A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices

Tommaso Proietti and Alessandro Giovannelli

CREATES Research Paper 2017-20

A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices

Tommaso Proietti[*]
Università di Roma "Tor Vergata" and CREATES

Alessandro Giovannelli
Università di Roma "Tor Vergata"

May 16, 2017

Abstract

We consider the problem of estimating the high-dimensional autocovariance matrix of a stationary random process, with the purpose of out of sample prediction and feature extraction. This problem has received several solutions. In the nonparametric framework, the literature has concentrated on banding and tapering the sample autocovariance matrix. This paper proposes and evaluates an alternative approach, based on regularizing the sample partial autocorrelation function, via a modified Durbin-Levinson algorithm that receives as input the banded and tapered partial autocorrelations and returns a sample autocovariance sequence which is positive definite. We show that the regularized estimator of the autocovariance matrix is consistent and its convergence rates is established. We then focus on constructing the optimal linear predictor and we assess its properties. The computational complexity of the estimator is of the order of the square of the banding parameter, which renders our method scalable for high-dimensional time series. The performance of the autocovariance estimator and the corresponding linear predictor is evaluated by simulation and empirical applications.

Keywords: Toeplitz systems; Optimal linear prediction; Partial autocorrelation function.

JEL Codes: C22, C53, C55.

1 Introduction

Let $\{Y_t, t = 1, \ldots\}$ be a real valued, discrete time stationary random process, with mean zero and autocovariance function $\gamma(j) = \mathrm{E}(Y_t \cdot Y_{t-j}), j = 0, \pm 1, \ldots$. Several essential characteristics of the process can be defined in terms of the autocovariance sequence, such as the spectral density, $f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-\iota\omega j}$, $\omega \in [-\pi, \pi]$, where $\iota^2 = -1$, and the long run variance $V = \sum_{j=-\infty}^{\infty} \gamma(j), V = 2\pi f(0))$.

The $n \times n$ autocovariance matrix

$$
\Gamma_n = \left[\begin{array}{cccc}
\gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\
\gamma(1) & \gamma(0) & \ddots & \gamma(n-2) \\
\vdots & \ddots & \ddots & \vdots \\
\gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0)
\end{array} \right],
\tag{1}
$$

plays an important role for optimal linear prediction and interpolation from a finite sample. In particular, the minimum mean squared error linear predictor of Y_{n+h} at time n, based on the information set $\{Y_1, \ldots, Y_n\}$, is obtained as $\hat{Y}_{n+h|n} = \sum_{j=1}^{n} \phi_{nj}^{(h)} Y_{n-j-1}$, where the coefficients $\phi_n^{(h)} = [\phi_{n1}^{(h)}, \phi_{n2}^{(j)}, \ldots, \phi_{nn}^{(j)}]'$ are the solution of the Yule-Walker system

$$
\phi_n^{(h)} = \Gamma_n^{-1} \gamma_n^{(h)},
$$

with $\gamma_n^{(h)} = [\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1)]'$.

The literature has investigated the problem of estimating Γ_n from a time series realization, $\{y_t, t = 1, \ldots, n\}$. Denoting the sample autocovariance at lag j by

$$
\hat{\gamma}(j) = \frac{1}{n} \sum_{t=j+1}^{n} y_t y_{t-j}, j = 0, 1, \ldots, n-1,
\tag{2}
$$

we can construct the sample autocovariance matrix $\hat{\Gamma}_n = \{\hat{\gamma}(|i-j|), i, j = 1, \ldots, n\}$, which is a positive definite Toeplitz matrix. However, as it was shown by Wu and Pourahmadi (2009), $\hat{\Gamma}_n$ is not consistent for Γ_n, as the operator norm (i.e. the largest eigenvalue) of the estimation error matrix, $\rho(\hat{\Gamma}_n - \Gamma_n)$, does not converge to zero as n tends to infinity.

Wu and Pourahmadi (2009) proposed the *banded* autocovariance matrix estimator

$$
\hat{\Gamma}_{n,\ell} = \{\hat{\gamma}(|i-j|) \cdot I(|i-j| \leq \ell), i, j = 1, \ldots, n\},
\tag{3}
$$

where ℓ is the banding parameter and $I(\cdot)$ is the indicator function.

E. J. Hannan and Deistler (1988) showed that, for linear ARMA processes and for $\ell \leq (\log n)^\alpha, \alpha < \infty$, the infinity norm of $\hat{\Gamma}_{n,\ell} - \Gamma_n$ is $O\left(n^{-1/2}\sqrt{\log\log n}\right)$. Wu and Pourahmadi (2009) proved the consistency of (3) for the class of non-linear short-range dependent processes considered by Wu (2005), and obtained an explicit upper bound for the operator norm of $\hat{\Gamma}_{n,\ell} - \Gamma_n$ See Appendix A for a review of the different matrix norms. Bickel and Gel (2011) obtained the consistency of (3) under the Frobenius norm. Bickel and Gel (2011), and proposed a cross-validation method for selecting the optimal banding parameter.

McMurry and Politis (2010), MP henceforth, proposed the *banded and tapered* autocovariance matrix estimator

$$
\hat{\Gamma}_{n,MP} = \{\hat{\gamma}(|i-j|) \cdot w(|i-j|), i, j = 1, \ldots, n\},
\tag{4}
$$

2

with $w(|i-j|) = \kappa((i-j)/\ell)$, where ℓ is the banding parameter and $\kappa(u)$ is the trapezoidal kernel

$$\kappa(u) = \begin{cases} 1, & |u| \leq 1 \\ 2 - |u|, & 1 < |u| \leq 2 \\ 0, & |u| > 2 \end{cases} \tag{5}$$

McMurry et al. (2015) further develop the theory of optimal linear prediction using the banded and tapered sample autocovariance sequence.

Both estimators (3) and (4) preserve the Toeplitz structure, but they are not necessarily positive definite. In general, the estimator $\{\hat{\gamma}(|i-j|) \cdot w(|i-j|)\}$ is positive definite if and only if $\sum_i \sum_j w(|i-j|)e^{-\imath\omega|i-j|} > 0$, $\forall \omega \in [-\pi, \pi]$, which does not holds in either cases. McMurry et al. (2015) present and compare several alternative positivity corrections.

This paper proposes an autocovariance matrix estimator based on the idea of banding and tapering the sample partial autocorrelation sequence and deriving the implied autocovariance by inverting the Durbin-Levinson (DL) algorithm. The advantages of modelling a covariance matrix via the partial autocorrelations have been considered by Daniels and Pourahmadi (2009). The DL algorithm (see Brockwell and Davis (1991) and Barndorff-Nielsen and Schou (1973)) maps the autocovariance sequence into the partial autocorrelations. It can be inverted so that the regularized sample partial autocorrelations, resulting from the application of a weighting function, such as the trapezoidal kernel with banding parameter ℓ, are mapped into regularized sample autocovariances, denoted by $\tilde{\gamma}_r(j), j = 0, \ldots, n-1$, satisfying the condition $\sum_j \tilde{\gamma}_r(j)e^{-\imath\omega j} > 0$, $\omega \in [-\pi, \pi]$. We name our method the regularized Durbin-Levinson (RDL) estimator of Γ_n. As a by-product, the RDL estimates the coefficients of the linear predictor of Y_{n+h} based on the sample realization $\{y_1, \ldots, y_t, \ldots, y_n\}$. The estimator has the following comparative advantages:

- It is positive definite by construction and no positivity correction is needed.

- The optimal linear predictor based on a sample of size n requires a number of operations which is $O(\ell^2)$. Hence, it can be computed for large n with a complexity increasing only with the square of the banding parameter, which is $o(n)$.

- It outperforms the banded and tapered autocovariance estimator for processes with high dynamic range, such as a cyclical process.

The plan of the paper is the following. In section 2 we present the RDL algorithm and derive its properties when it is applied to the true autocovariances of the process. Section 3 contains the main results, showing the consistency of the RDL estimator of the autocovariance matrix. Section 4 deals with optimal linear prediction using the RDL estimator. We consider the problem of selecting the banding parameter (section 5). In section 6 we assess the performance of our method by simulation experiments, dealing with the ability of estimating the autocovariance function of known processes belonging to the ARMA class, and by out-of-sample rolling forecast experiments dealing with the prediction of sea surface temperatures in the the Niño 3.4 region. Section 7 concludes the paper.

2   Regularized Durbin-Levinson Algorithm

The Durbin-Levinson algorithm (see Levinson (1946) Durbin (1960), and Brockwell and Davis (1991) and Brémaud (2014)) processes the autocovariances $\{\gamma(0), \ldots, \gamma(k)\}$, recursively for $k = 1, \ldots, n-1$, and computes the coefficients of the optimal one-step-ahead predictor at time $t > k$, $\hat{Y}_{t|t-1} = \sum_{j=1}^{k} \phi_{kj}Y_{t-j}$, based on k past values, where ϕ_{kk} is the partial autocorrelation between

Y_t and Y_{t-k}, as well as the prediction error variance $\varsigma_k = \mathrm{Var}(Y_t|Y_{t-1}, \ldots, Y_{t-k})$. It provides the solution of a Yule-Walker system of equations in $O(k^2)$ operation, by making an efficient use of the Toeplitz structure of Γ_{k+1}.

The following regularised Durbin-Levinson (RDL) algorithm augments the usual DL recursions by equations that i. regularize the partial autocorrelation at lag k by applying a weight $w_k = \kappa(k/\ell)$, where $\kappa(|u|)$ is the trapezoidal kernel given in (5) and ℓ is the banding parameter; ii. compute the regularized autocovariance $\gamma_r(k)$, corresponding to a process with partial autocorrellation function $|w_k\phi_{kk}| < 1$; iii. obtain the coefficients $\pi_{kj}, j = 1, \ldots, k$, of the regularized predictor.

The algorithm is initialised as follows: set $\gamma_r(0) = \gamma(0)$ and

$$
\begin{array}{llll}
\varsigma_0 & = & \gamma(0), & v_0 & = & \gamma_r(0), \\
\phi_{11} & = & \gamma(1)/\gamma(0), & \pi_{11} & = & w_1\phi_{11}, \\
& & & \gamma_r(1) & = & v_0\pi_{11}, \\
\varsigma_1 & = & (1 - \phi_{11}^2)\varsigma_0, & v_1 & = & (1 - \pi_{11}^2)v_0.
\end{array}
\tag{6}
$$

Then, for $k = 2, \ldots, n - 1$, the following recursions are run:

$$
\begin{array}{llll}
\phi_{kk} & = & \frac{\gamma(k) - \sum_{j=1}^{k-1} \phi_{k-1,j}\gamma(k-j)}{\varsigma_{k-1}}, & \pi_{kk} & = & w_k\phi_{kk}, \\
& & & \gamma_r(k) & = & \sum_{j=1}^{k-1} \pi_{k-1,j}\gamma_r(k-j) + v_{k-1}\pi_{kk}, \\
\phi_{kj} & = & \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}, & \pi_{kj} & = & \pi_{k-1,j} - \pi_{kk}\pi_{k-1,k-j}, \quad j = 1, 2, \ldots, k - 1, \\
\varsigma_k & = & (1 - \phi_{kk}^2)\varsigma_{k-1}, & v_k & = & (1 - \pi_{kk}^2)v_{k-1}.
\end{array}
\tag{7}
$$

The recursions on the left of (7) are the traditional DL recursions, mapping the autocovariances $\gamma(k)$ into the partial autocorrelations ϕ_{kk}. The recursions on the right perform the reverse mapping, from the regularized partial autocorrelations, $\pi_{kk} = w_k\phi_{kk}$, to the regularized autocovariances, $\gamma_r(k)$, $k = 0, \ldots, n - 1$. If $w_k = 1$, for all k, we recover the usual DL recursion yielding the raw partial autocorrelations, and $\gamma_r(k) = \gamma(k)$ for all k. If $w_k = 1$ for $k \leq p$, and $w_k = 0$ for $k > p$, the system (7) computes the usual Yule-Walker predictor based on p lagged values, i.e. it amounts to fitting an autoregressive model of order p. In this case $\gamma_r(k) = \gamma(k)$ for $k = 0, 1, \ldots p$ and $\gamma_r(k) = \sum_{j=1}^{p} \phi_{p,j}\gamma_r(k - j)$ for $|k| > p$.

Remark 1. *An alternative to tapering the sequence ϕ_{kk} is tapering the Fisher transform $\theta_k = \frac{1}{2}\ln\frac{1+\phi_{kk}}{1-\phi_{kk}}$ and obtaining $\pi_{kk} = \frac{\exp(2w_k\theta_k)-1}{\exp(2w_k\theta_k)+1}$. The solution does not differ relevantly, as $\frac{\partial\pi_{kk}}{\partial w_k} = 4\theta_k\exp(2w_k\theta_k)/[1 + \exp(2w_k\theta_k)]^2$, and $\frac{\partial\pi_{kk}}{\partial w_k}|_{w_k=0} = \theta_k$.*

The sequence $\{\pi_{kk}, k = 1, \ldots, n - 1\}$ is a proper partial autocorrelation function, as $|\pi_{kk}| < 1$ by construction. It is associated with an auxiliary process, Y_t^*, which is AR(L), with $L = \lfloor 2\ell \rfloor$. This is evident from $\pi_{kk} = 0, k > L$. An alternative equivalent expression for the regularized partial autocorrelations is

$$
\begin{aligned}
\pi_{kk} & = \frac{\gamma_r(k) - \sum_{j=1}^{k-1} \pi_{k-1,j}\gamma_r(k-j)}{\gamma_r(0) - \sum_{j=1}^{k-1} \pi_{k-1,j}\gamma_r(j)} \\
& = \frac{\mathrm{Cov}(Y_t^*, Y_{t-k}^*|Y_{t-1}^*, \ldots, Y_{t-k}^*)}{\mathrm{Var}(Y_t^*|Y_{t-1}^*, \ldots, Y_{t-k}^*)}.
\end{aligned}
$$

Notice also that the first ℓ autocovariances are identical to the original, i.e. $\gamma_r(j) = \gamma(j)$ for $j \leq \ell$, as for the trapezoidal kernel $w_j = 1, j \leq \ell$.

The following two lemmata deal with the properties of the RDL algorithm.

Lemma 1. *Let $\gamma_r(j)$ and $\pi_{n-1,j}, j = 1, \ldots, n-1$, be defined as in (7). Let $0 \le w_j \le 1, j = 1, \ldots, n$, be the trapezoidal weigthing functions with banding parameter ℓ. Then the following results hold.*

(i) *The regularized autocovariance function $\{\gamma_r(k), k = 0, \pm 1, \ldots\}$ is a positive definite sequence.*

(ii) *The roots of the polynomial $\Pi_k(z) = 1 - \sum_{j=1}^k \pi_{k,j} z^j$ lie strictly inside the unit circle, i.e. $\Pi_k(1/z)$ is a Schur, or minimum phase, polynomial.*

(iii)

$$\sum_{j=1}^{n-1} |\pi_{n-1,j} - \phi_{n-1,j}| \le 2 \sum_{j=\ell+1}^{n-1} |\phi_{jj}| \prod_{i=1}^{j-1} (1 + |\phi_{ii}|). \tag{8}$$

Proof. See Appendix B. □

The following lemma characterises the nature of the tapered approximation of Y_t.

Lemma 2. *Let the autocovariance function of Y_t be summable, with $\sum_{j=-\infty}^{\infty} |j|^r |\gamma(j)| < \infty, r \in \mathbb{N}$, and $\sum_j \gamma(j) e^{-i\omega j} > 0, \forall \omega \in [-\pi, \pi]$, and denote $Y_t = \sum_{j=1}^{\infty} \phi_j Y_{t-j} + \epsilon_t, \epsilon_t \sim WN(0, \sigma^2)$, the AR($\infty$) representation. Then,*

$$\sum_{j=1}^{n-1} |\pi_{n-1,j} - \phi_j| \le C \sum_{j=\ell+1}^{\infty} |\phi_{jj}|, \tag{9}$$

$$\ln v_{n-1} - \ln \sigma^2 \le \sum_{j=\ell+1}^{\infty} \frac{\phi_{jj}^2}{1 - \phi_{jj}^2}. \tag{10}$$

If further $n, \ell \to \infty$, so that $\ell = o(n)$

$$\sum_{j=1}^{\infty} j^r |\gamma(j)| < \infty \Rightarrow \lim_{n \to \infty} \sum_{j=1}^{n-1} j^r |\gamma_r(j)| < \infty. \tag{11}$$

Proof. See Appendix C. □

Remark 2. *The assumption $\sum_{j=-\infty}^{\infty} |j|^r |\gamma(j)| < \infty$ is a smoothness condition implying that the spectral density has r continuous derivatives. If Y_t is a stationary ARMA process, its autocovariance function decreases to zero at a geometric rate and the above condition holds for any positive integer r.*

3 The Autocovariance Matrix Estimator

Given the time series $\{y_t, t = 1, 2, \ldots, n\}$, our estimator of the autocovariance matrix is obtained by running the regularized Durbin-Levinson recursions (7) on the raw sample autocovariances, $\{\hat{\gamma}(k), k = 0, \pm 1, \ldots, \pm(n-1)\}$, where $\hat{\gamma}(k)$ is given in (2). The latter shrinks the sample partial autocorrelations $\hat{\phi}_{kk}$ towards zero by setting $\tilde{\pi}_{kk} = w_k \hat{\phi}_{kk}$, and yields the regularized autocovariance sequence $\{\tilde{\gamma}_r(k), k = 0, 1, \ldots, n-1\}$, along with the estimated AR coefficients $\tilde{\pi}_{k,j}, j = 1, \ldots, k-1$.

In the sequel, we will denote the Toeplitz RDL sample autocovariance matrix constructed from the regularized sample autocovariances by

$$\tilde{\Gamma}_n = \{\tilde{\gamma}_r(|i-j|), i, j = 1, \ldots, n\}. \tag{12}$$

The RDL algorithm delivers the elements of the UDU decomposition of the inverse autocovariance function

$$\tilde{\Gamma}_n^{-1} = C_n' D_n C_n,$$

where $D_n = \text{diag}\left(\tilde{v}_0^{-1}, \tilde{v}_1^{-1}, \ldots, \tilde{v}_{n-1}^{-1}\right),$

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\tilde{\pi}_{11} & 1 & 0 & \cdots & 0 \\ -\tilde{\pi}_{22} & -\tilde{\pi}_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ -\tilde{\pi}_{n-1,n-1} & -\tilde{\pi}_{n-1,n-2} & -\tilde{\pi}_{n-1,n-3} & \cdots & 1 \end{bmatrix}. \tag{13}$$

The sampling properties of the estimator (12) will be investigated under a suitable set of assumptions concerning the nature of the underlying random process (Assumption 1) and the design of the estimator (Assumption 2).

Assumption 1. *Let $\{Y_t\}$ be the stationary linear process $Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ with $\psi_0 = 1$, and ϵ_t satisfying*

i. $E(\epsilon_t | \mathcal{F}_{t-1}) = 0,$

ii. $E(\epsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2 < \infty,$ *and* $E(\epsilon_t^4) < \infty.$

iii. $f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-\imath \omega j} \neq 0, \forall -\pi < \omega \leq \pi.$

iv. $\sum_{j=-\infty}^{\infty} (1 + |j|) |\gamma(j)| < \infty$

Remark 3. *The assumption 1. ii can be relaxed to allow for conditional heteroscedasticity as in Goncalves and Kilian (2007), who replace it by the assumption that $E(\epsilon_t^2) = \sigma^2$ and the that the joint cumulants of ϵ_t up to the eight order are absolutely summable. The assumption that Y_t is a linear process can be relaxed as in Wu and Pourahmadi (2009) and McMurry and Politis (2010), who assume that Y_t is a non linear process of the form $Y_t = g(\epsilon_t, \epsilon_{t-1}, \ldots)$, where $g(\cdot)$ is a measurable function of the i.i.d. random variables $\epsilon_s, s = t, t - 1, \ldots,$ possessing finite fourth moment and finite physical dependence (see e.g. Wu (2005)).*

Remark 4. *Assumptions iii. and iv. imply that $\sum_j (1 + |j|) |\psi_j| < \infty$ and that the infinite AR representation $Y_t = \sum_{j=1}^{\infty} \phi_j Y_{t-j} + \epsilon_t$ has $\sum_j (1 + |j|) |\phi_j| < \infty$. See Brillinger (1981), Theorem 3.8.4, p. 78. Moreover, the spectral density is continuous and bounded and there exists two real positive numbers $m = \inf_\omega f(\omega)$ and $M = \sum_\omega f(\omega)$, such that the eigenvalues of Γ_n, $\lambda_1, \lambda_2, \ldots, \lambda_n$, satisfy ((Brockwell and Davis, 1991), Proposition 4.5.3)*

$$0 < m \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n < M < \infty.$$

Assumption 2. *The following assumptions are made concerning the banding parameter ℓ and a rate p_n, such that $\ell < p_n < n$:*

i. $p_n = O(n^\alpha), 0 < \alpha < 1.$

ii. $\kappa(\cdot)$ *is the trapezoidal kernel with banding parameter $\ell = o\left(n^{1/2}\right)$, and $L = \lfloor 2\ell \rfloor$ is such that with $p_n / \ell \to \infty$ as $n \to \infty$.*

iii. *The quantity*

$$\mathrm{r}_n = \frac{\ell}{\sqrt{n}} + \sum_{j=\ell}^{\infty} |\gamma(j)|, \tag{14}$$

converges to zero as $n \to \infty$.

iv. $\lim_{n\to\infty} \left(\frac{p_n}{n}\right)^{1/2} \ell = 0.$

v $\ell = O(n^\beta)$, $0 < \beta < 1/2$, *and* $\alpha + 2\beta < 1$.

The following theorem establishes the consistency the RDL estimator and its rate of convergence, and shows that inverse autocovariance matrix can be estimated consistently using $\tilde{\Gamma}_n^{-1}$.

Theorem 1. *Under the assumptions 1 and 2 (i.)–(iv.), $\tilde{\Gamma}_n$ converges in operator norm to Γ_n, and*

$$\rho(\tilde{\Gamma}_n - \Gamma_n) = O_p\left(p_n^{\frac{1}{2}}\mathrm{r}_n\right). \tag{15}$$

Moreover,

$$\rho(\tilde{\Gamma}_n^{-1} - \Gamma_n^{-1}) = O_p\left(p_n^{\frac{1}{2}}\mathrm{r}_n\right), \tag{16}$$

Proof. See Appendix F $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 5. *Wu and Pourahmadi (2009) and McMurry and Politis (2010) obtained the faster convergence rate r_n for the banded (and tapered) estimator. The introduction of $p_n > L$ is needed for having $\sum_{p_n+1}^{\infty} |\tilde{\gamma}_r(j)| = o_p(1)$. McMurry et al. (2015) also need to introduce p_n to establish the convergence of the best linear predictor to the oracle predictor. Bickel and Gel (2011) defined their ℓ-banded estimator in terms of a submatrix of $\hat{\Gamma}_n$ of dimension $p_n < n$, where $p_n = o(n)$.*

Remark 6. *The optimal choice of the rate for ℓ depends on the rate of decay of $\gamma(j)$. If $|\gamma(j)| = O(|j|^{-d})$, $d > 1$, the optimal rate for ℓ is as in Assumption 2 v. with $\beta = -1/(2d)$. If $|\gamma(j)| = O(|\xi|^j)$, $|\xi| < 1$, then $\ell = O(\ln(n))$ and $p_n = O(n^\alpha)$, $\alpha < 1/2$.*

Our second result deals with the consistency and the rate of convergence of the coefficients of the finite predictor, $\tilde{\pi}_{n-1,j}, j = 1, \ldots, n-1$.

Theorem 2. *Under the assumptions 1 and 2 (i.)–(iv.),*

$$\left[\sum_{j=1}^{n-1} (\tilde{\pi}_{n-1,j} - \phi_{n-1,j})^2\right]^{1/2} = O_p(\mathrm{r}_n), \tag{17}$$

Proof. See Appendix F $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4 Optimal linear prediction and interpolation

We turn now our attention to the problem of predicting Y_{n+1} from a sample realization of length n. The optimal linear predictor, $\hat{Y}_{n+1|n} = \sum_{j=1}^n \phi_{n,j} Y_{t-j+1}$ can be estimated by $\tilde{Y}_{n+1|n} = \sum_{j=1}^n \tilde{\pi}_{n,j} Y_{t-j+1}$, where the coefficients $\tilde{\pi}_{n,j}$ are computed at time n by the RDL algorithm using the sample autocovariances up to lag $n-1$ and setting $\pi_{kk} = 0$ for $k = L+1, \ldots, n$. Hence, $\tilde{Y}_{n+1|n} = \sum_{j=1}^L \tilde{\pi}_{L,j} Y_{t-j+1}$, which means that the computational complexity of the one-step-ahead predictor is $O(\ell^2)$ rather than $O(n^2)$.

Theorem 3. *Optimal linear prediction Under the assumptions 1 and 2 (i.)–(iv.), and*

$$n^{1/2} \sum_{L+1}^{\infty} |\phi_j| \to 0, \tag{18}$$

the RDL predictor $\tilde{Y}_{n+1|n} = \sum_{j=1}^{n} \tilde{\pi}_{n,j} Y_{n-j+1}$ converges in probability to the oracle predictor $\hat{Y}_{n+1|n} = \sum_{j=1}^{n} \phi_{n,j} Y_{t-j+1}$, and in particular $|\hat{Y}_{n+1|n} - \tilde{Y}_{n+1|n}| = O_p\left(\ell^{1/2} \mathrm{r}_n\right)$.

Proof. See Appendix G □

Notice that our results encompass Berk (1974), Bhansali (1978) and Gupta et al. (2013). These references consider the prediction and the estimation of the spectral density of a stationary process with infinite AR representation, by fitting an AR(ℓ) model by least squares, under different assumptions on the process ϵ_t. They also provide results for the consistency of $\hat{\Gamma}_\ell$ and $\hat{\Gamma}_\ell^{-1}$, where $\ell = O(n^\delta)$, $\delta < 1/3$. Lewis and Reinsel (1985), Saikkonen and Lütkepohl (1996) and Lütkepohl and Saikkonen (1997) extended the theory to multivariate time series, whereas Goncalves and Kilian (2007) derive the asymptotic theory when Y_t is a stationary linear process with martingale difference errors ϵ_t that are possibly conditionally heteroscedastic. Goncalves and Kilian (2007) also show the asymptotic validity of the sieve bootstrap under the same assumptions. In our framework, given the factorization $\tilde{\Gamma}_n^{-1} = C_n' D_n C_n$, we can envisage the following bootstrap algorithm. Defining $y = (y_1, \ldots, y_n)'$,

- Compute $e = D_n^{1/2} C_n y$, where e is the vector with elements

$$e_t = \left(y_t - \sum_{j=1}^{t-1} \tilde{\pi}_{t-1,j} y_{t-j} \right) / \sqrt{\tilde{v}_{t-1}}, t = 2, \ldots, n, \quad e_1 = y_1 / \sqrt{\tilde{v}_0}.$$

- Center and sample e with replacement; let e^* denote the resulting bootstrap sample.

- Compute the bootstrap sample $y^* = C_n^{-1} D_n^{-1/2} e^*$, where the generic element is

$$y_t^* = \sum_{j=1}^{t-1} \tilde{\pi}_{t-1,j} y_{t-j}^* + \sqrt{\tilde{v}_{t-1}} e_t^*, \quad y_1^* = \sqrt{\tilde{v}_0} e_1^*.$$

We leave to future research establishing the properties of the above bootstrap algorithm, along with the modifications required in the presence of conditional heteroscedasticity.

5 Estimation of the Banding Parameter

MP propose a data-based selection criterion for the banding parameter ℓ, which is chosen as the smallest $\hat{\ell}$ such that $|\hat{\rho}(\hat{\ell} + k)| < c\sqrt{\log n/n}$, $k = 1, 2, \ldots, K_n$, $K_n = o(\log n)$, where $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$. For the sample sizes typically used in applied work, MP recommend $c = 2$ and $K_n = 5$. The rule amounts to conducting an approximate 95% simultaneous test of $\rho(\hat{l} + k) = 0, k = 1, \ldots, K_n$. See also Politis (2003).

The empirical rule provides an effective and practical criterion which can be extended to our RDL estimator, by picking the smallest value of $\hat{\ell}$ such that

$$|\hat{\phi}_{kk}(\hat{\ell} + k)| < c\sqrt{\log n/n}, \ k = 1, 2, \ldots, K_n, \ K_n = o(\log n). \tag{19}$$

where $\hat{\phi}_{kk}$ is the sample partial autocorrelation computed by the RDL algorithm (7) applied to the sample autocovariances.

Other criteria could be devised. Denoting by $\tilde{v}_{n-1}(\ell)$, the RDL estimated one-step-ahead prediction error variance, as a function of the banding parameter ℓ, the Akaike Information Criterion is $\text{AIC}(\ell) = \ln v_{n-1}(\ell) + 2\sum_{k=0}^{n-1} w_k/n$, where $\sum_{k=0}^{n-1} w_k$ measures the complexity of the predictor based on ℓ.

In principle, the selection of the optimal banding parameter depends on the objective of the analysis and, for instance, what is optimal for short run forecasting may be suboptimal for long range prediction, which may call for a larger ℓ. Naturally, we can apply the MP rule (19) with K_n varying with the forecast lead time. Bickel and Gel (2011) propose a cross-validation method which divides the time series into two consecutive segments of length n_0 (e.g. $n/3$) and n_1, respectively. Given the reduced computational burden of the RDL predictor, which is $O(\ell^2)$, in the applications we will estimate the banding parameter by minimizing the mean square forecast error (MSFE) of the h-step ahead predictor, where h is the desired forecast horizon. The MSFE can be estimated by performing an out-of-sample rolling or recursive forecasting exercise, by dividing the sample into a training set and test set.

6 Simulations and Empirical Illustrations

We explore and compare the performance of the RDL estimator of the autocovariance sequence and the associated linear predictor by simulation and by an empirical case study dealing with the prediction of the sea surface temperature index in the Niño 3.4 region.

6.1 Simulations

We conducted three Monte Carlo simulations to assess the properties of our the RDL estimator of the true autocovariance matrix Γ_n in (1), when the data generating process is AR(1), MA(1), or ARMA(5,5). The main objective is to compare the performance with the banded and tapered estimator (4) by (McMurry and Politis, 2010), referred to as the MP estimator; the overall conclusions can be extended to the class of banded estimator (3) proposed by Wu and Pourahmadi (2009) and Bickel and Gel (2011)

All the reported results are based on 1,000 replications. The selection of the banding parameter is based on the empirical rule proposed by Politis (2003) and used by McMurry and Politis (2010), applied to the sample autocorrelations for the MP estimator and to the sample partial autocorrelations for the RDL estimator; see section 5. We have set $c = 2$ and $K_n = 5$.

The estimation accuracy of $\tilde{\Gamma}_n$ and $\hat{\Gamma}_{n,MP}$ is evaluated by computing the operator norms $\rho(\tilde{\Gamma}_n - \Gamma_n)$ and $\rho(\hat{\Gamma}_{n,MP} - \Gamma_n)$, i.e. the largest eigenvalue of the estimation error matrix, and the infinity norms $||\tilde{\Gamma}_n - \Gamma_n||_\infty$, $||\hat{\Gamma}_{n,MP} - \Gamma_n||_\infty$, i.e. the maximum row sum of the estimation error matrix. See Appendix A.1 for a review of matrix norms.

6.1.1 AR(1)

In the first simulation the data were generated by the autoregressive process $X_t = \phi X_{t-1} + \epsilon_t$, with ϵ_t being an i.i.d. sequence of $N(0, 1 - \phi^2)$ random variables, so that $\gamma(k) = \phi^k, k = 0, 1, \ldots, n - 1$. We consider three sample sizes, $n = 250, 500,$ and 750, and three values of the AR parameter, $\phi = 0.1, 0.5$ and 0.9. Table 1 reports the mean values of the selected banding parameter ℓ, and operator and infinity norms, averaged across 1,000 independent replications, along with their standard deviations. As far as the selection of ℓ is concerned, it is noticeable

from the table that Politis' empirical rule proves very effective in the RDL case in picking up the oracle value $\ell = 1$. For the MP estimator, we confirm the simulation results produced by McMurry and Politis (2010); in particular, the selected ℓ increases with the sample size.

The mean values of the operator and infinity norms illustrate that for $\phi = 0.5$ and $\phi = 0.9$ the RDL estimator performs better, and that the results improve if the sample size increases.

Table 1: Banding parameter and losses in the matrix infinity norm and operator norm for the autoregressive process $Y_t = \phi Y_{t-1} + \epsilon_t, \epsilon_t \sim \mathrm{N}(0, 1 - \phi^2)$.

| | | Regularized Durbin-Levinson | | | | | | McMurry and Politis | | | | | |
| | | ℓ | | Op norm | | ∞-norm | | ℓ | | Op norm | | ∞-norm | |
n	ϕ	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
250	0.1	1.01	0.14	0.26	0.14	0.30	0.16	1.01	0.22	0.18	0.10	0.20	0.11
250	0.5	1.01	0.11	0.67	0.48	0.71	0.47	1.98	0.98	0.84	0.51	0.95	0.48
250	0.9	1.02	0.23	7.79	5.77	8.03	5.93	14.22	8.30	8.97	4.90	10.23	5.50
500	0.1	1.00	0.06	0.19	0.10	0.22	0.11	1.00	0.06	0.13	0.07	0.15	0.07
500	0.5	1.03	0.32	0.48	0.34	0.51	0.33	2.44	1.02	0.61	0.34	0.70	0.32
500	0.9	1.00	0.00	5.76	4.59	5.84	4.59	17.46	8.43	7.18	5.30	8.11	5.60
750	0.1	1.01	0.14	0.15	0.08	0.18	0.09	1.01	0.13	0.11	0.06	0.12	0.07
750	0.5	1.02	0.32	0.37	0.25	0.39	0.24	2.58	0.88	0.49	0.26	0.57	0.25
750	0.9	1.01	0.19	5.02	3.86	5.07	3.84	20.92	10.79	6.64	4.90	7.63	5.41

6.1.2 MA(1)

When the data were generated by the moving average process $Y_t = \epsilon_t + \theta\epsilon_{t-1}$, with ϵ_t being an i.i.d. sequence of $N(0, 1)$ random variables, the true Γ_n is a banded matrix with $\gamma(k) = 0, k > 1$ and partial autocorrelation function is $\phi_{kk} = -(-\theta)^k(1 - \theta^2)/(1 - \theta^{2(k+1)})$. For each simulation θ assumes the values $0.1, 0.5, 0.9$ and we consider three sample sizes, $n = 250, 500, 750$. The results are reported in Table 2. As it is evident from the table, the MP estimator has the best performances: the empirical rule tends to select the oracle $\ell = 1$ and the accuracy measures are systematically better.

Table 2: Banding parameter and losses in the matrix infinity norm and operator norm for the moving average process $Y_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\epsilon_t \sim \mathrm{N}(0, 1)$.

| | | Regularized Durbin-Levinson | | | | | | McMurry and Politis | | | | | |
| | | ℓ | | Op norm | | ∞-norm | | ℓ | | Op norm | | ∞-norm | |
n	θ	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
250	0.1	1.03	0.40	0.27	0.15	0.33	0.19	1.04	0.42	0.19	0.12	0.20	0.15
250	0.5	1.58	0.70	0.52	0.30	0.70	0.38	1.08	0.57	0.26	0.23	0.27	0.29
250	0.9	4.25	1.53	1.52	0.96	2.24	1.28	1.15	0.79	0.48	0.50	0.52	0.59
500	0.1	1.00	0.00	0.18	0.09	0.22	0.10	1.01	0.13	0.12	0.07	0.12	0.07
500	0.5	2.05	0.62	0.42	0.23	0.55	0.27	1.06	0.43	0.20	0.18	0.20	0.20
500	0.9	5.94	1.69	1.21	0.57	1.88	0.82	1.09	0.57	0.31	0.31	0.33	0.34
750	0.1	1.00	0.00	0.15	0.07	0.18	0.08	1.00	0.00	0.12	0.06	0.12	0.06
750	0.5	2.10	0.94	0.39	0.20	0.55	0.28	1.00	0.00	0.20	0.10	0.20	0.10
750	0.9	6.50	1.50	1.85	1.39	2.56	1.44	1.60	1.80	0.49	0.64	0.58	0.87

6.1.3 ARMA(5,5)

In our third Monte Carlo experiment 1,000 series of length $n = 250, 500$, and 750 were generated according to the ARMA(5,5) process,

$$Y_t = \frac{1 - 0.8762B + 0.0184B^2 + 0.0197B^3 + 0.8591B^4 - 0.7491B^5}{1 - 0.6281B + 0.3597B^2 + 0.2634B^3 - 0.5322B^4 + 0.7900B^5}\epsilon_t, \epsilon_t \sim \text{i.i.d. N}(0,1)$$

This process is considered by Bickel and Gel (2011), section 5.1. Its spectral density is characterized by a very high dynamic range, i.e. a large excursion between $\max_\omega f(\omega)$ and $\min_\omega f(\omega)$, and by the presence of two distinctive spectral peaks at $\pi/2$ and π. The first 8 autocorrelations are $\rho(1) = -0.15, \rho(2) = -0.60, \rho(3) = -0.18, \rho(4) = 0.89, \rho(5) = -0.16, \rho(6) = -0.57, \rho(7) = -0.16, \rho(8) = 0.76$; moreover, $\rho(20) = 0.48, \rho(40) = 0.22, \rho(60) = 0.11$.

Table 3 reports the mean and the standard deviation of the selected banding parameter and the operator and infinity norm over the 1,000 independent replications. It shows that the RDL estimator selects a smaller banding parameter and is characterized by a greater accuracy in estimating the true Γ_n.

When coupled together with the evidence arising from the AR(1) experiment, we can conclude that banding and tapering the partial autocorrelation function is more effective than banding and tapering the autocovariance function, when the true generating process has high dynamic range, resulting from the presence of cyclical components and strong autoregressive components.

Table 3: Banding parameter and losses in the matrix infinity norm and operator norm for the ARMA(5,5) process considered by Bickel and Gel (2011).

| | RDL Empirical Rule | | | | | | McMurry and Politis | | | | | |
| | ℓ | | Op norm | | ∞-norm | | ℓ | | Op norm | | ∞-norm | |
n	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
250	10.02	2.65	170.63	124.06	252.73	174.69	56.05	37.04	189.05	141.13	294.10	208.61
500	15.66	3.17	142.00	98.65	208.61	135.51	74.97	52.40	179.02	152.83	285.17	237.04
750	18.55	3.28	128.74	79.78	182.46	103.37	80.79	59.83	172.84	147.24	270.50	238.97

6.2 Forecasting Niño 3.4 sea surface temperature

Our illustrative example deals with the ability to forecast the sea surface temperatures (SST) in the Niño 3.4 region. This time series has been analyzed by Bickel and Gel (2011) and provides an interesting case study, since its cyclical behaviour is a manifestation of the El Niño phenomenon, which determines an increase in SST in the eastern and central Pacific regions. The monthly time series, available at https://www.esrl.noaa.gov/psd/data/climateindices/list) for the period January 1950 - December 2016, is plotted in the top left panel of figure 1. The prediction of SST are an important input of global circulation models and are relevant for the prediction of El Niño Southern Oscillation (ENSO) events, which are the most dominant forcing factors of inter-annual climate variability. SST is also the most important surface condition affecting climate at longer run horizons.

The evaluation of the forecasting ability of the RDL estimator, in comparison with the MP estimator and traditional ARMA predictors, is carried out by a rolling forecast experiment such that, starting from January 1970, we compute the RDL and MP h-step-ahead predictors, $h = 1, 2, \ldots, 48$, for values of the banding parameters $\ell = 1, 2, \ldots, 48$, using a training sample of

dimension $n_0 = 288$ observations (20 years of monthly data); we proceed by adding one future observation and removing the initial one, until the end of the sample is reached, so that each forecast is based on a fixed number of observations, and re-estimating the coefficients of the linear predictors for each rolling window. The experiment yields $H = 564 - h + 1$ h-step-ahead prediction errors for each forecasting methodology. The predictive performances of the RDL and MP estimator depends on the forecast horizon h and the banding parameter ℓ; they are compared to the h-step ahead predictions arising from fitting ARMA(p, q) models, with $p \leq 13$ and $q \leq 2$.

Denoting by $\tilde{y}_{t|t-h}$ a generic prediction of y_t, for each forecast lead h and banding parameter ℓ we compute the mean square forecast error, $MSFE(h, \ell) = \frac{1}{H} \sum_{t=n_0+h}^{n} (y_t - \tilde{y}_{t|t-h})^2$, and the following measure of predictability:

$$P(h, \ell) = \max\left(0, 1 - \frac{MSFE(h, \ell)}{\hat{\gamma}(0)}\right), \quad h, \ell = 1, \dots 48.$$

Figure 2 is a contour plot of $P(h, \ell)$ for the RDL predictor, whereas figure 1b compares its maximum across ℓ, $P(h) = \max_{\ell=1,\dots,48} P(h, \ell)$, $h = 1, \dots 48$, with that characterizing the MP and ARMA predictor.

The value of ℓ that is optimal for different horizons, in the sense that it minimizes the MSFE, grows rapidly from 10 to values in the range (28,33). The empirical rule of section 5 selects a value around 10 with some variability across the rolling samples; as it can be seen from figure 2, choosing a small ℓ is detrimental to predictability at forecast lead times greater than 6. For MP, the optimal bandwidth goes from 42 to 48. For the ARMA models the $P(h)$ measure is defined as the maximum predictability across the AR and MA orders; the selected specifications in the majority of cases are AR(12) and ARMA(12,1).

The main evidence is that the predictability of SST is high at inter-annual horizons (up to lead time 6), and it decreases quite rapidly. However, there is some predictability at longer horizons. Secondly, the RDL predictor outperforms the MP and ARMA predictors. With respect to the latter, the performance is about the same for $h = 1$. Also, the MP outperforms the ARMA predictors for horizons $h > 6$. The superior performance with respect to the MP predictor is due to the fact that SST are characterized by the presence of strong cyclical components, that are better captured by regularizing the partial autocorrelations rather than the sample autocovariances.

Figure 1c displays the sample partial autocorrelation function and the regularized partial autocorrelations for $\ell = 28$; the estimated log-spectral density is plotted along the log-periodogram of the series in figure 1d. The two sharp peaks at occur at the annual and semiannual frequencies, $\pi/6$ and $\pi/3$, and are expression of the seasonal cycle in SST. The peak close to the origin is expression of the alternation of El Niño and La Niña episodes, with a periodicity of 4 years. Our RDL seems to capture the main stylized facts concerning the SST series quite well.

This illustration also shows that the selection of the banding parameter can be related to the forecast horizon that is relevant to the investigator. It can be determined by evaluating the forecasting performance of the predictor as a function of ℓ.

7 Conclusions

We have proposed an estimator of the autocovariance function of a random process based on a modification of the Durbin-Levinson algorithm, which regularizes the partial autocorrelation function, shrinking it towards zero according to a weighting function. In particular, we have

focused on the trapezoidal kernel, see Politis and Romano (1995) and McMurry and Politis (2010). We have shown the consistency of the regularized Durbin-Levinson autocovariance matrix estimator and the associated predictor and have derived its rate of convergence. The simulations and the empirical illustration have shown that the RDL method is very effective and more efficient than available tapered estimators for processes with high spectral dynamic range (cyclical processes). Due to its reduced computational complexity, the optimal bandwidth of the tapering function can be selected by crossvalidation.

(a) Time series

(b) Predictability $P(h)$
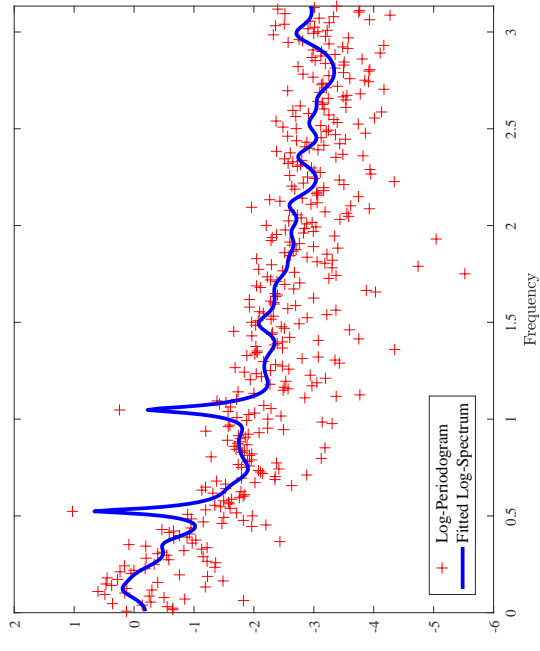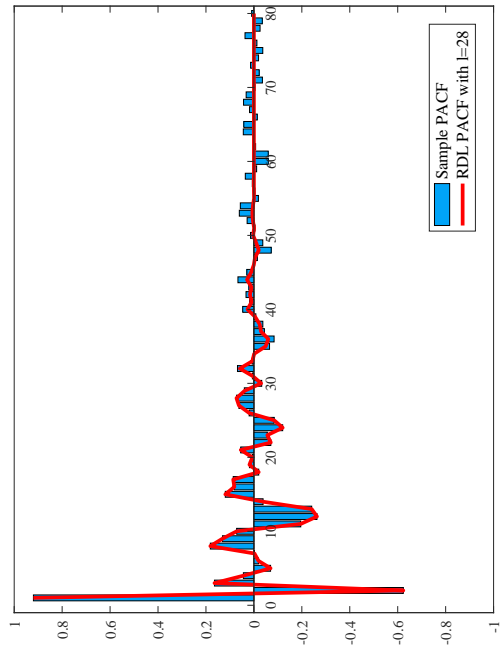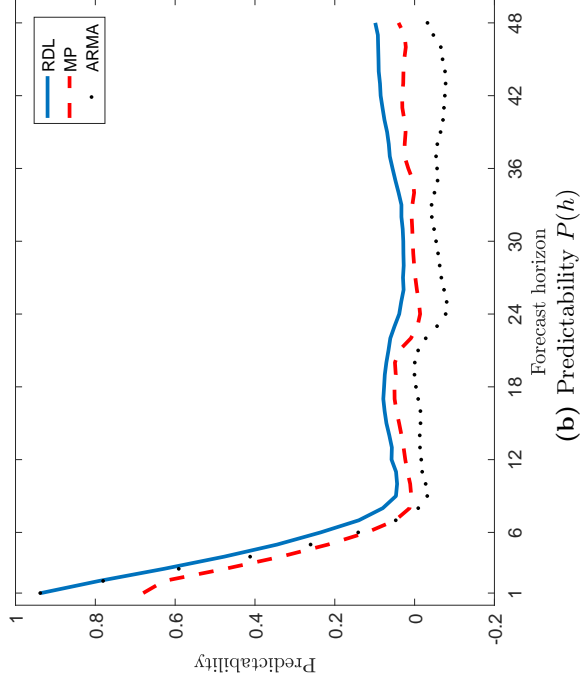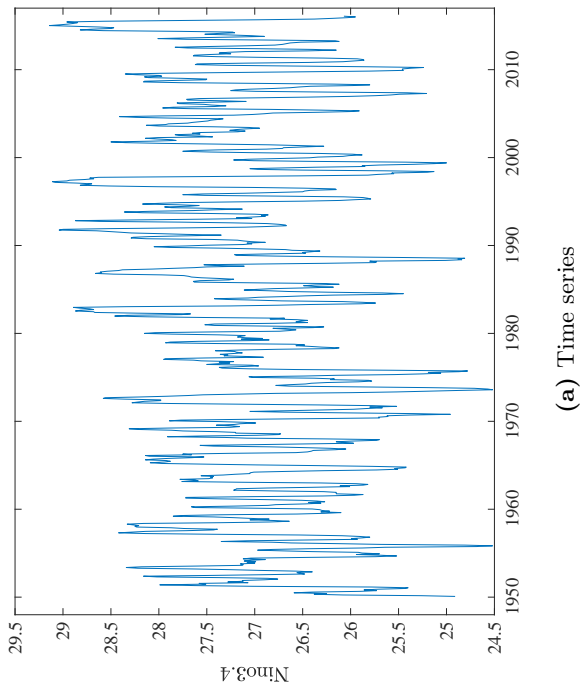
(c) Raw and regularized partial autocorrelations

(d) Raw and regularized sample spectrum

Figure 1: Niño 3.4 Sea Surface Temperature Series, 1950.1-2016.12 (Source: https://www.esrl.noaa.gov/psd/data/climateindices/list)

Figure 2: Niño 3.4 Sea Surface Temperature Series. RDL predictor: contour plot of predictability at forecast horizon h using banding parameter l, $P(h, \ell) = \max\left(0, 1 - \frac{MSFE(h, \ell)}{\hat{\gamma}(0)}\right)$, for $h = 1, \dots, 48$ and $\ell = 1, \dots, 48$.
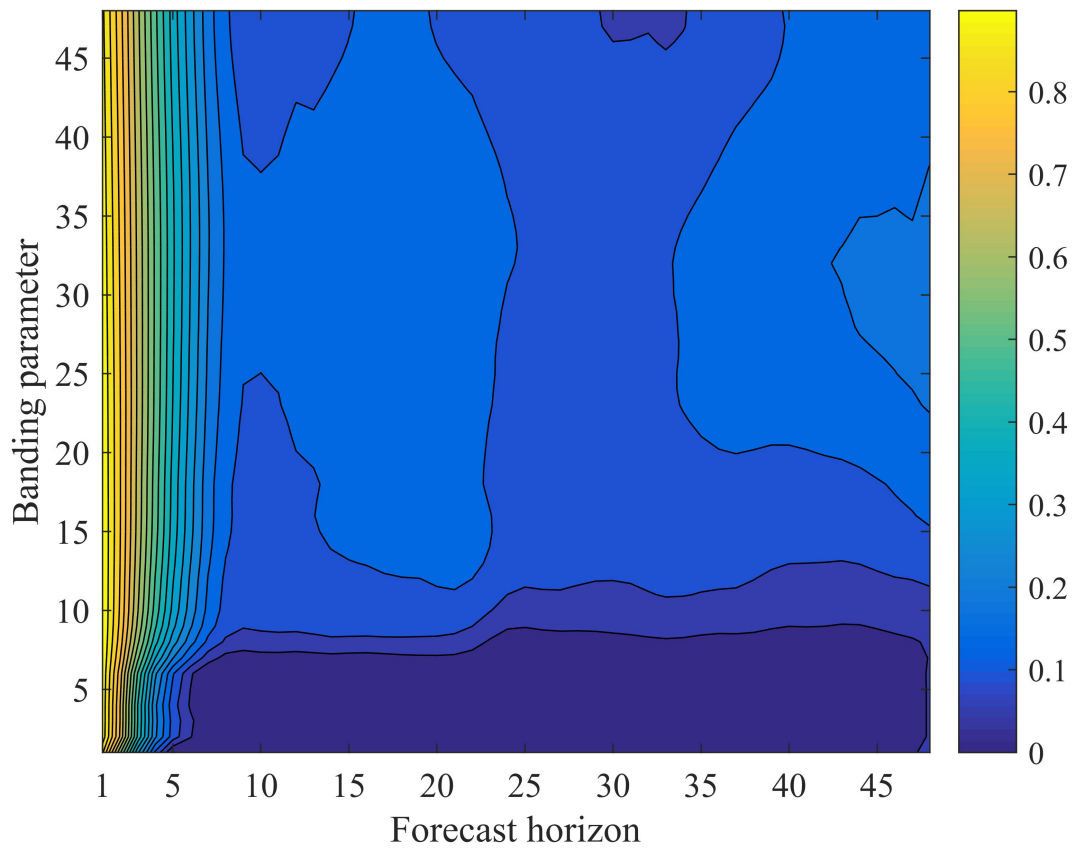
Appendix

A   Preliminaries

A.1   Vector and matrix norms

If x denotes an $m \times 1$ vector, the vector p-norm is defined as $||x||_p = (\sum_{i=1}^m x_i^p)^{1/p}$. For an $m \times n$ matrix A, $||\,A\,||_p = \sup_{x \neq 0} \frac{||Ax||_p}{||x||_p}$ The 2-norm $||A||_2 = \sqrt{\lambda_{max}(A'A)}$, the square root of the maximum eigenvalue of $A'A$, is also known as the operator norm and it will be denoted by $\rho(A) = ||A||_2$. If A is $n \times n$ and symmetric, then $||A||_2$ is the maximum eigenvalue of A. In the proofs we will make use of the following result, implied by corollary 2.3.2 in Golub and Van Loan (2012): if $A = A'$,

$$||A||_2 \leq \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|,$$

i.e. the operator norm is bounded by the maximum column sum of A (this is also known as the 1-norm, $||A||_1$). The operator norm has the following properties: let c be a scalar, B a conformable matrix and b an $n \times 1$ vector. Then $\rho(cA) = |c|\rho(A), \rho(A + B) \leq \rho(A) + \rho(B), \rho(AB) \leq \rho(A)\rho(B), \rho(Ab) \leq \rho(A)||b||_2$. The Frobenius 2-norm is defined as $||A||_F = \sqrt{\text{trace}(A'A)} = \sum_i \sum_j a_{ij}^2$. In the illustrations we also refer to the infinity norm, $||A||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$, the maximum row sum of A. If A is symmetric, $||A||_\infty = ||A||_1$.

A.2   Consistency of the sample autocovariances

Let $\hat{c}(k) = \frac{1}{n-k} \sum_{t=k+1}^n y_t y_{t-k}$, $\hat{c}(k) = n\hat{\gamma}(k)/(n-k)$. A fundamental result that will be used in the proofs is the following upper bound for the mean square estimation error of the lag k autocovariance, based on a sample of size $n > k$:

$$\mathrm{E}[(\hat{c}(k) - \gamma(k))^2] \leq \frac{C^*}{n-k}.$$

The above result is due to E. J. Hannan (1960) (page 39, under the i.i.d. assumption for ϵ_t), and E. Hannan et al. (1972), where the constant $C^* = 2\sum_{r=-\infty}^\infty \gamma^2(r) + \kappa_4(\sum_j \psi_j^2)^2$, where κ_4 is the 4-th cumulant of ϵ_t ($C < \infty$ is implied by the assumptions); an analogous result is implied by Lemma 1 in Wu and Pourahmadi (2009) (taking $\alpha = 4$), where the constant C^* depends on $[E(Y_t^4)]^{1/4}$ and the physical dependence measure of the process; see Wu (2011) for details. The same result is obtained by applying the Marcinkiewicz-Zygmund inequality, theorem 4.1 in Dedecker et al. (2007); see also Bickel and Gel (2011).

Using $\hat{\gamma}(k) = \frac{n-k}{n}\hat{c}(k)$, or equivalently $\hat{\gamma}(k) - \gamma(k) = \frac{n-k}{n}(\hat{c}(k) - \gamma(k)) - \frac{k}{n}\gamma(k)$, we have

$$|\hat{\gamma}(k) - \gamma(k)| \leq \frac{n-k}{n}|\hat{c}(k) - \gamma(k)| + \frac{k}{n}|\gamma(k)|.$$

Hence, for a suitable constant C, we have the following bound for the mean absolute error:

$$\mathrm{E}|\hat{\gamma}(k) - \gamma(k)| \leq \frac{C}{n^{1/2}} + \frac{k}{n}|\gamma(k)|. \tag{20}$$

B   Proof of Lemma 1

The positive definiteness of the regularized autocovariances can be proved by first noticing that the assumptions on $\{w_k, k = 0, 1, \ldots\}$ imply that $|\pi_{kk}| < 1$ and that $v_{k-1} = v_0 \prod_{j=1}^{k-1}(1 - \pi_{jj}^2)$ is equal to $\gamma_r(0) - \sum_{j=1}^{k-1} \pi_{k-1,j} \gamma_r(j)$. Let $\Gamma_{r,k}$ denote the Toeplitz matrix formed with $\{\gamma_r(0), \ldots, \gamma_r(k)\}$. Then, by a suitable partitioning of $\Gamma_{r,k}$,

$$
\begin{aligned}
|\Gamma_{r,k}| &= \left[\gamma_r(0) - \sum_{j=1}^{k-1} \pi_{k-1,j} \gamma_r(j) \right] |\Gamma_{r,k-1}| \\
&= v_{k-1} |\Gamma_{r,k-1}| \\
&= \prod_{j=0}^{k-1} v_j \\
&= v_0^k \prod_{j=1}^{k-1}(1 - \pi_{jj}^2)^{k-j}.
\end{aligned}
$$

Hence, $|\Gamma_{r,k}| > 0$, as $|\pi_{jj}| < 1$ and $v_0 > 0$.

The proof of statement (ii) is by induction. For $k = 1$, $\pi_{11} = w_1 \gamma(1)/\gamma(0)$ is such that $|\pi_{11}| < 1$, and thus the root of $\Pi_1(z) = 1 - \pi_{11}z$, $z = \pi_{11}^{-1}$, is greater than 1 in modulus. Assume now that $\Pi_{k-1}(z) \neq 0 \iff |z| \leq 1$, then $\Pi_k(z)$ has all its roots greater than 1 in modulus iff $|\pi_{kk}| < 1$: as a matter of fact, according to (7), we can write $\Pi_k(z) = \Pi_{k-1}(z) - \pi_{kk} z^k \Pi_{k-1}(1/z)$, so that $\Pi_k(z) = 0$ for all $\{z \in \mathbb{C} : \Pi_{k-1}(z) - \pi_{kk} z^k \Pi_{k-1}(1/z) = 0\}$. Hence, $\Pi_k(z) = 0 \iff z = \left[\frac{1}{\pi_{kk}} \frac{\Pi_{k-1}(1/z)}{\Pi_{k-1}(z)} \right]^{1/k}$ and, since $\frac{\Pi_{k-1}(1/z)}{\Pi_{k-1}(z)}$ has unit modulus, $|z| > 1 \iff |\pi_{kk}| < 1$. Iterating this result, we can conclude that $\Pi_k(z)$ has stationary roots iff $|\pi_{jj}| < 1, j = 1, 2, \ldots, k$. That $|\pi_{jj}| < 1$ follows directly from the assumptions on $\{w_j\}$ and from $|\phi_{jj}| < 1$.

To prove statement (iii), let us consider a generic k. According to the RDL algorithm (7), for $j < k$, $\pi_{kj} - \phi_{kj} = \pi_{k-1,j} - \phi_{k-1,j} - \pi_{kk}\pi_{k-1,k-j} + \phi_{kk}\phi_{k-1,k-j}$. If we take $w_k = 0$ (and thus $\pi_{kk} = 0$), the following inequality holds: $|\pi_{kj} - \phi_{kj}| \leq |\pi_{k-1,j} - \phi_{k-1,j}| + |\phi_{kk}||\phi_{k-1,k-j}|$. Summing from $j = 1$ to $j = k - 1$ and adding $|\phi_{kk}|$ to both sides, we have that (for $w_k = 0$):

$$
\sum_{j=1}^{k} |\pi_{kj} - \phi_{kj}| \leq \sum_{j=1}^{k-1} |\pi_{k-1,j} - \phi_{k-1,j}| + |\phi_{kk}| \left(1 + \sum_{j=1}^{k-1} |\phi_{k-1,j}| \right). \tag{21}
$$

Consider now the AR(ℓ) approximation to the process, which is the one having $w_k = 1$ for $0 \leq k \leq \ell$ and $w_k = 0, \ell < k \leq n-1$, and let us denote by π_{kj}^* the coefficients resulting from the RDL recursions (7), applying the above uniform weights truncated at ℓ. Since $\sum_{j=1}^{k} |\pi_{kj}^* - \phi_{kj}| = 0$ for $k \leq \ell$, the recursion (21) yields

$$
\sum_{j=1}^{n-1} |\pi_{n-1,j}^* - \phi_{n-1,j}| \leq \sum_{j=\ell+1}^{n-1} |\phi_{jj}| \left(1 + \sum_{i=1}^{j-1} |\phi_{j-1,i}| \right).
$$

The same reasoning leads to the following inequality:

$$
\sum_{j=1}^{n-1} |\pi_{n-1,j}^* - \pi_{n-1,j}| \leq \sum_{j=\ell+1}^{L} |\pi_{jj}| \left(1 + \sum_{i=1}^{j-1} |\pi_{j-1,i}| \right),
$$

where $\pi_{kj}, k, j = 1, \ldots, n-1$, results from applying a trapezoidal kernel ($L = \lfloor 2\ell \rfloor$). Hence, by the triangle inequality,

$$
\begin{aligned}
\sum_{j=1}^{n-1} |\pi_{n-1,j} - \phi_{n-1,j}| &\leq \sum_{j=1}^{n-1} |\pi_{n-1,j} - \pi_{n-1,j}^*| + \sum_{j=1}^{n-1} |\pi_{n-1,j}^* - \phi_{n-1,j}| \\
&\leq \sum_{j=\ell+1}^{n-1} |\phi_{jj}| \left(1 + \sum_{i=1}^{j-1} |\phi_{j-1,i}| \right) + \sum_{j=\ell+1}^{L} |\pi_{jj}| \left(1 + \sum_{i=1}^{j-1} |\pi_{j-1,i}| \right),
\end{aligned}
$$

and by Lemma 3 in Appendix D both terms are bounded by $\sum_{j=\ell+1}^{n-1} |\phi_{jj}| \prod_{i=1}^{j-1} (1 + |\phi_{ii}|)$, giving (8).

C   Proof of Lemma 2

The proof of statement (i) is based on Baxter's Theorem (Baxter (1962), see also (Bingham, 2012), p. 302), extended by Inoue (2008), according to which the stated assumptions imply the summability the partial autocorrelations ϕ_{kk}, $\sum_j |\phi_{jj}| < \infty$, and therefore $\sum_j |\pi_{jj}| < \infty$; morever, the AR coefficients ϕ_j are also absolutely summable by Theorem 3.8.4 in Brillinger (1981).

By Baxter inequality, $\sum_{j=1}^{n-1} |\phi_{n-1,j} - \phi_j| \le C \sum_{j=n}^{\infty} |\phi_j|$, for some positive constant C. The right hand side can be expressed in terms of the partial autocorrelations: in fact, applying the same argument used in the proof of lemma 1, we have that for $s > r$, setting $\phi_{rj} = 0$ for $j > r$, $\sum_{j=1}^{s} |\phi_{rj} - \phi_{sj}| \le \sum_{j=r+1}^{s} |\phi_{jj}| \prod_{i=1}^{j-1} (1 + |\phi_{ii}|) \le \prod_{i=1}^{\infty} (1 + |\phi_{ii}|) \sum_{j=r}^{s} |\phi_{jj}|$; letting $r = n-1$, and $s \to \infty$, we get

$$\sum_{j=1}^{n-1} |\phi_{n-1,j} - \phi_j| \le C^* \sum_{j=n}^{\infty} |\phi_{jj}|, \tag{22}$$

where $C^* = \prod_{i=1}^{\infty} (1 + |\phi_{ii}|) < \infty$ as $\sum_j |\phi_{jj}| < \infty$. Expression (22) is an alternative effective way of expressing Baxter's inequality in terms of the partial autocorrelations of the process.

By the triangle inequality

$$
\begin{array}{rcl}
\sum_{j=1}^{n-1} |\pi_{n-1,j} - \phi_j| & \le & \sum_{j=\ell+1}^{n-1} |\phi_{jj}| \prod_{i=1}^{j-1} (1 + |\phi_{ii}|) + C \sum_{j=n}^{\infty} |\phi_j| \\
& \le & \sum_{j=\ell+1}^{n-1} |\phi_{jj}| \prod_{i=1}^{\infty} (1 + |\phi_{ii}|) + C^* \sum_{j=n}^{\infty} |\phi_{jj}| \\
& \le & C^* \sum_{j=\ell+1}^{\infty} |\phi_{jj}|.
\end{array}
$$

The bound (10) follows from $v_{n-1} = \gamma(0) \prod_{j=1}^{n-1} (1 - \pi_{jj}^2)$, $\sigma^2 = \gamma(0) \prod_{j=1}^{\infty} (1 - \phi_{jj}^2)$, $\pi_{jj} = \phi_{jj}, j \le \ell$ and $-\ln(1 - x) < x/(1 - x), x < 1$.

The proof of statement (11) is based on the fact that $\sum_{j=1}^{\infty} j^r |\gamma(j)|$ implies $\sum_{j=1}^{\infty} j^r |\phi_{jj}| < \infty$, by Theorem 2.3 in Baxter (1962). This in turn implies that $\lim_{n\to\infty} \sum_{j=1}^{\infty} j^r |\phi_{n-1,j}| < \infty$ (and $\sum_{j=1}^{\infty} j^r |\phi_j| < \infty$), as we are going to prove. Starting from the inequality $j^r |\phi_{kj}| \le j^r |\phi_{k-1,j}| + j^r |\phi_{kk}||\phi_{k-1,k-j}|$ for $j < k$, as implied by (7), summing for $j = 1, \ldots, k-1$, and adding $k^r |\phi_{kk}|$ to both sides, yields the recursion

$$
\begin{array}{rcl}
\sum_{j=1}^{k} j^r |\phi_{kj}| & \le & \sum_{j=1}^{k-1} j^r |\phi_{k-1,j}| + k^r |\phi_{kk}| + |\phi_{kk}| \sum_{j=1}^{k-1} j^r |\phi_{k-1,k-j}| \\
& \le & \sum_{j=1}^{k-1} j^r |\phi_{k-1,j}| + k^r |\phi_{kk}| + |\phi_{kk}| \sum_{j=1}^{k-1} (k-j-1)^r |\phi_{k-1,j}| \\
& \le & \sum_{j=1}^{k-1} j^r |\phi_{k-1,j}| + k^r |\phi_{kk}| + k^r |\phi_{kk}| \sum_{j=1}^{k-1} |\phi_{k-1,j}| \\
& \le & \sum_{j=1}^{k-1} j^r |\phi_{k-1,j}| + k^r |\phi_{kk}| \left(1 + \sum_{j=1}^{k-1} |\phi_{k-1,j}|\right) \\
& \le & \sum_{j=1}^{k-1} j^r |\phi_{k-1,j}| + C^* k^r |\phi_{kk}|,
\end{array}
$$

where $C^* = \prod_{j=1}^{\infty} (1 + |\phi_{jj}|)$ is a finite constant (as implied by the summability of the partial autocorrelation); the last line follows from Lemma 3 of Appendix D. The above recursion implies $\sum_{j=1}^{k} j^r |\phi_{kj}| \le C^* \sum_{j=1}^{k} j^r |\phi_{jj}|$, and thus

$$\lim_{n\to\infty} \sum_{j=1}^{n-1} j^r |\phi_{n-1,j}| = \sum_{j=1}^{\infty} j^r |\phi_j| < \infty.$$

Applying these results to the tapered DL process, the PACF is power summable as $\sum_j j^r |\pi_{jj}| < \sum_j j^r |\phi_{jj}|$, which in turns implies that the AR coefficients $\pi_{n-1,j}$ are power summable. This in turn implies, by Theorem 3.8.4 in Brillinger (1981) and Lemma 2.2. in Bühlmann (1995), that the tapered autocovariances $\gamma_r(j)$ are power summable.

D   Lemma 3

Lemma 3. *Let $\phi_{kj}, j = 1, \ldots, k$, be generated according to the DL recursions (7). Then*

$$1 + \sum_{j=1}^{k} |\phi_{kj}| \leq \prod_{j=1}^{k} (1 + |\phi_{jj}|), \tag{23}$$

and

$$1 - \sum_{j=1}^{k} \phi_{kj} = \prod_{j=1}^{k} (1 - \phi_{jj}). \tag{24}$$

Proof. By (7), $|\phi_{kj}| \leq |\phi_{k-1,j}| + |\phi_{kk}||\phi_{k-1,k-j}|$ for $j < k$. Adding for $j = 1, \ldots, k-1$ $\sum_{j=1}^{k-1} |\phi_{kj}| \leq (1 + |\phi_{kk}|) \sum_{j=1}^{k-1} |\phi_{k-1,j}|$ and adding $(1 + |\phi_{kk}|)$ to both sides yields the recursion

$$1 + \sum_{j=1}^{k} |\phi_{kj}| \leq (1 + |\phi_{kk}|) \left(1 + \sum_{j=1}^{k-1} |\phi_{k-1,j}| \right).$$

Also, summing both sides of $\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$ with respect to j, adding ϕ_{kk} to both sides and subtracting from 1, gives the recursion:

$$1 - \sum_{j=1}^{k} \phi_{kj} = (1 - \phi_{kk}) \left(1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \right),$$

which can be solved to give (24). $\qquad \square$

E   Proof of Theorem 1

From Corollary 2.3.2 in (Golub and Van Loan, 2012) (see also Appendix A) and the Toeplitz nature of the estimation error matrix, it follows that

$$
\begin{aligned}
\rho(\tilde{\Gamma}_n - \Gamma_n) &\leq \max_{1 \leq j \leq n} \sum_{i=1}^{n} |\tilde{\gamma}_r(|i-j|) - \gamma(|i-j|)| \\
&\leq \sum_{j=-(n-1)}^{n-1} |\tilde{\gamma}_r(j) - \gamma(j)| \\
&\leq 2 \sum_{j=0}^{n-1} |\tilde{\gamma}_r(j) - \gamma(j)| \\
&\leq 2 \sum_{j=0}^{n-1} |\tilde{\gamma}_r(j) - \gamma_r(j)| + 2 \sum_{j=0}^{n-1} |\gamma_r(j) - \gamma(j)| \\
&\leq T_1 + T_2 \\
T_1 &= 2 \sum_{j=0}^{n-1} |\tilde{\gamma}_r(j) - \gamma_r(j)| \\
T_2 &= 2 \sum_{j=0}^{n-1} |\gamma_r(j) - \gamma(j)|
\end{aligned}
$$

Exploiting the notion that $\tilde{\gamma}_r(j)$ is the sample autocovariance function of the process with theoretical autocovariance function $\gamma_r(j)$, we can apply the distributional results of appendix A.2 to the regularized sample autocovariances ($\tilde{\gamma}_r(j) = \hat{\gamma}(j)$ for $|j| \leq \ell$), to characterize it as an estimator of $\gamma_r(j)$: recalling that $L = \lfloor 2\ell \rfloor$.

$$
\begin{aligned}
T_1 &= 2 \sum_{j=0}^{L} |\tilde{\gamma}_r(j) - \gamma_r(j)| + 2 \sum_{j=L+1}^{p_n} |\tilde{\gamma}_r(j) - \gamma_r(j)| + 2 \sum_{j=p_n+1}^{n-1} |\tilde{\gamma}_r(j) - \gamma_r(j)| \\
&= T_{11} + T_{12} + T_{13}
\end{aligned}
$$

Using Appendix A.2 and the fact that $\gamma_r(j) = \gamma(j)$ for $j \leq \ell$, we have

$$E|T_{11}| \leq [E(T_1^2)]^{1/2} \leq C \frac{L+1}{\sqrt{n}} + \frac{2}{n} \sum_{j=1}^{L+1} j|\gamma(j)| + \frac{2}{n} \sum_{j=\ell+1}^{L+1} j|\gamma_r(j)|.$$

and thus, $T_{11} = O_p(\mathbf{r}_n)$.

For $j > L$ the RDL sample and theoretical autocovariance functions are generated according to the following homogeneous difference equations of order L:

$$\tilde{\gamma}_r(j) = \sum_{k=1}^{L} \tilde{\pi}_{Lk} \tilde{\gamma}_r(j-k), \gamma_r(j) = \sum_{k=1}^{L} \pi_{Lk} \tilde{\gamma}_r(j-k).$$

Hence, for $j > L$,

$$\begin{aligned}
|\tilde{\gamma}_r(j) - \gamma_r(j)| &\leq \sum_{k=1}^{L} |\tilde{\pi}_{Lk}||\tilde{\gamma}_r(j-k) - \gamma_r(j-l)| + \sum_{k=1}^{L} |\gamma_r(j-k)||\tilde{\pi}_{Lk} - \pi_{Lk}| \\
&\leq \sum_{k=1}^{L} |\tilde{\pi}_{Lk}| \sum_{k=1}^{L} |\tilde{\gamma}_r(j-k) - \gamma_r(j-l)| + \sum_{k=1}^{L} |\gamma_r(j-k)| \sum_{k=1}^{L} |\tilde{\pi}_{Lk} - \pi_{Lk}| \\
&= O_p(\mathbf{r}_n),
\end{aligned}$$

since $\sum_{k=1}^{L} |\tilde{\pi}_{Lk}|$ and $\sum_{k=1}^{L} |\gamma_r(j-k)|$ are bounded for $L \to \infty$. As a result, $T_{12} = O_p(p_n^{1/2}\mathbf{r}_n)$.

As far as the component T_{13} is concerned, the rate $p_n > L$ is chosen according to Assumption 2 such that $\sum_{p_n+1}^{n-1} |\tilde{\gamma}_r(j)| = o(1)$ and $\sum_{p_n+1}^{n-1} |\gamma_r(j)| = o(1)$, so that we can write

$$T_{13} = \sum_{j=L+1}^{p_n} |\tilde{\gamma}_r(j) - \gamma_r(j)| + o(1).$$

As a matter of fact, notice that by Brockwell and Davis (1991), page 94, $\gamma_r(j)$ follows a homogeneous linear difference equation with general solution $\gamma_r(j) = \sum_{i=1}^{k} \sum_{h=0}^{m_i-1} \beta_{ih} \rho_i^j j^h$, where $\rho_i, i = 1, \ldots, k$, are the reciprocals of the roots of the polynomial $\pi_L(z)$, which by Lemma 1 (ii) satisfy $|\rho_i| < 1$, m_i is the multiplicity of the i-th root, $\sum_i m_i = L$, and the coefficients β_{ih} are determined from the initial L autocovariances. Hence, denoting by $\varrho = \max_{1 \leq i \leq k} |\rho_i|, 0 < \varrho < 1$, $m_L = \max_{1 \leq i \leq k} m_i, m_L \leq L, |\gamma_r(j)| \leq C_L \varrho^j j^{m_L}$, so that the autocovariances tend to zero at a geometric rate and $\sum_{p_n+1}^{n-1} |\gamma_r(j)| = o(1)$. An analogous results holds for $\tilde{\gamma}_r(j)$.

In conclusion, $T_1 = O_p(p_n^{1/2}\mathbf{r}_n)$. For the component T_2,

$$\begin{aligned}
T_2 &= 2 \sum_{j=0}^{n-1} |\gamma_r(j) - \gamma(j)| \\
&= 2 \sum_{j=\ell+1}^{n-1} |\gamma_r(j) - \gamma(j)| \\
&\leq 2 \sum_{j=\ell+1}^{n-1} |\gamma_r(j)| + 2 \sum_{j=\ell+1}^{L} |\gamma(j)| \\
&\leq C_2 \sum_{j=\ell+1}^{n-1} |\gamma_r(j)|,
\end{aligned}$$

where $C_2 > 0$ is a suitable constant. Lemma 1 implies that $T_2 = o(1)$ under assumptions 1 and 2.

Hence, $\tilde{\Gamma}_n$ converges to Γ_n in operator norm under the stated assumptions and its rate of convergence is $p_n^{1/2}\mathbf{r}_n$.

Finally, to prove (16) we write $\tilde{\Gamma}_n^{-1} - \Gamma_n^{-1} = \tilde{\Gamma}_n^{-1} \left(\Gamma_n - \tilde{\Gamma}_n \right) \Gamma_n^{-1}$. By the properties of the operator norm (see Appendix A.1), $\rho(\tilde{\Gamma}_n^{-1} - \Gamma_n^{-1}) \leq \rho(\tilde{\Gamma}_n^{-1})\rho\left(\Gamma_n - \tilde{\Gamma}_n \right) \rho(\Gamma_n^{-1})$. The result follows from (15), and the fact that both $\rho(\Gamma_n^{-1})$ and $\rho(\tilde{\Gamma}_n^{-1})$ are bounded. See Remark 4 for the former; the latter results from the positive definiteness of the sample RDL autocovariance sequence (Lemma 1), which guarantees that the smallest eigenvalue of $\tilde{\Gamma}_n$ is strictly positive.

F   Proof of Theorem 2

Let $\tilde{\pi}_{(n-1)} = (\tilde{\pi}_{n-1,1}, \ldots, \tilde{\pi}_{n-1,n-1})'$ denote the vector containing the coefficients of the finite predictor of length $n-1$. Then, $\tilde{\pi}_{kk} = 0, k > L$, implies $\tilde{\pi}_{(n-1)} = [\tilde{\pi}'_{(L)}, 0, \ldots, 0]'$, where $\tilde{\pi}_{(L)} =$

$(\tilde{\pi}_{L,1}, \ldots, \tilde{\pi}_{L,L})'$ is obtained as the solution of the Yule-Walker linear system $\tilde{\Gamma}_L \tilde{\pi}_{(L)} = \tilde{\gamma}_{(L)}$, with $\tilde{\gamma}_{(L)} = (\tilde{\gamma}_r(1), \ldots, \tilde{\gamma}_r(L))'$.

Hence, denoting $\phi_{(n-1)} = (\phi_{n-1,1}, \ldots, \phi_{n-1,n-1})'$, $\pi_{(n-1)} = (\pi_{n-1,1}, \ldots, \pi_{n-1,n-1})'$, $\gamma_{(L)} = (\gamma_r(1), \ldots, \gamma_r(L))'$, and by $\Gamma_{r,L}$ the $L \times L$ Toeplitz matrix formed from $\gamma_r(j), j = 0, \ldots, L-1$,

$$
\begin{aligned}
\tilde{\pi}_{(n-1)} - \phi_{(n-1)} &= \tilde{\pi}_{(n-1)} - \pi_{(n-1)} + \pi_{(n-1)} - \phi_{(n-1)} \\
&= \tilde{\pi}_{(L)} - \pi_{(L)} + \pi_{(n-1)} - \phi_{(n-1)},
\end{aligned}
$$

so that, by Lemma 2 and Jensen's inequality,

$$
||\tilde{\pi}_{(n-1)} - \phi_{(n-1)}||_2 \leq ||\tilde{\pi}_{(L)} - \pi_{(L)}||_2 + \sum_{j=\ell+1}^{n-1} |\phi_{jj}|.
$$

The first term on the right hand side can be written:

$$
\begin{aligned}
\tilde{\pi}_{(L)} - \pi_{(L)} &= \tilde{\Gamma}_L^{-1} \tilde{\gamma}_{(L)} - \Gamma_{r,L}^{-1} \gamma_{(L)} \\
&= \tilde{\Gamma}_L^{-1} \left(\tilde{\gamma}_{(L)} - \gamma_{(L)} \right) + \tilde{\Gamma}_L^{-1} \left(\Gamma_{r,L} - \tilde{\Gamma}_L \right) \Gamma_{r,L}^{-1} \gamma_{(L)}
\end{aligned}
$$

so that, applying the properties of the operator norm (see also Appendix A.1),

$$
||\tilde{\pi}_{(L)} - \pi_{(L)}||_2 \leq \rho\left(\tilde{\Gamma}_L^{-1}\right) ||\tilde{\gamma}_{(L)} - \gamma_{(L)}||_2 + \rho\left(\tilde{\Gamma}_L^{-1}\right) \rho\left(\Gamma_{r,L} - \tilde{\Gamma}_L\right) \rho(\Gamma_{r,L}^{-1}) ||\gamma_{(L)}||_2
$$

which shows that $||\tilde{\pi}_{(L)} - \pi_{(L)}||_2 = O_p(r_n)$, as $\rho\left(\tilde{\Gamma}_L^{-1}\right)$ and $||\gamma_{(L)}||_2$ are both bounded, and $||\tilde{\gamma}_{(L)} - \gamma_{(L)}||_2 = O_p(r_n)$.

G   Proof of Theorem 3

Writing
$$
\begin{aligned}
\hat{Y}_{n+1|n} - \tilde{Y}_{n+1|n} &= \sum_{j=1}^n (\phi_{n,j} - \tilde{\pi}_{n,j}) Y_{n+1-j} \\
&= \sum_{j=1}^L (\phi_{n,j} - \tilde{\pi}_{n,j}) Y_{n+1-j} + \sum_{j=L+1}^n \phi_{n,j} Y_{n+1-j} \\
&= A + B,
\end{aligned}
$$

it follows from the Cauchy-Schwarz inequality and Theorem 2

$$
\begin{aligned}
|A| &= \left| \sum_{j=1}^L (\phi_{n,j} - \tilde{\pi}_{n,j}) Y_{n+1-j} \right| \\
&\leq ||\phi_{(n)} - \tilde{\pi}_{(n)}||_2 \left(\sum_{j=1}^L Y_{n+1-j}^2 \right)^{1/2} \\
&= O_p(L^{1/2} r_n)
\end{aligned}
$$

$$
\begin{aligned}
|B| &= \left| \sum_{j=L+1}^n \phi_{n,j} Y_{n+1-j} \right| \\
&\leq \left[\sum_{j=L+1}^{n-1} \phi_{n,j}^2 \right]^{1/2} \left(\sum_{j=1}^L Y_{n+1-j}^2 \right)^{1/2} \\
&\leq \sum_{j=L+1}^{n-1} |\phi_{n,j}| \left(\sum_{j=1}^L Y_{n+1-j}^2 \right)^{1/2} \\
&\leq C_2 n^{1/2} \sum_{j=L+1}^\infty |\phi_j|,
\end{aligned}
$$

where the last inequality follows from writing $\sum_{j=L+1}^n |\phi_{n,j}| = \sum_{j=L+1}^n |\phi_{n,j} - \phi_j + \phi_j|$ and applying Baxter's inequality (see Pourahmadi (2001), Theorem 7.22). Hence, given two constants $C_1, C_2 > 0$,

$$
|\hat{Y}_{n+1|n} - \tilde{Y}_{n+1|n}| \leq C_1 L^{1/2} r_n + C_2 \sqrt{n} \sum_{j=L+1}^\infty |\phi_j|.
$$

References

Barndorff-Nielsen, O., and Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *Journal of multivariate Analysis*, *3*(4), 408–419.

Baxter, G. (1962). An asymptotic result for the finite predictor. *Mathematica Scandinavica*, *10*, 137–144.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, 489–502.

Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals of Statistics*, 224–231.

Bickel, P. J., and Gel, Y. R. (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(5), 711–728.

Bingham, N. (2012). Szegö's theorem and its probabilistic descendants. *Probability Surveys*, *9*.

Brémaud, P. (2014). *Fourier analysis of stochastic processes*. Springer.

Brillinger, D. R. (1981). *Time series: data analysis and theory* (Vol. 36). Siam.

Brockwell, P. J., and Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics.

Bühlmann, P. (1995). Moving-average representation of autoregressive approximations. *Stochastic processes and their applications*, *60*(2), 331–342.

Daniels, M. J., and Pourahmadi, M. (2009). Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis*, *100*(10), 2352–2363.

Dedecker, J., Doukhan, P., Lang, G., Rafael, L. R. J., Louhichi, S., and Prieur, C. (2007). Weak dependence. In *Weak dependence: With examples and applications* (pp. 9–20). Springer.

Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 233–244.

Golub, G. H., and Van Loan, C. F. (2012). *Matrix computations* (Vol. 3). JHU Press.

Goncalves, S., and Kilian, L. (2007). Asymptotic and bootstrap inference for ar (âˆž) processes with conditional heteroskedasticity. *Econometric Reviews*, *26*(6), 609–641.

Gupta, S. D., Mazumdar, R. R., and Glynn, P. (2013). On the convergence of the spectrum of finite order approximations of stationary time series. *Journal of Multivariate Analysis*, *121*, 1–21.

Hannan, E., Heyde, C., et al. (1972). On limit theorems for quadratic functions of discrete time series. *The Annals of Mathematical Statistics*, *43*(6), 2058–2066.

Hannan, E. J. (1960). *Time series analysis.* Methuen& Co., Ltd., John Wiley& Sons.

Hannan, E. J., and Deistler, M. (1988). *The statistical theory of linear systems* (Vol. 70). SIAM.

Inoue, A. (2008). Ar and ma representation of partial autocorrelation functions, with applications. *Probability Theory and Related Fields*, *140*(3-4), 523–551.

Levinson, N. (1946). The wiener (root mean square) error criterion in filter design and prediction. *Studies in Applied Mathematics*, *25*(1-4), 261–278.

Lewis, R., and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of multivariate analysis*, *16*(3), 393–411.

Lütkepohl, H., and Saikkonen, P. (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics*, *81*(1), 127–157.

McMurry, T. L., and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, *31*(6), 471–482.

McMurry, T. L., Politis, D. N., et al. (2015). High-dimensional autocovariance matrices and optimal linear prediction. *Electronic Journal of Statistics*, *9*, 753–788.

Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, *15*(4-5), 517–533.

Politis, D. N., and Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of time series analysis*, *16*(1), 67–103.

Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory* (Vol. 379). John Wiley & Sons.

Saikkonen, P., and Lütkepohl, H. (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric Theory*, *12*(05), 814–844.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(40), 14150–14154.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface*, *4*(2), 207–226.

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

Research Papers
2016

CREATES
Center for Research in Econometric
Analysis of Time Series

2017-03:	Peter Exterkate and Oskar Knapik: A regime-switching stochastic volatility model for forecasting electricity prices
2017-04:	Timo Teräsvirta: Sir Clive Granger's contributions to nonlinear time series and econometrics
2017-05:	Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperatures and Co–Shifting: A Vector Shifting–Mean Autoregressive Analysis
2017-06:	Tobias Basse, Robinson Kruse and Christoph Wegener: The Walking Debt Crisis
2017-07:	Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market
2017-08:	Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold: Insight into the Female Longevity Puzzle: Using Register Data to Analyse Mortality and Cause of Death Behaviour Across Socio-economic Groups
2017-09:	Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations
2017-10:	Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability
2017-11:	Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models
2017-12:	Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term
2017-13:	Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence
2017-14:	Martin M. Andreasen and Anders Kronborg: The Extended Perturbation Method: New Insights on the New Keynesian Model
2017-15:	Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: A Non-Structural Investigation of VIX Risk Neutral Density
2017-16:	Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Does the ARFIMA really shift?
2017-17:	Massimo Franchi and Søren Johansen: Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles
2017-18:	Matias D. Cattaneo, Michael Jansson and Kenichi Nagasawa: Bootstrap-Based Inference for Cube Root Consistent Estimators
2017-19:	Daniel Borup and Martin Thyrsgaard: Statistical tests for equal predictive ability across multiple forecasting methods
2017-20:	Tommaso Proietti and Alessandro Giovannelli: A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices