all technical proofs.

2 A tapered covariance matrix estimator

In the present section, we establish convergence rates for the tapered sample covariance matrix to Σ_n in the operator norm, defined by

$$\rho(A) = \max_{x \in \mathbb{R}^n : |x|=1} |Ax|, \tag{1}$$

where $|\cdot|$ denotes the usual Euclidean norm on $\mathbb{R}^n$. It is worth noting that $\rho(A) = \sqrt{\lambda_{\max}(A^*A)}$, where $\lambda_{\max}(A^*A)$ is the largest eigenvalue of A^*A, and where A^* denotes the conjugate transpose of A; see Horn and Johnson (1990), p. 296.

We propose estimating Σ_n by the matrix $\hat{\Sigma}_{\kappa,l} := \left[w_{|i-j|}\hat{\gamma}_{|i-j|}\right]_{i,j=1}^n$, where $w_{|i-j|}$ is a weight function which down-weighs the values of $\hat{\gamma}_{|i-j|}$ when $|i-j|$ is large; this is desirable because estimated covariances with large values of $|i-j|$ are known to be less reliable (see, for example, Brockwell and Davis, 1991).

The motivation for our approach lies in the relationship between this problem and that of spectral density estimation. The spectral density is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

and is nonparametrically estimated by

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{j=-l}^{l} w_j \hat{\gamma}_j e^{-i\omega j},$$

where the w_j are weights that play a role analogous to those used in the present problem. In the context of spectral density estimation, the weighting scheme we propose here has shown to provide optimal convergence rates (Politis and Romano, 1995) and to allow for a straightforward method of banding parameter selection (Politis, 2003a); we show that these advantages carry over to the present setting.

With this motivation in mind, we denote our weight function by $\kappa(\cdot)$ and define it as follows.

Definition 1. The tapered weight function κ is given by

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ g(|x|) & \text{if } 1 < |x| \leq c_\kappa \\ 0 & \text{if } |x| > c_\kappa, \end{cases} \tag{2}$$

where $|g(x)| < 1$. The l-scaled version of $\kappa(\cdot)$ will be denoted by

$$\kappa_l(x) := \kappa(x/l).$$

With this notation, our tapered estimator of Σ_n is given by

$$\hat{\Sigma}_{\kappa,l} = \left[\kappa_l(i-j)\hat{\gamma}_{|i-j|}\right]_{i,j=1}^{n}. \tag{3}$$

A simple example of a weight function satisfying Definition 1 is the trapezoid proposed by Politis and Romano (1995), i.e.,

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ 2 - |x| & \text{if } 1 < x \leq 2 \\ 0 & \text{if } |x| > 2 \end{cases} \tag{4}$$

but other choices are possible. For example, McMurry and Politis (2004) consider an infinitely differentiable weight function, and Politis (2007) considers several smooth tapers.

Remark 1. The function $g(x)$ will typically also be decreasing in $|x|$ in such a way that $\kappa(x)$ is continuous; these restrictions do not impact asymptotic convergence rates but they tend to improve finite sample results. The banded estimator of Wu and Pourahmadi (2009) can be put in the framework of our general tapered estimator (3) with the choice $c_\kappa = 1$ and no function g, i.e., a rectangular window $\kappa(x)$. However, the rectangular window does not perform well for spectral estimation, and similarly here the use of a non-rectangular window is recommended.

In order to establish convergence rates of $\hat{\Sigma}_{\kappa,l}$ to Σ_n, we need to impose some short range dependence assumptions on the time series. We follow Wu and Pourahmadi (2009) in adopting the physical dependence measure of Wu (2005). Let ϵ_i, $i \in \mathbb{Z}$ be a sequence of i.i.d. random variables. Moreover, assume that X_i is a causal process of the form

$$X_i = f(\ldots, \epsilon_{i-1}, \epsilon_i),$$

where f is a measurable function such that X_i is well defined and $E\left[X_i^2\right] < \infty$.

In order to quantify the dependence, let ϵ_i' be an independent copy of ϵ_i, $i \in \mathbb{Z}$. Let $\xi_i = (\ldots, \epsilon_{i-1}, \epsilon_i)$, $\xi_i' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_i)$, and $X_i' = g(\xi_i')$. For $\alpha > 0$, we define the physical dependence measure

$$\delta_\alpha(i) := ||X_i - X_i'||_\alpha,$$

where $||Y||_\alpha := E\left[|Y|^\alpha\right]^{1/\alpha}$.

Note that the difference between X_i and X_i' is due only the difference between ϵ_0 and ϵ_0', and therefore $\delta_\alpha(i)$ measures the dependence of X_i on an event i units of time in the past. To measure the cumulative dependence across all time, the quantity

$$\Delta_\alpha := \sum_{i=1}^{\infty} \delta_\alpha(i)$$

is helpful. We will say that $\{X_i\}$ is short-range dependent with moment α if $\Delta_\alpha < \infty$.

Lemma 1. *(Wu and Pourahmadi, 2009) Assume that $\{X_i\}$ satisfies $\Delta_{2q} < \infty$ with $1 < q \leq 2$. Then for any $j \in \mathbb{Z}$,*

$$||\sum_{i=1}^{n} X_i X_{i+|j|} - n\gamma_j||_q \leq 2B_q n^{1/q} ||X_1||_{2q} \Delta_{2q}$$

where

$$B_q = \begin{cases} \frac{18q^{3/2}}{(q-1)^{1/2}} & \text{if } q \neq 2 \\ 1 & \text{if } q = 2. \end{cases}$$

The above lemma is used to establish our main result, which gives an upper bound for the rate of convergence of $\hat{\Sigma}_{\kappa,l}$ to Σ_n.

Theorem 1. *Let $1 < q \leq 2$. Assume $||X_1||_{2q} < \infty$, $\Delta_{2q} < \infty$, and $0 \leq c_\kappa l < n - 1$. Then*

$$||\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)||_q \leq d_q(\lfloor c_\kappa l \rfloor + 1) n^{-(q-1)/q} + \frac{2}{n} \sum_{i=1}^{\lfloor c_\kappa l \rfloor} i|\gamma_i| + 2 \sum_{i=l+1}^{n} |\gamma_j|, \tag{5}$$

where d_q is a constant depending on $||X_1||_{2q}$, Δ_{2q}, and q, and c_κ is as given in (2).

Theorem 1 generalizes Theorem 2 of Wu and Pourahmadi (2009) who consider an estimate Σ_n of the same form as that given in equation (3) but with the weights restricted to being only 1 or 0, i.e., a rectangular window. The additional generality is achieved without changing the overall rate of convergence.

Remark 2. Theorem 1 is stated for mean zero data, but the result applies equally well to the centered data $X_1 - \bar{X}, \ldots, X_n - \bar{X}$.

The inequality (5) suggests approximately optimal rates for l depending on the rate at which $|\gamma_i| \to 0$ as $i \to \infty$.

Corollary 1. *The convergence rate for the bound in inequality (5) can be optimized by minimizing the bound (5) as a function of l. The optimal bounds are found to be:*

i. If $|\gamma_i| = O(i^{-d})$ for some $d > 1$, then the rate for the bound in Theorem 1 is optimized by choosing $l \propto \left(n^{(q-1)/(dq)}\right)$, and the bound (5) becomes of order $O\left(n^{-(d-1)(q-1)/(dq)}\right)$.

ii. If $|\gamma_i| = O(\theta^i)$ for some θ with $|\theta| < 1$ and if $l = \lfloor a \log n \rfloor$ for a large enough, then the bound (5) becomes of order $O\left(n^{-(q-1)/q} \log n\right)$.

iii. If there exists B such that $\gamma_i = 0$ for all $i > B$, then if $l = B$, the bound (5) becomes of order $O\left(n^{(-q-1)/q}\right)$.

In all three cases above, the second term of the bound (5) is dominated by the other two terms.

3 Positive definite autocovariance matrix estimation

Under some additional conditions, Theorem 1 implies that $\hat{\Sigma}_{\kappa,l}$ is asymptotically invertible and provides a bound for the convergence rate of $\hat{\Sigma}_{\kappa,l}^{-1}$ to Σ_n^{-1}.

Theorem 2. *Assume l grows fast enough to ensure the convergence (5) and that $l = o(n^{(q-1)/q})$. Also assume that the spectral density*

$$f(\omega) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k},$$

satisfies $0 < c_1 \leq f(\omega) \leq c_2 < \infty$ for some positive constants c_1 and c_2. Then, under the conditions of Theorem 1, $\hat{\Sigma}_{\kappa,l}$ is positive definite with probability tending to 1, and

$$\rho\left(\hat{\Sigma}_{\kappa,l}^{-1} - \Sigma_n^{-1}\right) = O_p(r_n), \text{ where } r_n = l n^{-(q-1)/q} + \sum_{i=l}^{\infty} |\gamma_j|.$$

However, $\hat{\Sigma}_{\kappa,l}$ is not guaranteed to be positive definite for finite samples. If positive definiteness of $\hat{\Sigma}_{\kappa,l}$ is desired, a modified estimator achieves this goal without compromising accuracy. In particular, consider the spectral decomposition $\hat{\Sigma}_{\kappa,l} = T_n D T_n^t$ where T_n is an orthogonal matrix, and $D = \text{diag}(d_1, \ldots, d_n)$, a diagonal matrix containing the eigenvalues of $\hat{\Sigma}_{\kappa,l}$. Now let

$$\hat{\Sigma}^{\epsilon}_{\kappa,l} := T_n D^{\epsilon} T_n^t,$$

where $D^{\epsilon} = \text{diag}(d_1^{\epsilon}, \ldots, d_n^{\epsilon})$ and $d_i^{\epsilon} = \max(d_i, \epsilon\hat{\gamma}_0/n^{\beta})$; here β and ϵ are user-defined positive constants to be discussed below. The presence of the term $\hat{\gamma}_0$ in the definition of d_i^{ϵ} is in order to make $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ scale-equivariant.

It is obvious that $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ is positive definite by construction. The following is the analog of Theorem 1 for the modified estimator $\hat{\Sigma}^{\epsilon}_{\kappa,l}$.

Theorem 3. *Let $1 < q \leq 2$. Assume $||X_1||_{2q} < \infty$, $\Delta_{2q} < \infty$, and $0 \leq c_\kappa l < n - 1$. Then*

$$\begin{aligned} ||\rho(\hat{\Sigma}^{\epsilon}_{\kappa,l} - \Sigma_n)||_q \quad \leq \quad & 2d_q(\lfloor c_\kappa l \rfloor + 1)n^{-(q-1)/q} + \frac{4}{n}\sum_{i=1}^{\lfloor c_\kappa l \rfloor} i|\gamma_i| + 4\sum_{i=l+1}^{n}|\gamma_j| \\ & + \epsilon\gamma(0)/n^{\beta} + O(n^{1/q-1-\beta}), \end{aligned} \tag{6}$$

where d_q, $||X_1||_{2q}$, Δ_{2q}, and q, and c_κ are as in Theorem 1.

The two last terms on the right hand side of (6) are dominated by the first term when $\beta > 1/2$. The following corollary ensues showing that the modified estimator $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ maintains the same asymptotic rate of convergence as $\hat{\Sigma}_{\kappa,l}$.

Corollary 2. *Assume the conditions of Theorem 3 and that $\beta > 1/2$. Then,*

$$||\rho(\hat{\Sigma}^{\epsilon}_{\kappa,l} - \Sigma_n)||_q = O(||\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)||_q).$$

For practical use, it is advisable not to take β close to the threshold $1/2$. In simulation, we found taking $\beta = 1$ in conjunction with $\epsilon = 1$ worked well. Taking $\epsilon = 0$ will result into an estimator that is non-negative definite but not necessarily positive definite.

An immediate corollary of the two preceding theorems is that the inverse positive definite version of the estimator also achieves the same convergence rates as given in Theorem 2.

Corollary 3. *Under the conditions of Theorems 2 and 3,*

$$\rho\left((\hat{\Sigma}^{\epsilon}_{\kappa,l})^{-1} - \Sigma_n^{-1}\right) = O_p(r_n).$$

4 Banding parameter selection

In this section we recall the rule introduced in Politis (2003a) for estimating the bandwidth in spectral density estimation using flat-top kernels.

Empirical Rule of Picking l. (Politis, 2003a) Let $\varrho(k) = \gamma_k/\gamma_0$ and $\hat{\varrho}(k) = \hat{\gamma}_k/\hat{\gamma}_0$. Let $\hat{l}$ be the smallest positive integer such that $|\hat{\varrho}(\hat{l}+k)| < c\sqrt{\log n/n}$ for $k = 1, \ldots, K_N$ where $c > 0$ is a fixed constant, and K_n is a positive, nondecreasing sequence that satisfies $K_n = o(\log n)$.

The rates of increase of $\hat{l}$ chosen by the above rule vary according to how quickly the autocorrelation function of the process decays; they are summarized in the following theorem.

Theorem 4. *(Politis, 2003a) Assume conditions strong enough to ensure that for all finite m*

$$\max_{i=1,\ldots,m} |\hat{\varrho}(s+i) - \varrho(s+i)| = O_P(1/\sqrt{n})$$

uniformly in s, and

$$\max_{i=0,1,\ldots,n-1} |\hat{\varrho}(i) - \varrho(i)| = O_P\left(\sqrt{\log n/n}\right).$$

Also assume there exists a positive i_0 such that $|\gamma_i| > 0$ for all $i < i_0$.

i. Assume that $\gamma_i = Ci^{-d}$ for $i > i_0$, and for some $C > 0$, and a positive integer d. Then,

$$\hat{l} \overset{P}{\sim} \frac{A_1 n^{1/2d}}{(\log n)^{1/2d}}$$

ii. Assume $\gamma_i = C\theta^i$ for $i > i_0$, where $C > 0$, and $|\theta| < 1$ are some constants. Then

$$\hat{l} \overset{P}{\sim} A_2 \log n$$

where $A_2 = -1/\log|\theta|$.

iii. If $\gamma_i = 0$ for all $k > B \equiv i_0$, but $\gamma_B \neq 0$, then

$$\hat{l} = B + o_p(1).$$

Note $\hat{l}$ automatically adapts to the underlying correlation structure by switching its rate of increase without any decision from the practitioner.

An immediate consequence of Theorem 4 is that, in the case where $q = 2$, the above rule proves close to optimal in the present setting of estimating Σ_n. As a matter of fact, except for the slowly

varying factor $(\log n)^{1/2d}$ in case i, the rates of increase of $\hat{l}$ are the *same* as the optimal rates for $q = 2$ given in Corollary 1. We thus have the following Corollary that gives credence to the applicability of $\hat{l}$ for use in estimating the autocovariance matrix.

Corollary 4. *Assume* $||X_1||_4 < \infty$, $\Delta_4 < \infty$, $0 \leq c_\kappa l < n - 1$, *and let* $\hat{l}$ *be picked by the above empirical rule. Then*

i. If $\gamma_i = Ci^{-d}$ *for* $i > i_0$ *for some* $C > 0$ *and positive integer* d, *then,*

$$||\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)||_2 = O_P\left((n/\log n)^{(-1/2)(1-d^{-1})}\right).$$

ii. If $\gamma_k = C\theta^i$ *for* $i > i_0$ *for some* $C > 0$ *and* $|\theta| < 1$, *then*

$$||\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)||_2 = O_P\left(n^{-1/2}\log n\right).$$

iii. If $\gamma_i = 0$ *for all* $k > B \equiv i_0$, *but* $\gamma_B \neq 0$, *then*

$$||\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)||_2 = O_P\left(n^{-1/2}\right).$$

5 Linear process bootstrap

There are several bootstraps for time series data; see, for example, Lahiri (2003), Politis (2003b), or Bühlmann (2002) for reviews. The most popular methods in the literature are the block bootstrap and the AR sieve. The block bootstrap of Künsch (1989) and Liu and Singh (1992) create bootstrap pseudo-data by resampling from blocks of b consecutive observations. If b, which is assumed to grow with n, is sufficiently large, the pseudo-data will have a dependence structure which closely mimics that of the original process. The AR sieve bootstrap of Kreiss (1992), Paparoditis and Streitberg (1992), and Bühlmann (1997) fits an AR(p) model to the original data, and then uses the fitted model in conjunction with a residual bootstrap to simulate pseudo-data. Letting p grow with n allows the sieve bootstrap to asymptotically capture the covariance structure of the original time series.

A natural extension of the AR sieve would be an MA sieve, which models the observed time series by fitting increasingly high order MA(q) processes to the data; this has not been done because of the relative difficulty of fitting MA models. MA models are either fit by numerical optimization, which is not feasible for large values of q, or by algorithms such as the innovations algorithm presented in Theorem 8.3.1 of Brockwell and Davis (1991). Unfortunately, the innovations algorithm requires

estimating MA coefficients of orders much greater than q in order to assess the stability of the first q fitted parameters; see the discussion following Theorem 8.3.1 in Brockwell and Davis (1991).

Below, we propose a new bootstrap, termed the linear process bootstrap (LPB) which is an alternative to an MA sieve; it works because knowledge of $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ makes it possible to generate an MA process without knowing the MA coefficients. The LPB is also more general because one could, in principle, use a taper $\kappa(\cdot)$ that is not identically zero after a point, but just tends to zero, (see, for example Politis, 2007); in that case the LPB is generating linear MA(∞) rather than MA(q) processes. We prove the validity of the LPB for the mean, and we conjecture its validity for all statistics whose asymptotic distribution depends only on the mean and covariance of the data.

The LPB algorithm is as follows.

1. Let $Y_i = X_i - \bar{X}$ for $i = 1, \ldots, n$, and let $Y = (Y_1, \ldots, Y_n)^t$.
2. Let $W = (\hat{\Sigma}^{\epsilon}_{\kappa,l})^{-1/2} Y$.
3. Let Z be the standardized version of W, with $Z_i = (W_i - \bar{W})/\hat{\sigma}_W$, where $\bar{W} = n^{-1} \sum_{i=1}^{n} W_i$ and $\hat{\sigma}^2_W = n^{-1} \sum_{i=1}^{n} (W_i - \bar{W})^2$.
4. Generate $Z^*_1, \ldots Z^*_n$ by an i.i.d. bootstrap of $Z_1, \ldots, Z_n$.
5. Compute $Y^* = (\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} Z^*$, where $(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2}$ is taken to be the lower triangular matrix L in the Cholesky decomposition $\hat{\Sigma}^{\epsilon}_{\kappa,l} = LL^t$.

Remark 3. The matrix square root $(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{-1/2}$ in step 2 can be any matrix square root that converges to $\Sigma_n^{-1/2}$ at the same rate as $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ converges to Σ_n, such as those obtained by the Cholesky or spectral decompositions (see, for example, Horn and Johnson, 1990, p. 411). We conjecture that the same is true of the square root used in step 5, but our proof of Theorem 5 (below) is specific to the Cholesky decomposition. For reasons of symmetry, it seems preferable to use the same square root in step 2 as in step 5.

Under assumptions of the preceding theorems, the algorithm above can be used to produce confidence intervals for the mean which are justified by the following theorem.

Theorem 5. *Let $E[X_i] = \mu$. Then under the conditions of Theorems 1, 2, and 3, with $q = 2$,*

$$\sup_x \left| \mathrm{P}\left[n^{1/2}(\bar{X} - \mu) \le x \right] - \mathrm{P}^*\left[n^{1/2} \bar{Y}^* \le x \right] \right| \to_P 0, \tag{7}$$

and

$$\mathrm{var}^*\left[n^{1/2} \bar{Y}^* \right] \to_P \sigma^2,$$

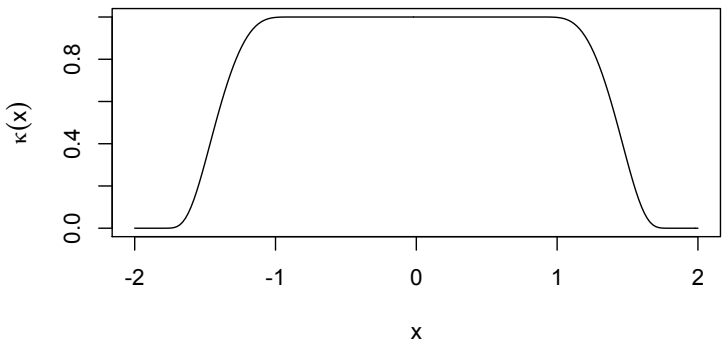

Figure 1: An infinitely differentiable weight function

where $\sigma^2 = (\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k) = \lim_{n\to\infty} \text{var}\left[n^{1/2}\bar{X}\right]$.

Surprisingly, the assumptions of Theorem 5 do not include linearity of the original time series, i.e., an MA(∞) model. Thus our theorem is more general than one would expect.

6 Simulations

6.1 Covariance matrix estimation

We conducted several simulations with the aim of a direct comparison between our estimator and that of Wu and Pourahmadi (2009). They use a subsampling rule to estimate l, whereas we employ the empirical rule of Section 4. We also use three different weight functions: rectangular, defined by $\kappa(x) = 1\{|x| < 1\}$; trapezoidal as defined in equation (4); and infinitely differentiable, as defined in McMurry and Politis (2004) and shown in Figure 1. All simulations were performed in R (R Development Core Team, 2009). Our results show that the trapezoid is consistently the best performer. All results are based on $N = 100$ replications, and the parameters in the bandwidth choice rule were chosen to be $c = 2$, and $K_n = 5$.

We also tested the adjustment to positive definiteness given in Theorem 3 using the trapezoid weight. While negative eigenvalues were occasionally observed, they were so close to zero that, to the two digit resolution given in the following tables, the results were numerically identical to the unadjusted trapezoid estimator. For this reason, these losses are omitted from the table. The rectangular weight function is expected to produce many more nonpositive matrices, as the negative sidelobes of its Fourier transform are more pronounced than those of the trapezoid's.

n	$\hat{l}$	∞-WP	∞-Rect	∞-Trap	∞-InfD	Op-Rect	Op-Trap	Op-InfD
250	1.04 (0.32)	2.7	0.28 (0.21)	0.28 (0.24)	0.28 (0.23)	0.28 (0.21)	0.28 (0.23)	0.28 (0.23)
500	1.07 (0.05)	2.2	0.19 (0.14)	0.19 (0.15)	0.19 (0.15)	0.18 (0.14)	0.19 (0.14)	0.19 (0.14)
750	1.07 (0.54)	1.9	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)	0.16 (0.10)

Table 1: Banding parameter and losses in the matrix infinity norm and operator norm for the moving average processes. Losses are calculated for the trapezoid, rectangular, and infinitely differentiable weight functions. The column ∞-WP contains the losses reported in Wu and Pourahmadi (Wu and Pourahmadi, 2009). Standard deviations are shown in parentheses.

6.1.1 MA(1)

In the first simulation, the data was generated by the moving average process $X_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\theta = 0.5$, and ϵ_t and iid sequence of $N(0,1)$ random variables. Our results are provided in Table 1. Wu and Pourahmadi (2009) estimate losses in the matrix infinity norm

$$|||A|||_\infty := \max_{i\in\{1,\ldots,n\}} \sum_{j=1}^{n} |a_{ij}|,$$

so this is included along with operator norm losses. The only difference between our approach with rectangular kernels and their approach is in the selection of $\hat{l}$. In infinity norm, our methodology reduces the loss by more than a factor of 10 for all sample sizes, and we are close to achieving the theoretically best possible results presented in Table 1 of Wu and Pourahmadi (2009).

6.1.2 AR(1)

In the second experiment data was simulated from the AR(1) process $X_t = \phi X_{t-1} + \epsilon_t$, where the ϵ_t were iid $N(0, 1-\phi^2)$, for $\phi = 0.1, 0.5,$ and 0.9; the error variance was chosen to make var $[X_t] = 1$ for all simulations. Results are presented in Table 2. Wu and Pourahmadi (2009) do not provide numeric results for this case.

6.1.3 Absolute value AR(1)

For the final simulation, data were simulated from the model $X_t = \phi|X_{t-1}| + \epsilon_t$, where ϵ_t were iid $N(0,1)$. The autocovariance function for the absolute AR model does not have a simple closed form, so we approximate the true Σ_n with its empirical version using a very large amount of data. While

n	ϕ	$\hat{l}$	∞-Rect	∞-Trap	∞-InfD	Op-Rect	Op-Trap	Op-InfD
250	0.1	0.17 (0.77)	0.29 (0.07)	0.29 (0.07)	0.29 (0.07)	0.27 (0.07)	0.27 (0.07)	0.27 (0.07)
250	0.5	2.07 (1.17)	0.91 (0.31)	0.84 (0.35)	0.87 (0.34)	0.77 (0.35)	0.73 (0.38)	0.74 (0.37)
250	0.9	12.95 (6.52)	10.16 (4.25)	9.55 (4.64)	9.56 (4.65)	9.18 (4.45)	8.81 (4.78)	8.82 (4.79)
500	0.1	0.13 (0.34)	0.27 (0.05)	0.27 (0.05)	0.27 (0.05)	0.25 (0.05)	0.25 (0.05)	0.25 (0.05)
500	0.5	2.54 (1.27)	0.79 (0.27)	0.71 (0.34)	0.74 (0.32)	0.65 (0.29)	0.61 (0.35)	0.63 (0.34)
500	0.9	17.61 (9.63)	8.89 (3.92)	8.48 (4.51)	8.48 (4.49)	7.89 (4.29)	7.77 (4.76)	7.76 (4.74)
750	0.1	0.20 (0.40)	0.24 (0.05)	0.24 (0.05)	0.24 (0.05)	0.22 (0.06)	0.22 (0.06)	0.22 (0.06)
750	0.5	2.69 (1.03)	0.67 (0.20)	0.60 (0.25)	0.62 (0.25)	0.56 (0.24)	0.52 (0.28)	0.54 (0.28)
750	0.9	19.72 (8.99)	7.37 (3.71)	6.98 (4.34)	6.98 (4.29)	6.46 (3.78)	6.33 (4.43)	6.32 (4.38)

Table 2: Banding parameter and losses in the matrix infinity norm and operator norm for the auto-regressive processes. Losses are calculated for the trapezoid, rectangular, and infinitely differentiable weight functions. Standard deviations are shown in parentheses.

this provides (crude) estimates of the loss, it induces some additional, and difficult to quantify, uncertainty into the values shown in Table 3. Nonetheless, our results again show significant improvement over those presented in Wu and Pourahmadi (2009), particularly for smaller values of ϕ.

6.2 Linear process bootstrap

Finally, we ran several simulation experiments to assess the performance of the linear process bootstrap and for comparison we also tested each simulated data set using two other popular resampling schemes. First, we considered the block bootstrap of Künsch (1989) and Liu and Singh (1992), as implemented in Canty and Ripley (2009), using the block length selection of Politis and White (2004) (see also Patton, Politis, and White, 2009). We also considered the sieve bootstrap of Bühlmann (1997). Each experiment was repeated 1000 times using 1000 bootstrap replications. The results are shown in Table 4.

For the absolute AR and AR models, the LPB's performance was comparable to that of the block bootstrap, and slightly less efficient than the sieve bootstrap; this is unsurprising as the sieve approximates the time series with a best fit AR model, while the LPB essentially approximates the time series with an MA model. In the case of an AR(1) series, fitting an AR model should be close to optimal, and in the case of the absolute AR model, it seems reasonable that an AR time series

n	ϕ	$\hat{l}$	∞-WP	∞-Rect	∞-Trap	∞-InfD	Op-Rect	Op-Trap	Op-InfD
250	0.1	0.09 (0.53)	1.7	0.17 (0.05)	0.17 (0.06)	0.17 (0.06)	0.11 (0.05)	0.11 (0.06)	0.11 (0.06)
250	0.5	0.44 (0.72)	2.2	0.59 (0.15)	0.59 (0.16)	0.59 (0.16)	0.47 (0.17)	0.48 (0.18)	0.47 (0.17)
250	0.9	6.50 (4.15)	14.3	11.39 (4.77)	11.02 (6.10)	11.09 (6.02)	9.87 (4.82)	9.92 (6.07)	9.96 (6.00)
500	0.1	0.00 (0.00)	1.3	0.20 (0.04)	0.20 (0.04)	0.20 (0.04)	0.09 (0.04)	0.09 (0.04)	0.09 (0.04)
500	0.5	0.76 (0.75)	1.8	0.52 (0.17)	0.52 (0.17)	0.52 (0.17)	0.34 (0.17)	0.34 (0.17)	0.34 (0.17)
500	0.9	7.70 (4.59)	13.4	10.04 (5.87)	9.47 (6.69)	9.53 (6.62)	8.39 (6.20)	8.10 (6.91)	8.12 (6.86)
750	0.1	0.00 (0.00)	1.1	0.24 (0.03)	0.24 (0.03)	0.24 (0.03)	0.07 (0.03)	0.07 (0.03)	0.07 (0.03)
750	0.5	0.85 (0.36)	1.5	0.51 (0.13)	0.51 (0.13)	0.51 (0.13)	0.27 (0.14)	0.27 (0.14)	0.27 (0.14)
750	0.9	9.01 (3.85)	14.6	9.31 (4.37)	8.74 (5.08)	8.82 (5.00)	7.33 (4.60)	7.28 (5.26)	7.30 (5.18)

Table 3: Banding parameter and approximate losses in the matrix infinity norm and operator norm for the absolute value auto-regressive processes. Losses are calculated for the trapezoid, rectangular, and infinitely differentiable weight functions. The column ∞-WP contains the estimated losses reported in Wu and Pourahmadi (2009). Standard deviations are shown in parentheses.

might provide a better fit. The results for the absolute AR model are only approximate, as the true mean for the model was estimated using a very large simulated data set. Perhaps the most surprising feature of the simulations was the sieve bootstrap's relatively poor performance when the AR coefficient was 0.9. Although the AR sieve bootstrap is expected to break down when ϕ is close to 1, the bad behavior for $\phi = 0.9$ was unexpected.

For the MA model, the LPB was slightly better than the other two when the MA coefficient was large. A more comprehensive comparison of the LPB with the block bootstrap and AR sieve would include the case of studentized sample mean and is part of future work on the subject.

7 Technical proofs

Proof of Theorem 1: By Problem 21, p. 313 in Horn and Johnson (1990), and since $\hat{\Sigma}_{\kappa,l} - \Sigma_n$ is symmetric,

$$\begin{aligned}
\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n) &\leq \max_{1\leq j\leq n} \sum_{i=1}^{n} |\hat{\gamma}_{i-j}\kappa_l(|i-j|) - \gamma_{i-j}| \\
&\leq \sum_{i=1-n}^{n-1} |\hat{\gamma}_i \kappa_l(i) - \gamma_i| \\
&\leq 2\sum_{i=0}^{l} |\hat{\gamma}_i - \gamma_i| + 2\sum_{i=l+1}^{\lfloor c_\kappa l \rfloor} |\hat{\gamma}_i\kappa_l(i) - \gamma_i| + 2\sum_{i=\lfloor c_\kappa l\rfloor+1}^{n} |\gamma_i| \\
&= T_1 + T_2 + T_3.
\end{aligned}$$

We first examine T_1. By Lemma 1, there exists a constant d'_q depending on $||X_1||_{2q}$ and Δ_{2q}, but not l or n, such that

$$\begin{aligned}
||\hat{\gamma}_i - \gamma_i||_q &\leq ||\hat{\gamma}_i - E\left[\hat{\gamma}_i\right]||_q + \frac{i}{n}|\gamma_i| \\
&\leq \frac{d'_q(n-i)^{1/q}}{n} + \frac{i|\gamma_i|}{n}.
\end{aligned}$$

Therefore

$$||T_1||_q \leq d_q(l+1)n^{-(q-1)/q} + \frac{2}{n}\sum_{i=1}^{l} i|\gamma_i|,$$

where $d_q = d'_q/2$.

The second term, T_2, follows in a similar fashion.

$$T_2 \leq 2\sum_{i=l+1}^{\lfloor c_\kappa l\rfloor} \kappa_l(i)|\hat{\gamma}_i - \gamma_i| + 2\sum_{i=l+1}^{\lfloor c_\kappa l\rfloor} |\kappa_l(i)-1||\gamma_i|.$$

Therefore,

$$||T_2||_q \leq d_q(\lfloor c_\kappa l\rfloor - l)n^{-(q-1)/q} + \frac{2}{n}\sum_{i=l+1}^{\lfloor c_\kappa l\rfloor} i|\gamma_i| + 2\sum_{i=l+1}^{\lfloor c_\kappa l\rfloor} |\gamma_i|.$$

□

Proof of Theorem 2: All the eigenvalues of Σ_n lie in the interval $[2\pi c_1, 2\pi c_2]$ (see Grenander and Szegö, 1958, Section 5.2). By Theorem 1, $\rho\left(\hat{\Sigma}_{\kappa,l} - \Sigma_n\right) = O_p(r_n)$. Since r_n tends to zero, the probability $\hat{\Sigma}_{\kappa,l}$ is positive definite tends to 1.

Let $A_n = \Sigma_n^{-1/2}$ and $\Gamma_n = A_n \hat{\Sigma}_{\kappa,l} A_n$. Then

$$\begin{aligned} \rho\left(\Gamma_n - I_n\right) &\leq \rho(A_n)^2 \, \rho\left(\hat{\Sigma}_{\kappa,l} - \Sigma_n\right) \\ &= O_p(r_n). \end{aligned}$$

Similarly,

$$\begin{aligned} \rho\left(\Gamma_n^{-1} - I_n\right) &\leq \rho(\Gamma_n^{-1}) \, \rho\left(\Gamma_n - I_n\right) \\ &= O_p(r_n). \end{aligned}$$

Since $\hat{\Sigma}_{\kappa,l}^{-1} - \Sigma_n^{-1} = A_n(\Gamma_n^{-1} - I_n)A_n$, the result follows. □

Proof of Theorem 3: By the triangle inequality,

$$\rho(\hat{\Sigma}_{\kappa,l}^{\epsilon} - \Sigma_n) \leq \rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n) + \rho(\hat{\Sigma}_{\kappa,l}^{\epsilon} - \hat{\Sigma}_{\kappa,l}). \tag{8}$$

Recall that $\hat{\Sigma}_{\kappa,l} = T_n D T_n^t$, where without loss of generality, we assume that the eigenvalues of $\hat{\Sigma}_{\kappa,l}$ have been ordered so that $D = \text{diag}(d_1, \ldots, d_n)$, where $d_1 \geq d_2 \geq \ldots \geq d_n$. Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ respectively denote the largest and smallest eigenvalues of a symmetric matrix A. Then,

$$\begin{aligned} d_n &= \lambda_{\min}(\hat{\Sigma}_{\kappa,l}) \\ &= -\lambda_{\max}(-\hat{\Sigma}_{\kappa,l}) \\ &\geq -\lambda_{\max}(\Sigma_n - \hat{\Sigma}_{\kappa,l}) \\ &\geq -\rho(\Sigma_n - \hat{\Sigma}_{\kappa,l}), \end{aligned} \tag{9}$$

where the first inequality follows because Σ_n is non-negative definite (see Corollary 4.3.3, p.182 in Horn and Johnson, 1990).

We now focus on the second term of (8).

$$\hat{\Sigma}_{\kappa,l}^{\epsilon} - \hat{\Sigma}_{\kappa,l} = T_n D^{-} T_n^t$$

where $D^{-} = \text{diag}\left(\max(d_1, \epsilon\hat{\gamma}_0/n^{\beta}) - d_1, \ldots, \max(d_n, \epsilon\hat{\gamma}_0/n^{\beta}) - d_n\right)$. By the above spectral decomposition and inequality (9),

$$\begin{aligned} \rho(\hat{\Sigma}_{\kappa,l}^{\epsilon} - \hat{\Sigma}_{\kappa,l}) &= \max\left(0, \epsilon\hat{\gamma}_0/n^{\beta} - d_n\right) \\ &\leq \max\left(0, \epsilon\hat{\gamma}_0/n^{\beta} + \rho(\Sigma_n - \hat{\Sigma}_{\kappa,l})\right) \\ &\leq \epsilon\hat{\gamma}_0/n^{\beta} + \rho(\Sigma_n - \hat{\Sigma}_{\kappa,l}). \end{aligned}$$

The result now follows from Theorem 1 and Lemma 1. □

Proof of Theorem 5: By Theorem 3 in Wu (2005),

$$n^{1/2}(\bar{X} - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

We establish (7) by showing $n^{1/2}\bar{Y}^*$ has the same limiting normal distribution. For clarity of exposition, the proof proceeds through a sequence of lemmas.

Lemma 2. *Define $\tilde{Z}^*$ to be the equivalent bootstrap resample to Z^*, except the resample is drawn from the standardized values of $\Sigma_n^{-1/2}Y$ rather than its data driven counterpart $(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{-1/2}Y$. Let $\mathbf{1}$ be the n-vector of 1's. Under the conditions of Theorem 5,*

$$\begin{aligned} n^{1/2}\bar{Y}^* &= n^{-1/2}\mathbf{1}^t(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2}Z^* \\ &= n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^* + n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})(Z^* - \tilde{Z}^*) + n^{-1/2}\mathbf{1}^t\left[(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} - \Sigma_n^{1/2}\right]Z^* \\ &= n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^* + R_1 + R_2 \qquad (10) \\ &= n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^* + o_P(1). \end{aligned}$$

Proof of Lemma 2. We first consider R_2. It has bootstrap mean 0 and variance

$$\begin{aligned} &\mathrm{var}^*\left[n^{-1/2}\mathbf{1}^t\left((\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} - \Sigma_n^{1/2}\right)Z^*\right] \\ &\qquad = E^*\left[n^{-1}\mathbf{1}^t\left[(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} - \Sigma_n^{1/2}\right]Z^*(Z^*)^t\left[(\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} - \Sigma_n^{1/2}\right]\mathbf{1}\right] \\ &\qquad = o_P(1), \end{aligned}$$

where the final equality follows because $E^*\left[Z^*(Z^*)^t\right] = I$ and $\rho\left((\hat{\Sigma}^{\epsilon}_{\kappa,l})^{1/2} - \Sigma_n^{1/2}\right) \to_P 0$.

For R_1, we can write, $Z^* = \hat{\sigma}_W^{-1}M^*(I - n^{-1}\mathbf{1}_n)\hat{\Sigma}^{-1/2}_{\kappa,l}Y$ where $\mathbf{1}_n$ is the $n \times n$ matrix of ones, and M^* is a random $n \times n$ matrix, where each row is independently and uniformly selected from the standard basis vectors $e_1, \ldots, e_n$. With this notation, $\tilde{Z}^* = \hat{\sigma}_{\tilde{W}}^{-1}M^*(I - n^{-1}\mathbf{1}_n)\Sigma_n^{-1/2}Y$. Since $|\hat{\sigma}^2_W - \hat{\sigma}^2_{\tilde{W}}| = o_P(1)$, and both $\hat{\sigma}^2_W$ and $\hat{\sigma}^2_{\tilde{W}}$ are bounded away from 0 and from above with probability tending to 1,

$$\begin{aligned} n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})(Z^* - \tilde{Z}^*) &= n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})M^*(I - n^{-1}\mathbf{1}_n)\left[\hat{\sigma}_W^{-1}\hat{\Sigma}^{-1/2}_{\kappa,l} - \hat{\sigma}_{\tilde{W}}^{-1}\Sigma_n^{-1/2}\right]Y \\ &= n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})M^*(I - n^{-1}\mathbf{1}_n)\left[\hat{\sigma}_{\tilde{W}}^{-1}\hat{\Sigma}^{-1/2}_{\kappa,l} - \hat{\sigma}_{\tilde{W}}^{-1}\Sigma_n^{-1/2}\right]Y + o_P(1) \\ &= R_3 + o_P(1). \end{aligned}$$

It is clear by construction that $E^*[R_3] = 0$. Its bootstrap variance is

$$\begin{aligned}
&\mathrm{var}^*[R_3] \\
&= \hat{\sigma}_{\tilde{W}}^{-2} E^*\left[n^{-1}\mathbf{1}^t\Sigma_n^{1/2}M^*(I-n^{-1}\mathbf{1}_n)(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})YY^t(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})(I-n^{-1}\mathbf{1}_n)(M^*)^t\Sigma_n^{1/2}\mathbf{1}\right] \\
&= \hat{\sigma}_{\tilde{W}}^{-2} E^*\left[n^{-1}\mathbf{1}^t\Sigma_n^{1/2}V^*(V^*)^t\Sigma_n^{1/2}\mathbf{1}\right]
\end{aligned}$$

where V^* is an n-vector of bootstrap resamples of the elements of $(I-n^{-1}\mathbf{1}_n)(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})Y$. Since the sample is i.i.d., $E^*\left[V^*(V^*)^t\right] = \sigma_V^2 I$, where

$$\begin{aligned}
\sigma_V^2 &= n^{-1}Y^t(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})(I-n^{-1}\mathbf{1}_n)^2(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})Y \\
&= n^{-1}Y^t(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2})Y(1+O_P(1)).
\end{aligned}$$

Since $Y^tY = O_P(n)$ (see Wu, 2005) and $\rho(\hat{\Sigma}_{\kappa,l}^{-1/2}-\Sigma_n^{-1/2}) = O_P(r_n)$, we have $\sigma_V^2 = o_P(1)$ which implies $\mathrm{var}^*[R_3] = o_P(1)$ and therefore $R_3 = o_P(1)$. □

Lemma 3.

$$\mathrm{var}^*\left[n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^*\right] = \left(n^{-1}\sum_{i=-(n-1)}^{n-1}(n-|i|)\gamma_i\right)$$

Proof of Lemma 3.

$$\begin{aligned}
\mathrm{var}^*\left[n^{-1/2}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^*\right] &= E^*\left[n^{-1}\mathbf{1}^t(\Sigma_n^{1/2})\tilde{Z}^*\tilde{Z}^{*t}\Sigma_n^{-1/2}\mathbf{1}\right] \\
&= \left(n^{-1}\sum_{i=-(n-1)}^{n-1}(n-|i|)\gamma_i\right),
\end{aligned}$$

where the final equality follows because $E^*\left[\tilde{Z}^*\tilde{Z}^{*t}\right] = I$. □

Lemma 4. *Let A_n and B_n be sequences of $n \times n$ symmetric matrices bounded in operator norm and satisfying $\rho(A_n - B_n) \to 0$. Then $n^{-1/2}\mathbf{1}^tA_n\tilde{Z}^* = n^{-1/2}\mathbf{1}^tB_n\tilde{Z}^* + o_P(1)$.*

Proof of Lemma 4. Since $\tilde{Z}^*$ has bootstrap mean 0, it is sufficient to show the variance converges to 0 in probability.

$$\begin{aligned}
\mathrm{var}^*\left[n^{-1/2}\mathbf{1}^t(A_n-B_n)\tilde{Z}^*\right] &= \sigma_{\tilde{Z}^*}^2 n^{-1}\mathbf{1}^t(A_n-B_n)^2\mathbf{1} \\
&\leq \rho(A_n-B_n)(1+o_p(1)) \\
&= o_p(1)
\end{aligned}$$

□

Lemma 5. *Under the assumptions of Theorem 5,* $E\left[\tilde{Z}_i^4\right]$ *uniformly bounded in* i*.*

Proof of Lemma 5. Denote the entries of $\Sigma_n^{-1/2} = [a_{ij}]_{i,j=1}^n$. With this notation, $\tilde{Z}_i = \sum_{j=1}^n a_{ij} Y_j$. Following Theorem 2 in Wu (2005), define the projection operator $P_k Y = E[Y|\xi_k] - E[Y|\xi_{k-1}]$, and define $M_{k,n} = \sum_{j=1}^n a_{ij} P_{j-k} Y_j$. Then, $\tilde{Z}_i = \sum_{k=0}^\infty M_{k,n}$. By Proposition 4 in Dedecker and Doukhan (2003),

$$||M_{k,n}||_4 \leq ||P_0 Y_k||_4 \left(8 \sum_{j=1}^n a_{ij}^2\right)^{1/2}$$

Since $\sum_{j=1}^n a_{ij}^2 = e_i^t \Sigma_n^{-1} e_i$, where e_i is the i'th standard basis vector, for large enough n, $\sum_{j=1}^n a_{ij}^2$ is bounded from above and away from 0 uniformly in i. By Theorem 1 in Wu (2005), $\sum_{k=0}^\infty ||P_0 Y_k||_4 < \infty$. Therefore, $||\tilde{Z}_i||_4$ is uniformly bounded in i. □

Lemma 6. *(Horn and Johnson, 1990, p. 411) Let* A *and* B *be symmetric and positive definite. Let* $A^{1/2}$ *and* $B^{1/2}$ *denote the lower triangular square roots obtained through the Cholesky decomposition of* A *and* B*. Then* $\rho(A^{1/2} - B^{1/2}) \leq \rho(A^{-1/2})\rho(A - B)$*.*

We are now in a position to complete the proof of Theorem 5. We do so by approximating $\Sigma_n^{1/2}$ in (10) in the following manner. Let $\Sigma_{n,k} = \left[\gamma_{|i-j|} 1_{|i-j| \leq k}\right]_{i,j=1}^n$ be the k-banded version of Σ_n. By Horn and Johnson (1990) p. 313,

$$\rho(\Sigma_{n,k} - \Sigma_n) \leq 2 \sum_{i=k+1}^\infty |\gamma_i|.$$

Therefore $\rho(\Sigma_{n,k} - \Sigma_n) \to 0$ for any sequence $k \to \infty$. Let $L_{n,k}$ and $\Sigma_n^{1/2}$ be the lower-triangular matrices associated with the Cholesky decompositions of $\Sigma_{n,k}$ and Σ_n respectively. By Lemma 6, $\rho(L_{n,k} - \Sigma_n^{1/2}) \to 0$, so by Lemma 4 we can approximate $\Sigma_n^{1/2}$ in (10) by $L_{n,k}$.

Matrix multiplication shows that $L_{n,k}$ is nonzero only on the main diagonal and the first k diagonals below the main, and that the entries of $L_{n,k}$ are bounded in absolute value by $\gamma_0^{1/2}$. Letting $c_{1,n}, \ldots, c_{n,n}$ denote the column sums of $L_{n,k}$, we immediately see $\sum_{i=1}^n c_{i,n}^4 = O(k^4 n) = O(n(\log n)^4)$, if we choose $k \propto \log n$. We can now establish the main result.

In order to proceed with the proof of Theorem 5, we use $L_{n,k}$ to approximate $\Sigma_n^{1/2}$ in the first term of (10).

$$n^{-1/2} \mathbf{1}^t (L_{n,k}) \tilde{Z}^* = n^{-1/2} (c_{1,n} \tilde{Z}_1^* + \ldots + c_{n,n} \tilde{Z}_n^*).$$

We establish the desired result via the central limit theorem for triangular arrays (Resnick, 1999, p. 321) which is implied by the Liapunov condition

$$\frac{1}{(c_{1,n}^2 + \ldots + c_{n,n}^2)^{1+\delta/2}} \sum_{i=1}^{n} E^* \left[|c_{i,n} \tilde{Z}_i^*|^{2+\delta} \right] \to 0, \tag{11}$$

for some $\delta > 0$. Convergence (11) will be shown to hold for $\delta = 2$. We first examine the numerator.

$$\sum_{i=1}^{n} E^* \left[|c_{i,n} \tilde{Z}_i^*|^{2+\delta} \right] = \left(\sum_{i=1}^{n} c_{i,n}^4 \right) \left(n^{-1} \sum_{i=1}^{n} Z_i^4 \right) = O_P(k^4 n)$$

by Lemma 5 and the preceding calculation. We now turn our attention to the denominator of (11).

$$\begin{aligned} c_{1,n}^2 + \ldots + c_{n,n}^2 &= \mathbf{1}^t L_{n,k} L_{n,k}' \mathbf{1} \\ &= \mathbf{1}^t \Sigma_{n,k} \mathbf{1}. \end{aligned}$$

By Problem 21, p. 313 in Horn and Johnson (1990), $\rho(\Sigma_{n,k} - \Sigma_n) \leq 2 \sum_{i=k+1}^{n} |\gamma_i|$. Therefore $\rho(\Sigma_{n,k} - \Sigma_n) \to 0$ for any sequence $k \to \infty$. Since the eigenvalues of Σ_n lie in the interval $[2\pi c_1, 2\pi c_2]$ for c_1 and c_2 as in Theorem 2, for any $\epsilon > 0$ there exists k large enough such that $(2\pi c_1 - \epsilon)n < \mathbf{1}^t \Sigma_{n,k} \mathbf{1} < (2\pi c_2 + \epsilon)n$. This establishes that the denominator of (11) is $O(n^2)$ when $\delta = 2$.

The only requirement on k is that $k \to \infty$, so we now choose $k = \log n$, and it immediately follows that the left side of (11) converges to 0 in probability. With Lemma 3, this completes the proof. □

References

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer, New York, 1991.

P. Bühlmann. Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148, 1997.

P. Bühlmann. Bootstraps for time series. *Statist. Sci.*, 17(1):52–72, 2002.

A. Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2009. R package version 1.2-41.

J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic Process. Appl*, 106(1):63–80, 2003.

U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications.* University of California Press, Los Angeles, 1958.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1990.

J.-P. Kreiss. Bootstrap procedures for AR(∞)-processes. In *Bootstrapping and related techniques (Trier, 1990)*, volume 376 of *Lecture Notes in Econom. and Math. Systems*, pages 107–113. Springer, Berlin, 1992.

H. R. Künsch. The jackknife and the bootstrap for general stationary observations. *Ann. Statist*, 17:1217–1241, 1989.

S. N. Lahiri. *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer-Verlag, New York, 2003.

R. Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard, editors, *Exploring the Limits of Bootstrap*, pages 225–248. Wiley, New York, 1992.

T. McMurry and D. N. Politis. Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics*, 16(3-4):549–562, 2004.

E. Paparoditis and B. Streitberg. Order identification statistics in stationary autoregressive moving-average models: vector autocorrelations and the bootstrap. *J. Time Ser. Anal.*, 13(5):415–434, 1992.

A. Patton, D. N. Politis, and H. White. CORRECTION TO "Automatic block-length selection for the dependent bootstrap". *Econometric Reviews*, 28(4):372–375, 2009.

D. N. Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 25(4-5):517–533, 2003a.

D. N. Politis. The impact of bootstrap methods on time series analysis. *Statist. Sci.*, 18(2):219–230, 2003b.

D. N. Politis. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *UCSD Deptartment of Economics, Discussion Paper 2005-03*, 2007.

D. N. Politis and J. P. Romano. Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis*, 16(1):67–103, 1995.

D. N. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

S. I. Resnick. *A Probability Path*. Birkhauser, Boston, 1999.

W. B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102:14150–14154, 2005.

W. B. Wu and M. Pourahmadi. Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 19(4):1755–1768, 2009.

			Coverage			Avg. Length		
Model	n	ϕ	LPB	Blk	Sieve	LPB	Blk	Sieve
AbsAR(1)	250	0.1	0.95	0.95	0.94	0.25	0.25	0.25
AbsAR(1)	250	0.5	0.90	0.92	0.93	0.28	0.30	0.31
AbsAR(1)	250	0.9	0.89	0.85	0.88	0.99	0.90	1.01
AbsAR(1)	500	0.1	0.93	0.93	0.93	0.18	0.18	0.18
AbsAR(1)	500	0.5	0.93	0.94	0.94	0.21	0.21	0.22
AbsAR(1)	500	0.9	0.91	0.88	0.91	0.75	0.69	0.76
AbsAR(1)	750	0.1	0.93	0.93	0.93	0.14	0.14	0.14
AbsAR(1)	750	0.5	0.94	0.95	0.95	0.17	0.18	0.18
AbsAR(1)	750	0.9	0.92	0.89	0.92	0.62	0.57	0.62
AR(1)	250	0.1	0.93	0.94	0.95	0.25	0.26	0.27
AR(1)	250	0.5	0.93	0.92	0.95	0.39	0.39	0.42
AR(1)	250	0.9	0.88	0.82	0.89	0.90	0.77	0.96
AR(1)	500	0.1	0.92	0.92	0.94	0.18	0.19	0.19
AR(1)	500	0.5	0.93	0.92	0.94	0.29	0.28	0.30
AR(1)	500	0.9	0.92	0.89	0.94	0.70	0.61	0.72
AR(1)	750	0.1	0.94	0.95	0.96	0.15	0.15	0.16
AR(1)	750	0.5	0.94	0.94	0.95	0.24	0.23	0.25
AR(1)	750	0.9	0.91	0.88	0.92	0.58	0.52	0.59
MA(1)	250	0.1	0.92	0.93	0.94	0.25	0.26	0.27
MA(1)	250	0.5	0.95	0.94	0.95	0.37	0.35	0.36
MA(1)	250	0.9	0.94	0.92	0.91	0.46	0.44	0.44
MA(1)	500	0.1	0.92	0.93	0.94	0.18	0.19	0.19
MA(1)	500	0.5	0.95	0.93	0.94	0.26	0.25	0.26
MA(1)	500	0.9	0.95	0.94	0.93	0.33	0.32	0.32
MA(1)	750	0.1	0.94	0.95	0.95	0.15	0.15	0.16
MA(1)	750	0.5	0.95	0.94	0.93	0.21	0.21	0.21
MA(1)	750	0.9	0.94	0.93	0.92	0.27	0.26	0.26

Table 4: Simulation of bootstrap confidence intervals for the mean at nominal 95% coverage. Coverage is shown in the first 3 columns and average interval length is shown in the rightmost 3 columns.