Original Article

Journal of Time Series Analysis

First version received 31 March 2010

Published online in Wiley Online Library: 22 August 2010

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2010.00679.x

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

Timothy L. McMurry[a] and Dimitris N. Politis[b,*,†]

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.

Keywords: 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 this article is to estimate the $n \times n$ autocovariance matrix

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

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

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

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 this article, we propose a more general estimator of Σ_n that 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 this article 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 (LPB), a new bootstrap for stationary processes; Section 6 provides a small simulation study; and Section 7 contains all technical proofs.

[a]*DePaul University*
[b]*University of California*
**Correspondence to: Dimitris N. Politis, Department of Mathematics, University of California, San Diego, USA.*
[†]*E-mail: dpolitis@ucsd.edu*

2. A TAPERED COVARIANCE MATRIX ESTIMATOR

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 for a real-valued matrix A, $\rho(A) = \sqrt{\lambda_{\max}(A^{\mathsf{T}}A)}$, where $\lambda_{\max}(M)$ is the largest eigenvalue of the matrix M; see Horn and Johnson (1990, p. 296).

We propose estimating Σ_n by the matrix $\hat{\Sigma}_{\kappa,l} := [w_{|i-j|}\hat{\gamma}_{|i-j|}]_{i,j=1}^n$, where $w_{|i-j|}$ is a weight function that 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 non-parametrically 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*

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

where $g(\cdot)$ is a function satisfying $|g(x)| < 1$, and c_κ a constant satisfying $c_\kappa \ge 1$. 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

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

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

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

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 iid 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 $E[X_i^2] < \infty$.

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

Note that the difference between X_i and X_i' is due only to 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 $\Delta_\alpha < \infty$.

This notion of short-range dependence provides conditions strong enough 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 eqn (2).

Theorem 1 generalizes Thm 2 of Wu and Pourahmadi (2009) who consider an estimate Σ_n of the same form as that given in eqn (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 centred 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 $l \propto (n^{(q-1)/(dq)})$, and the bound (5) becomes of order $O(n^{-(d-1)(q-1)/(dq)})$.*
- [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 $O(n^{-(q-1)/q} \log n)$.*
- [iii] *If there exists B such that $\gamma_i = 0$ for all $i > B$, then if $l = B$, the bound* (5) *becomes of order $O(n^{(-q-1)/q})$.*

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

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

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 $\hat{\Sigma}_{\kappa,l} = T_n D T_n^{\mathrm{T}}$ where T_n is an orthogonal matrix, and $D = \mathrm{diag}(d_1,\ldots,d_n)$, a diagonal matrix containing the eigenvalues of $\hat{\Sigma}_{\kappa,l}$. Now let

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

where $D^{\epsilon} = \mathrm{diag}(d_1^{\epsilon}, \ldots, d_n^{\epsilon})$ 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} makes $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ scale equivariant.

It is obvious that $\hat{\Sigma}^{\epsilon}_{\kappa,l}$ is positive definite by construction. The following is the analogue 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*

$$\|\rho(\hat{\Sigma}^{\epsilon}_{\kappa,l} - \Sigma_n)\|_q \leq 2d_q(\lfloor c_\kappa l \rfloor + 1)n^{-(q-1)/q} + \frac{4}{n}\sum_{i=1}^{\lfloor c_\kappa l \rfloor} i|\gamma_i| + 4\sum_{i=l+1}^{n}|\gamma_j| + \epsilon\gamma_0/n^{\beta} + O(n^{1/q-1-\beta}), \tag{6}$$

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 eqn (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 $\beta = 1$ 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

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

4.1. Empirical rule for 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, non-decreasing sequence that satisfies $K_n = o(\log n)$.

The rates of increase in $\hat{l}$ 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*

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

uniformly in s, and

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

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

[i] *Assume that* $\gamma_i \sim Ci^{-d}$, *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* $\gamma_i \sim C\theta^i$, *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.

REMARK 3. Theorem 4 remains valid for all $c > 0$ and $1 \leq K_n \leq n$, although different choices of c and K_n lead to very different finite sample performances. Nonetheless, there are some guidelines for practically useful choices of these tuning parameters. The factor $\sqrt{\log n}$ varies slowly, so it has little influence. For example, if log is taken to denote base 10 logarithm, then for sample sizes between 100 and 1000, as is quite typical, $\sqrt{\log n}$ varies between 1.41 and 1.73. Thus, if c is chosen to be around 2 and K_n about 5, the event $\varrho(\hat{l}+k) < c\sqrt{\log n/n}$ for $k = 1,\ldots,K_n$ corresponds by the Bonferroni inequality to an approximate 95% hypothesis test for $\varrho(\hat{l}+1),\ldots,\varrho(\hat{l}+K_n)$ simultaneously equal to zero. We have found that values in this range work well in practice.

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 Corollary 4 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* $\gamma_i \sim Ci^{-d}$ *for some* $C > 0$ *and positive integer* d, *then*

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

[ii] *If* $\gamma_k \sim C\theta^i$ *for some* $C > 0$ *and* $|\theta| < 1$, *then*

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

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

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

5. LINEAR PROCESS BOOTSTRAP

There are several bootstraps for time series data [see, for example, Bühlmann (2002), Lahiri (2003), or Politis (2003b) 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 that 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. The 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 Thm 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 Thm 8.3.1 in Brockwell and Davis (1991).

Below, we propose a new bootstrap, termed 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 generates 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 $Y = (Y_1,\ldots,Y_n)^{\mathsf{T}}$.
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 iid 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 $\hat{\Sigma}^{\epsilon}_{\kappa,l} = LL^{\mathsf{T}}$.

REMARK 4. 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 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 LPB algorithm can be used to produce confidence intervals for the mean, which are justified by Theorem 5.

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

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

and

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

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

REMARK 5. Surprisingly, the assumptions of Theorem 5 do not include linearity of the original time series, i.e. an MA(∞) model. The LPB manages to generate a linear process with covariance structure approaching that of the original time series while requiring only the weak dependence conditions used in estimating Σ_n. Thus, the range of applicability of the LPB is quite wide and includes a vast class of nonlinear processes. The name LPB therefore only alludes to the fact that a linear process is being generated in the bootstrap world. Furthermore, we conjecture that the LPB is applicable to all statistics that depend only on the first two moments of the process.

6. SIMULATIONS

6.1. Covariance matrix estimation

We conducted several simulations to assess the performance of our estimator and to draw a direct comparison with the method of Wu and Pourahmadi (2009); they use a subsampling rule to estimate the banding parameter 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 eqn (4); and infinitely differentiable, as defined in McMurry and Politis (2004) and shown in Figure 1. Simulations were performed in R (R Development Core Team, 2009). Results are based on $n = 100$ replications, and the parameters in the bandwidth choice rule were chosen to be $c = 2$, and $K_n = 5$. Each model we considered was tested with sample sizes $n = 250$, 500, and 750.

We assessed the the performance by computing a matrix norm of the difference between the estimated matrix and the true, i.e. $\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n)$, and then averaged these numbers over all 100 replications. We used two matrix norms, the operator norm $\rho(\cdot)$ and, to allow for a direct comparison with Wu and Pourahmadi (2009), the matrix infinity norm

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

The averaged errors in the two norms will be referred to respectively as *operator norm losses* and *infinity norm losses*.

We also tested the adjustment to positive definiteness given in Theorem 3 using the trapezoid weight function. While negative eigenvalues were occasionally observed, they were so close to zero that the results were numerically identical in operator norm loss to the unadjusted trapezoid estimator. For this reason, these results will not be further discussed. The rectangular weight function is

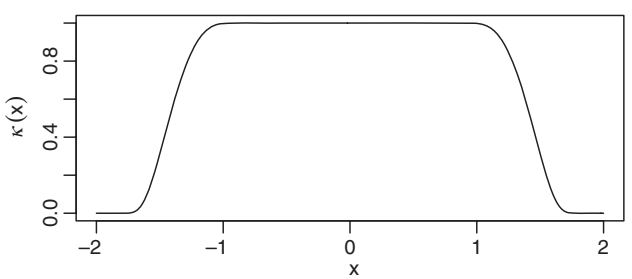

Figure 1. An infinitely differentiable weight function

expected to produce many more nonpositive matrices, as the negative sidelobes of its Fourier transform are more pronounced than those of the trapezoid's.

In what follows, we will qualitatively describe the experiments we ran and highlight the notable results. Full numeric results can be found in tables 1–3 of McMurry and Politis (2010). While most results were specific to individual simulations, across simulations the trapezoid weight function was consistently the best performing of the three weight functions. The infinitely differentiable weight function was often close but never better, and both were substantially better than the rectangular weight function.

6.1.1. MA(1)

In the first simulation, the data were generated by the moving average process $X_t = \epsilon_t + 0.5\epsilon_{t-1}$, with ϵ_t an iid sequence of N(0,1) random variables. For all three weight functions, the infinity norm losses given in Wu and Pourahmadi (2009) are more than 10 times our losses, and our method comes close to achieving the oracle bounds given in their table 1.

6.1.2. AR(1)

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

In all simulations, the performance of our estimator was substantially worse for $\phi = 0.9$ than for the two smaller values; this is to be expected, as off diagonal terms of Σ_n are substantially larger for large ϕ.

Losses in this case were also somewhat larger than losses for the MA model; this is also unsurprising as the MA covariance matrix can be estimated without tapering or truncating any non-zero entries, whereas tapering non-zero entries cannot be avoided for an AR process. Finally, as expected, results consistently improve with increasing sample size.

6.1.3. Absolute value AR(1)

For the final simulation, data were simulated from the model $X_t = \phi|X_{t-1}| + \epsilon_t$, for $\phi = 0.1$, 0.5, and 0.9, and 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 simulated data. While this provides (crude) estimates of the loss, it induces some additional difficulty to quantify uncertainty. Nonetheless, our results again show significant improvement over those presented in Wu and Pourahmadi (2009), particularly for smaller values of ϕ. With $n = 250$, our infinity norm losses for $\phi = 0.1$ were always less than 25% of theirs, for $\phi = 0.5$ our losses were always less than 50% of theirs, and for $\phi = 0.9$, our loss was about 90% of theirs; our method's advantage improved with increasing n.

6.2. Linear process bootstrap

Finally, we ran several simulation experiments to assess the performance of the LPB. For comparison, we also tested each simulated data set using two other popular resampling schemes. First, we considered the block bootstrap as implemented in Canty and Ripley (2010), using the block length selection of Politis and White (2004) (see also Patton *et al.*, 2009). Second, we also used the AR sieve bootstrap with AIC for model selection. Reviews of the block and AR sieve bootstraps can be found in Bühlmann (2002) and Politis (2003b). Each experiment was repeated 1000 times using 1000 bootstrap replications. Full numeric results can be found in table 4 of McMurry and Politis (2010).

6.2.1. MA(1)

The first model we considered was the moving average process $X_t = \epsilon_t + \phi\epsilon_{t-1}$ for $\phi = 0.1$, 0.5, and 0.9, and ϵ_i iid N(0,1). We used sample sizes $n = 250$, 500, and 750. All three bootstraps performed well for most combinations of ϕ and n, with typical coverage of 93–94% for nominal 95% confidence intervals. The LPB seemed to be slightly better than its two competitors for the two larger values of ϕ, and comparable or very slightly worse for $\phi = 0.1$.

It is expected that the LPB should excel for MA processes, as its banded covariance matrix structure can produce data that exactly possesses an MA covariance structure. The AR sieve should have the most difficulty in this setting because an MA(1) process can only be well represented as a very high order AR process; this was evident when $n = 250$ and $\phi = 0.9$, where the AR sieve only produced coverage of 90%. In all other situations, the AR-sieve's performance was very similar to the other bootstrap estimators.

6.2.2. AR(1)

The second model we considered was the AR(1) process considered in Section 6.1.2, with the same values of ϕ and n. The LPB performed slightly worse than the block bootstrap for the two smaller values of ϕ and noticeably better when $\phi = 0.9$.

The AR sieve consistently outperformed the other two methods; this was expected because an AR model is exactly correct in this setting. Perhaps the most surprising feature of the simulations was that the sieve bootstrap performed somewhat poorly when the AR coefficient was 0.9. In this case, for nominal 95% confidence intervals, we observed coverages of 88% with $n = 250$, 91% with $n = 500$, and 94% with $n = 750$. Although the AR sieve bootstrap is expected to break down when ϕ is close to 1, the bad behaviour for $\phi = 0.9$ was unexpected.

6.2.3. Absolute value AR(1)

Finally, we considered the absolute value AR(1) process described in Section 6.1.3. In this case, the LPB's coverage was identical to the block bootstrap's when $\phi = 0.1$, noticeably worse when $\phi = 0.5$, and somewhat better when $\phi = 0.9$.

The AR sieve again consistently outperformed the other two estimators in this setting. Surprisingly, in this situation, the AR sieve did as well as or better than it did for the AR(1) processes, for which it should be ideal. Here, the AR sieve's coverage was never more than 3% different from the LPB's.

7. TECHNICAL PROOFS

The proof of Theorem 1 rests on Lemma 1.

LEMMA 1 (WU AND POURAHMADI, 2009). *Assume that* $\{X_i\}$ *satisfies* $\Delta_{2q} < \infty$ *with* $1 < q \leq 2$. *Then for any* $j \in \mathbb{Z}$,

$$\left\| \sum_{i=1}^{n} X_i X_{i+|j|} - n\gamma_j \right\|_q \leq 2B_q n^{1/q} \|X1\|_{2q} \Delta_{2q},$$

where

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

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

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

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, can be studied 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,

$$\|T_2\|_q \leq d_q (\lfloor c_\kappa l \rfloor - l) n^{-(q-1)/q} + \frac{2}{n} \sum_{i=l+1}^{\lfloor c_\kappa l \rfloor} i|\gamma_i| + 2 \sum_{i=l+1}^{\lfloor c_\kappa l \rfloor} |\gamma_i|. \tag{8}$$

The final term of eqn (8) can now be combined with T_3. Since $c_\kappa \geq 1$,

$$2 \sum_{i=l+1}^{\lfloor c_\kappa l \rfloor} |\gamma_i| + T_3 = 2 \sum_{i=l+1}^{n} |\gamma_i|,$$

the last term in eqn (5). □

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, $\rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n) = O_p(r_n)$. 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

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

Similarly,

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

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,

$$\rho(\hat{\Sigma}_{\kappa,l}^{\epsilon} - \Sigma_n) \le \rho(\hat{\Sigma}_{\kappa,l} - \Sigma_n) + \rho(\hat{\Sigma}_{\kappa,l}^{\epsilon} - \hat{\Sigma}_{\kappa,l}). \tag{9}$$

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

$$\begin{aligned}d_n &= \lambda_{\min}(\hat{\Sigma}_{\kappa,l}) \\ &= -\lambda_{\max}(-\hat{\Sigma}_{\kappa,l}) \\ &\ge -\lambda_{\max}(\Sigma_n - \hat{\Sigma}_{\kappa,l}) \\ &\ge -\rho(\Sigma_n - \hat{\Sigma}_{\kappa,l}),\end{aligned} \tag{10}$$

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

We now focus on the second term of eqn (9).

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

where $D^{-} = \mathrm{diag}(\max(d_1, \epsilon\hat{\gamma}_0/n^{\beta}) - d_1, \ldots, \max(d_n, \epsilon\hat{\gamma}_0/n^{\beta}) - d_n)$. By the above spectral decomposition and inequality (10),

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

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

PROOF OF THEOREM 5. By Theorem 3 in Wu (2005),

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

We establish eqn (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 $(\hat{\Sigma}_{\kappa,l}^{\epsilon})^{-1/2}Y$. Let $\mathbf{1}$ be the n-vector of 1s. Under the conditions of Theorem 5,*

$$\begin{aligned}n^{1/2}\bar{Y}^* &= n^{-1/2}\mathbf{1}^{\mathsf{T}}(\hat{\Sigma}_{\kappa,l}{}^{\epsilon})^{1/2} Z^* \\ &= n^{-1/2}\mathbf{1}^{\mathsf{T}}(\Sigma_n^{1/2})\tilde{Z}^* + n^{-1/2}\mathbf{1}^{\mathsf{T}}(\Sigma_n^{1/2})(Z^* - \tilde{Z}^*) + n^{-1/2}\mathbf{1}^{\mathsf{T}}[(\hat{\Sigma}_{\kappa,l}{}^{\epsilon})^{1/2} - \Sigma_n^{1/2}]Z^* \\ &= n^{-1/2}\mathbf{1}^{\mathsf{T}}(\Sigma_n^{1/2})\tilde{Z}^* + R_1 + R_2 \\ &= n^{-1/2}\mathbf{1}^{\mathsf{T}}(\Sigma_n^{1/2})\tilde{Z}^* + o_P(1).\end{aligned} \tag{11}$$

PROOF OF LEMMA 2. We first consider R_2. It has bootstrap mean 0 and variance

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

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

For R_1, we can write, $Z^* = \hat{\sigma}_W^{-1}M^*(I - n^{-1}\mathbf{1}_n)\hat{\Sigma}_{\kappa,l}^{-1/2}Y$ 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, $\tilde{Z}^* = \hat{\sigma}_{\bar{W}}^{-1}M^*(I - n^{-1}\mathbf{1}_n)\Sigma_n^{-1/2}Y$. Since $|\hat{\sigma}_W^2 - \hat{\sigma}_{\bar{W}}^2| = o_P(1)$, and both $\hat{\sigma}_W^2$ and $\hat{\sigma}_{\bar{W}}^2$ are bounded away from zero and from above with probability tending to 1,

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

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

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

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 iid, $E^*[V^*(V^*)^{\mathrm{T}}] = \sigma_V^2 I$, where

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

Since $Y^{\mathrm{T}}Y = O_P(n)$ (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.

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

PROOF OF LEMMA 3.

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

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

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 $n^{-1/2}\mathbf{1}^{\mathrm{T}}A_n\tilde{Z}^* = n^{-1/2}\mathbf{1}^{\mathrm{T}}B_n\tilde{Z}^* + o_P(1)$.*

PROOF OF LEMMA 4. Since $\tilde{Z}^*$ has bootstrap mean zero, it is sufficient to show the variance converges to zero in probability.

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

□

LEMMA 5. *Under the assumptions of Theorem 5, $E[\tilde{Z}_i^4]$ is 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 Thm 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 Prop. 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 $\sum_{j=1}^n a_{ij}^2 = e_i^{\mathrm{T}} \Sigma_n^{-1} e_i$, where e_i is the ith standard basis vector, for large enough n, $\sum_{j=1}^n a_{ij}^2$ is bounded from above and away from zero uniformly in i. By Thm 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.

PROOF OF THEOREM 5. We do so by approximating $\Sigma_n^{1/2}$ in eqn (11) in the following manner. Let $\Sigma_{n,k} = [\gamma_{|i-j|} 1_{|i-j| \leq k}]_{i,j=1}^n$ 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 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 eqn (11) by $L_{n,k}$.

Matrix multiplication shows that $L_{n,k}$ is non-zero 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 eqn (11).

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

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

$$\frac{1}{(c_{1,n}^2 + \cdots + c_{n,n}^2)^{1+\delta/2}} \sum_{i=1}^n E^* \left[|c_{i,n} \tilde{Z}_i^*|^{2+\delta} \right] \to 0, \tag{12}$$

for some $\delta > 0$. Convergence (12) 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 (12).

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

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 $(2\pi c_1 - \epsilon)n < \mathbf{1}^{\mathrm{T}} \Sigma_{n,k} \mathbf{1} < (2\pi c_2 + \epsilon)n$. This establishes that the denominator of (12) 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 eqn (12) converges to 0 in probability. With Lemma 3, this completes the proof. □

Acknowledgements

The authors thanks Mohsen Pourahmadi, Wei Biao Wu, and an anonymous referee for their comments, critiques, and helpful suggestions.

REFERENCES

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer.
Bühlmann, P. (1997) Sieve bootstrap for time series. *Bernoulli* **3**(2):123–48.
Bühlmann, P. (2002) Bootstraps for time series. *Statistical Science* **17**(1):52–72.
Canty, A. and Ripley, B. D. (2010) *Boot: Bootstrap R (S-Plus) Functions*. R Package Version 1.2–42.
Dedecker, J. and Doukhan, P. (2003) A new covariance inequality and applications. *Stochastic Processes and their Applications* **106**(1):63–80.
Grenander, U. and Szegö, G. (1958) *Toeplitz Forms and Their Applications*. Los Angeles: University of California Press.
Horn, R. A. and Johnson, C. R. (1990) *Matrix Analysis*. New York: Cambridge University Press.
Kreiss, J.-P. (1992) Bootstrap procedures for AR(∞)-processes. In *Bootstrapping and related techniques (Trier, 1990), Lecture Notes in Economies and Mathematical Systems*, vol. 376, Berlin: Springer, pp. 107–13.
Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**:1217–41.
Lahiri, S. N. (2003) *Resampling Methods for Dependent Data*, Springer Series in Statistics. New York: Springer-Verlag.
Liu, R. Y. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (eds R. LePAGE and BILLARD), New York: Wiley, pp. 225–48.
McMurry, T. and Politis, D. N. (2004) Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics* **16**(3–4):549–62.
McMurry, T. and Politis, D. N. (2010) Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. Available at: http://www.math.ucsd.edu/~politis/PAPER/covmat3-2010.pdf. Accessed on 1 August 2010.
Paparoditis, E. and Streitberg, B. (1992) Order identification statistics in stationary autoregressive moving-average models: vector autocorrelations and the bootstrap. *Journal of Time Series Analysis* **13**(5):415–34.
Patton, A., Politis, D. N. and White, H. (2009) Correction to 'Automatic block-length selection for the dependent bootstrap'. *Econometric Reviews* **28**(4):372–5.
Politis, D. N. (2003a) Adaptive bandwidth choice. *Journal of Nonparametric Statistics* **25**(4–5):517–33.
Politis, D. N. (2003b) The impact of bootstrap methods on time series analysis. *Statistical Science* **18**(2):219–30.
Politis, D. N. (2007) Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *UCSD Department of Economics, Discussion Paper* 2005. Accepted to Econometric Theory.
Politis, D. N. and Romano, J. P. (1995) Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* **16**(1):67–103.
Politis, D. N. and White, H. (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* **23**(1):53–70.
R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at URL: `http://www.R-project.org`. ISBN 3-900051-07-0.
Resnick, S. I. (1999) *A Probability Path*. Boston: Birkhauser.
Wu, W. B. (2005) Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences* **102**:14150–54.
Wu, W. B. and Pourahmadi, M. (2009) Banding sample autocovariance matrices of stationary processes. *Statistica Sinica* **19**(4):1755–68.