Consistent Autoregressive Spectral Estimates: Nonlinear Time Series and Large Autocovariance Matrices

Jiang Wang and Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, 92036-0112

Abstract

We consider the problem of using an autoregressive (AR) approximation to estimate the spectral density function and the $n \times n$ autocovariance matrix based on stationary data $X_1, \ldots, X_n$. The consistency of the autoregressive spectral density estimator has been proven since the 1970s under a linearity assumption. We extend these ideas to the non-linear setting, and give an application to estimating the $n \times n$ autocovariance matrix. Under mild assumptions on the underlying dependence structure and the order p of the fitted $AR(p)$ model, we are able to show that the autoregressive spectral estimate and the associated AR-based autocovariance matrix estimator are consistent. We are also able to establish an explicit bound on the rate of convergence of the proposed estimators.

Keywords: Nonlinear time series, Spectral Density, Covariance Matrix Estimation

1 Introduction

Let $X_1, \ldots, X_n$ be an observed realization of a second order stationary time series with mean zero and autocovariance function $\gamma_k = E[X_t X_{t-k}]$. Based on these data, we wish to estimate the spectral density

$$f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_k e^{\mathrm{i}k\lambda} \tag{1}$$

for some $\lambda \in [-\pi, \pi]$, as well as the $n \times n$ autocovariance matrix.

$$\Gamma_n = [\gamma_{|i-j|}]_{i,j=1}^n \tag{2}$$

One may be tempted to estimate $f(\lambda)$ and Γ_n by the plug-in estimators, which replace γ_k with $\hat{\gamma}_k = \frac{1}{n}\sum_{i=1}^{n-k}(X_i - \bar{X})(X_{i+k} - \bar{X})$ is the the sample autocovariance at lag k. However, as $n \to \infty$, these naive plug-in estimators are not consistent; see Wu and Pourahmadi(2009), and also Theorem 2 of Xiao and Wu(2011).

In the setting of a linear time series, Berk(1974) showed the consistency of the autoregressive (AR) estimate of spectral density. In a possibly nonlinear setting, kernel smoothed estimators of the spectral density have been the go-to method since the seminal work of Parzen (1957). It is well-known that these estimators are equivalent to the so-called 'lag-window' estimators that taper the $\hat{\gamma}_k$ estimates before plugging them in the sum given in eq. (1); see e.g. Politis and Romano (1995).

The problem of estimating the $n \times n$ autocovariance matrix Γ_n shares some similarities with estimating the spectral density. Wu and Pourahmadi(2009) constructed a consistent estimator of Γ_n by plugging in $\hat{\gamma}_k \mathbb{1}_{\{|k| \le \ell\}}$ in place of the unknown γ_k; here, $\mathbb{1}_A$ is the indicator of set A. This proposal results into a banded matrix estimator, i.e., the (i,j) entry of the matrix estimator is zero when $|i-j| > \ell$ which is the banding parameter. Subsequently, McMurry and Politis(2010) replaced the hard banding produced by the indicator $\mathbb{1}_{\{|k| \le \ell\}}$ with a tapering function that smoothly goes to zero as k approaches ℓ; this results into a banded and tapered matrix estimator, hereafter denoted by $\hat{\Gamma}_n^{BT}$, with better finite-sample properties.

Recall that the autocovariance structure of a Moving Average (MA) model of order q is characterized by vanishing autocovariances for lags bigger than q. Hence, by being banded at ℓ, the matrix $\hat{\Gamma}_n^{BT}$ corresponds to the autocovariance structure of an MA(q) model with order $q = \ell$. However, when the data are generated by an $AR(p)$ model, it is conceivable that an estimator can be devised that captures the AR–structure better than lag window/tapered estimates. Indeed, one may need an MA(q) with high values of q to approximate well even a low order AR model. Consequently, McMurry and Politis (2015, Rejoinder) proposed an AR-based estimator of Γ_n, hereafter denoted by $\hat{\Gamma}_n^{AR}$, and empirically studied its performance comparing it to that of $\hat{\Gamma}_n^{BT}$. Notably, no asymptotic results on $\hat{\Gamma}_n^{AR}$ are available to-date.

Showing $\hat{\Gamma}_n^{AR}$ is consistent is not trivial even in a linear time series setting, as it requires careful arguments based on Berk's (1974) consistent autoregressive spectral estimates. But to make the matrix estimator $\hat{\Gamma}_n^{AR}$ competitive with $\hat{\Gamma}_n^{BT}$, it must be shown consistent even in a possibly nonlinear setting (since the latter is). Hence, in this paper we revisit Berk's (1974) seminal paper, and establish consistency (and rate of convergence) of AR-based spectral estimation in a possibly nonlinear time series setting; we then apply these results to show consistency of $\hat{\Gamma}_n^{AR}$.

The remainder of the paper is structured as follows: Section 2 presents the structure of the autoregressive spectral density estimator and the AR-based autocovariance matrix estimator; Section 3 presents our main results for consistency rates of spectral density estimator; Section 4 presents the application of our result to the autocovariance matrix estimator; Section 5 discusses the simulation study and our findings. All technical proofs are placed in the Appendix.

2 Construction of AR-based spectral density and autocovariance estimators

By the well-known Wold representation, every purely nondeterministic stationary and zero-mean stochastic process $\{X_t\}$ can be expressed as

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ where } Z_t \sim \ WN\ (0, \sigma^2) \text{ and } \psi_0 = 1; \tag{3}$$

the notation $WN\ (0, \sigma^2)$ means white noise with zero mean and variance σ^2. Further, for all purely nondeterministic, stationary and zero-mean time series with nonvanishing spectral density, there exist unique autoregressive coefficients $(\phi_i,\ i = 0, 1, 2, \ldots)$ that only depend on the autocovariance function of the time series $\{X_t\}$. Under the additional assumption that the coefficients $(\phi_i,\ i = 0, 1, 2, \ldots)$ are absolute summable, i.e., $\sum_{i=0}^{\infty} |\phi_i| < \infty$, one then obtains an autoregressive, Wold-type representation of the underlying process given by

$$\sum_{j=0}^{\infty} \phi_j X_{t-j} = Z_t \text{ where } \phi_0 = 1 \tag{4}$$

and Z_t are the innovations appearing in eq. (3); for details, see see eq. (2.1), (2.2) and (2.3) of Kreiss, Paparoditis and Politis (2011), or Corollary 6.1.17 of McElroy and Politis (2020).

For some choice of p, we compute the sample autocovariances $\hat{\gamma}_k = n^{-1} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X})$ from the data $X_1, \ldots, X_n$, and use them to fit an $AR(p)$ model via the Yule-Walker equations:

$$\hat{\Gamma}_p \underline{x} = \underline{\hat{\gamma}}(p)$$

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \underline{x}(p)'\underline{\hat{\gamma}}(p)$$

where $\hat{\Gamma}_p = [\hat{\gamma}_{|i-j|}]_{i,j=1}^p$ and $\underline{\hat{\gamma}}(p) = (\hat{\gamma}_1, \ldots, \hat{\gamma}_p)'$. Denote the solution of the above system of equations as $\underline{\hat{\phi}}(p) = (\hat{\phi}_1, \ldots, \hat{\phi}_p)'$. It is well-known that the fitted $AR(p)$ model is causal; see e.g. Brockwell and Davis (1991, Problem 8.3).

The spectral density of the fitted $AR(p)$ model is

$$\hat{f}^{AR}(\lambda) = \frac{\hat{\sigma}^2}{2\pi|\hat{\Phi}(e^{i\lambda})|^2} \tag{5}$$

which is the AR-based spectral density estimator; as usual, $\hat{\Phi}(z) = 1 - \hat{\phi}_1 z - \cdots - \hat{\phi}_p z^p$.

Furthermore, denote $\hat{\gamma}_k^{AR}$ the lag-k autocovariance implied by the fitted $AR(p)$ model, i.e., the lag-k autocovariance of an $AR(p)$ model with AR-coefficients $\underline{\hat{\phi}}(p) = (\hat{\phi}_1, \ldots, \hat{\phi}_p)'$, and error variance $\hat{\sigma}^2$. It is easy to see that $\hat{\gamma}_k^{AR} = \hat{\gamma}_k$ for $k = 0, 1, \ldots, p$. However, for lags k higher than p, $\hat{\gamma}_k^{AR}$ is obtained by solving (or just iterating) the difference equation that the fitted $AR(p)$ model implies for its autocovariance; see e.g. Brockwell and Davis (1991, Example 3.3.4).

By definition of spectral density, we can write the AR-based spectral density estimator in a different (but equivalent) form, namely:

$$\hat{f}^{AR}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \hat{\gamma}_j^{AR} e^{-\mathrm{i}\lambda \mathrm{j}}. \tag{6}$$

Finally, we let

$$\hat{\Gamma}_n^{AR} = [\hat{\gamma}_{|i-j|}^{AR}]_{i,j=1}^n \tag{7}$$

be the $n \times n$ autocovariance matrix corresponding to our fitted $AR(p)$ model. We will use this matrix as our AR-based autocovariance matrix estimator as proposed by McMurry and Politis (2015, Rejoinder).

Recall that the banded matrix estimator of Wu and Pourahmadi(2009) and the banded and tapered matrix estimator of McMurry and Politis(2010) have (i, j) entries that are zero if $|i - j| > \ell$. By contrast, the (i, j) entries of $\hat{\Gamma}_n^{AR}$ are generally nonzero even when $|i - j|$ is large. To see why, note that the first row of $\hat{\Gamma}_n^{AR}$ is given by $\hat{\gamma}_k^{AR}$ for $k = 0, 1, \ldots, n-1$ which will display the (often oscillatory) pattern associated with an $AR(p)$ model.

In the next Section, we proceed to establish the convergence of the autocovariance spectral density estimator $\hat{f}^{AR}(\lambda)$.

3 Consistency of the AR-based spectral estimator

Every purely nondeterministic stationary and zero-mean series is represented by its Wold-type representation coefficients by (4). In order to prove consistency of the

autoregressive spectral estimator $\hat{f}^{AR}(\lambda)$ to the true spectral density $f(\lambda)$, the bridge is to quantify the distance between $\underline{\hat{\phi}}(p)$ and $\underline{\phi}(p) = (\phi_1, \ldots, \phi_p)'$. The above two vectors could be connected through the coefficients of an optimal (in the mean square sense) autoregressive fit of order p, or equivalently, through coefficients based on the finite past. To be precise, denote the solution $\underline{x}(p)$ of the following Yule-Walker linear equations

$$\Gamma_p \underline{x}(p) = \underline{\gamma}(p) \tag{8}$$

as $\underline{\phi}^{AR}(p) = (\phi_1^{AR}, \ldots, \phi_p^{AR})'$. Recall from Brockwell and Davis(1991, Proposition 5.1.1) that the covariance matrix Γ_p on the left hand side is invertible for all p provided $\gamma_0 > 0$ and $\gamma_h \to 0$ as $h \to \infty$. Next we will talk about the distance between ϕ(p) and $\underline{\phi}^{AR}(p)$ and the distance between $\underline{\hat{\phi}}(p)$ and $\underline{\phi}^{AR}(p)$ respectively.

The distance between $\underline{\phi}(p)$ and $\underline{\phi}^{AR}(p)$ is straightforward. By slight modifications of Baxter (1962, Theorem 2.2) —see also Pourahmadi (2001, Theorem 7.22)— we obtain the following helpful result relating the coefficients $\underline{\phi}^{AR}(p)$ to $\underline{\phi}(p)$.

Theorem 1. *Assume that $f(\lambda)$ is strictly positive and continuous, and that $\sum_{h=0}^{\infty}(1+h)^r|\gamma_h| < \infty$ for some $r \geq 0$. Then there exists p_o and $C > 0$(both depending on $f(\lambda)$ only) such that for all $p \geq p_o$:*

$$||\underline{\phi}^{AR}(p) - \underline{\phi}(p)||_2 \leq C \cdot \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k| \tag{9}$$

as well as

$$\sum_{k=1}^{\infty}(1+k)^r|\phi_k| < \infty \tag{10}$$

where $||\cdot||_2$ stands for vector 2-norm.

In order to quantify the distance between $\underline{\hat{\phi}}(p)$ and $\underline{\phi}^{AR}(p)$, we have to make some further assumptinos on our series. Let us introduce the following physical dependence measure conditions from Xiao and Wu (2012). Let $\{\epsilon_i\}$ be a sequence of i.i.d. random variables. Assume that X_t is a causal function of $\{\epsilon_s, s \leq t\}$, i.e.,

$$X_t = g(\epsilon_t, \epsilon_{t-1}, \ldots) \tag{11}$$

where g is some measurable function such that X_t is well defined and has finite second moment. Let $\epsilon_i^{'}$ be an independent copy of ϵ_i, and let $X_t^{'} = g(\epsilon_t, \ldots, \epsilon_1, \epsilon_o', \epsilon_{-1}, \ldots)$.

Following Wu(2005), for some $\alpha > 0$, define the physical dependence measure

$$\delta_\alpha(t) = E[|X_t - X_t^{'}|^\alpha]^{\frac{1}{\alpha}}$$

To measure the cumulative dependence across times, the quantity

$$\Theta_\alpha(m) = \sum_{t=m}^{\infty} \delta_\alpha(t)$$

is helpful. We adopt the convention that $\delta_\alpha(t) = 0$ for $t < 0$. We will say that $\{X_t\}$ is short-range dependent with moment α if $\Theta_\alpha = \Theta_\alpha(0) < \infty$. For example, if the series X_t has a stationary ARMA representation, then condition $\Theta_\alpha(m) = O(m^{-\beta})$ for any $\beta > 0$; details can be found in Shao and Wu(2007).

The short-range dependence assumption allows us to establish the consistency rate of $\hat{\gamma}_k$ to γ_k. The consistency of $\hat{\Gamma}_p$ to Γ_p and $\underline{\hat{\gamma}}(p)$ to $\underline{\gamma}(p)$ are therefore established as well. Using Lemma 3 and Theorem 4 of Xiao and Wu(2012), we have got a useful lemma characterizing the eigenvalues of $\hat{\Gamma}_p$ and the consistency of $\hat{\Gamma}_p$ to Γ_p.

Lemma 1. *Assume that the spectral density is continuous and $0 < F_1 \le f(\lambda) \le F_2$ for some real numbers F_1, F_2 and $\sum_{h=0}^{\infty}(1+h)^r|\gamma_h| < \infty$ for some $r \ge 0$. Denote by $\hat{\lambda}_1 \le \hat{\lambda}_2 \le \cdots \le \hat{\lambda}_p$ the eigenvalues of $\hat{\Gamma}_p$. Choose p such that $p = O(n^\gamma)$, and assume that for some $\alpha > 4$ we have $\Theta_\alpha < \infty$ and $\Theta_\alpha(m) = O(m^{-\beta})$, where the real numbers β, γ satisfy*

$$0 < \gamma < 1, \quad \gamma < \frac{\beta\alpha}{2} \quad \textit{and} \quad (1-2\beta)\gamma < \frac{\alpha-4}{\alpha}. \tag{12}$$

Then, for $b_n = \frac{2}{n}\sum_{k=1}^{p} k|\gamma_k| + 2\sum_{k=p+1}^{\infty} k|\gamma_k|$ and $c_\alpha = (\alpha+4)e^{\frac{\alpha}{4}}\Theta_4^2$, we have:

$$\lim_{n\to\infty} P(\hat{\lambda}_p \le 2\pi F_2 + 12c_\alpha\sqrt{\frac{p\log p}{n}} + 2\pi b_n) = 1 \tag{13}$$

as well as

$$\lim_{n\to\infty} P(\hat{\lambda}_1 \ge 2\pi F_1 - 12c_\alpha\sqrt{\frac{p\log p}{n}} - 2\pi b_n) = 1 \tag{14}$$

In the next part, we establish convergence rates for the matrix estimator in the operator norm, defined by

$$\rho(A) = \max_{||\underline{x}(n)||_2=1} ||A\underline{x}(n)||_2$$

It is worth noting that $\rho(A) = \sqrt{\lambda_{max}(A'A)}$. We then have the following Corollary:

Corollary 1. *Choose p such that $p = O(n^\gamma)$, and assume that for some $\alpha > 4$ we have $\Theta_\alpha < \infty$ and $\Theta_\alpha(m) = O(m^{-\beta})$ such that eq. (12) holds.*

Then, for $b'_n = \frac{2}{n}\sum_{k=1}^{p} k|\gamma_k|$ and $c_\alpha = (\alpha+4)e^{\frac{\alpha}{4}}\Theta_4^2$, we have:

$$\lim_{n\to\infty} P(\rho(\hat{\Gamma}_p - \Gamma_p) \le 12c_\alpha\sqrt{\frac{p\log p}{n}} + b'_n) = 1 \tag{15}$$

Remark 1. We are assuming (for simplicity) that X_t has mean zero; however, in practice all time series data are centered at the sample mean before processing. This is why we are working with the data-centered autocovariance estimator $\hat{\gamma}_k = \frac{1}{n}\sum_{i=1}^{n-k}(X_i - \bar{X})(X_{i+k} - \bar{X})$ which is common in the literature. Notably, Xiao and Wu(2012) are using $\hat{\gamma}_k^* = \frac{1}{n}\sum_{i=1}^{n-k} X_i X_{i+k}$, i.e., implying they have knowledge that X_t has mean zero. However, it is not difficult to show that $|\hat{\gamma}_k^* - \hat{\gamma}_k| = O_p(\frac{1}{n})$ in our setting, which implies that $\hat{\gamma}_k$ could replace $\hat{\gamma}_k^*$ without influencing any of our theoretical results.

Now under the short-range dependent structure, with the help of Lemma 1 and Corollary 1, we can establish the convergence of $\hat{\underline{\phi}}(p)$ to $\underline{\phi}^{AR}(p)$:

Theorem 2. *Under the combined assumptions of Lemma 1 and Corollary 1, we further assume that n and p satisfy $\frac{p^2 \log p}{n} \to 0$. Then we have:*

$$\lim_{n\to\infty} P(||\hat{\underline{\phi}}(p) - \underline{\phi}^{AR}(p)||_2 \leq c_p(c_\alpha \sqrt{\frac{p \log p}{n}} + b_n')) = 1 \tag{16}$$

where $c_p = \frac{1}{\pi F_1}(\sqrt{p} + \frac{\sum_{k=0}^{\infty}|\gamma|_k}{2\pi F_1})$.

The following corollary is immediate.

Corollary 2. *Under the assumptions of Theorem 2, we have*

$$||\hat{\underline{\phi}}(p) - \underline{\phi}(p)||_2 = O_p(\sqrt{\frac{p^2 \log p}{n}}) + C \cdot p^{\frac{1}{2}} \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|. \tag{17}$$

Corollary 3. *Under the assumptions of Theorem 2, we further assume that for every positive d, $||\Gamma_d^{-1}||_1 < M$ for some positive number M. We then have*

$$||\hat{\underline{\phi}}(p) - \underline{\phi}(p)||_1 = O_p(\sqrt{\frac{p^2 \log p}{n}}) + C \cdot \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|. \tag{18}$$

Remark 2. *To assess the strength of the assumptions of Lemma 1 and Theorem 2, note that the condition $\Theta_\alpha(m) = O(m^{-\beta})$ implies that $\sum_{k=m}^{\infty}|\gamma_i| = O(m^{-\beta})$. This further implies that the smoothness condition $\sum_{k=0}^{\infty}(k+1)^r|\gamma_k| < \infty$ is satisfied as long as $\beta \geq \frac{1}{2} + r$, in which case the last two equations of condition (12) could be dropped; for details, see Theorem 3 and its proof.*

We now return to the main focus $\hat{f}^{AR}(\lambda)$. Kreiss, Pararoditis and Politis (2011) have discovered that, under some mild conditions, there exists some p_o such that when $p > p_o$, $|\hat{\Phi}(e^{i\lambda})|^{-1}$ is bounded above. This in turn leads to that (5) is bounded. Therefore equivalently, we only need to show that both the numerator and the denominator in (5) converge to their estimands. Putting it all together, we reach our main result below; for a detailed proof, see the Appendix.

Theorem 3. *Assume that the spectral density is continuous, positive and bounded. Further, assume that for some $\alpha > 4$ we have $\Theta_\alpha < \infty$ and $\Theta_\alpha(m) = O(m^{-\beta})$ with $\beta > r + 1/2$. Choose p such that $p = O(n^\gamma)$ with $0 < \gamma < \frac{1}{3}$. Then, as $n \to \infty$,*

$$\sup_{\lambda \in [0,2\pi]} |\hat{f}^{AR}(\lambda) - f(\lambda)| \to 0 \tag{19}$$

in probability. In particular, we have

$$\sup_{\lambda \in [0,2\pi]} |\hat{f}^{AR}(\lambda) - f(\lambda)| = O_p(\sqrt{\frac{p^3 \log p}{n}}) + C \cdot p^{\frac{1}{2}} \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|. \tag{20}$$

If we further assume that for every positive d, $||\Gamma_d^{-1}||_1 < M$ for some positive number M, then we also have

$$\sup_{\lambda \in [0,2\pi]} |\hat{f}^{AR}(\lambda) - f(\lambda)| = O_p(\sqrt{\frac{p^2 \log p}{n}}) + C \cdot \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|. \tag{21}$$

Theorem 3 will typically be applied with $p \to \infty$ as $n \to \infty$, i.e., using a higher and higher order of the approximating AR model in order to better estimate the underlying spectral density. However, if the spectral density at hand is that of a true AR model of finite order p_{true} (say), then we do not need to take $p \to \infty$; in this case, Theorem 3 even applies for finite p (as long as it is greater or equal to p_{true}), since this is sufficient to make the second term on the right-hand-side of (20) vanish.

Remark 3. *Under a linear time series assumption, Berk(1974) has shown the convergence rate in (19) is bounded by $O_p(\sqrt{\frac{p^3}{n}} + p^{\frac{1}{2}} \sum_{k=p+1}^{\infty} |\phi_k|)$. Eq. (20) shows that a similar (up to a logarithmic term) upper bound holds true even without the linearity assumption.*

4 Consistency of AR-based autocovariance matrix estimator

The application to autocovariance matrices is natural. In the proof of of Corollary 1, we have already made use of the relationship between a Toeplitz matrix and its corresponding 'spectral density'; this directly leads to our next theorem.

Theorem 4. *Under all assumptions in Theorem 3, then $\rho(\hat{\Gamma}_n^{AR} - \Gamma_n)$ converges to 0 in probability. Further we have*

$$\rho(\hat{\Gamma}_n^{AR} - \Gamma_n) = O_p(\sqrt{\frac{p^3 \log p}{n}}) + C \cdot p^{\frac{1}{2}} \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|.$$

If we further assume that for every positive d, $||\Gamma_d^{-1}||_1 < M$ for some positive number M, then we have

$$\rho(\hat{\Gamma}_n^{AR} - \Gamma_n) = O_p(\sqrt{\frac{p^2 \log p}{n}}) + C \cdot \sum_{k=p+1}^{\infty} (1+k)^r |\phi_k|.$$

Remark 4. *The above rate can be compared with the rate associated with the banded and tapered estimator $\hat{\Gamma}_n^{BT}$ that satisfies $\rho(\hat{\Gamma}_n^{BT} - \Gamma_n) = O_p(\sum_{j=p+1}^{\infty} |\gamma_j| + p^{\frac{1}{2}} (\log p)^{\frac{1}{2}} n^{-\frac{1}{2}})$. If the true underlying time series satisfies an AR(p) model with a finite $p(= p_*$, say), then in constructing $\hat{\Gamma}_n^{AR}$ we can take $p = p_*$ (finite), and the two rates coincide up to a logarithmic term. However, if the true underlying time series satisfies an MA(q) model with a finite q, then the banded/tapered estimator $\hat{\Gamma}_n^{BT}$ achieves a better rate than $\hat{\Gamma}_n^{AR}$ since now p must tend to infinity as $n \to \infty$ but the banding parameter ℓ can remain finite. These asymptotic remarks are born out in finite-sample simulations given in McMurry and Politis (2015, Rejoinder), as well as in our own empirical work presented in the next Section.*

5 Simulation Study and Discussion

5.1 Simulation and Results

We conducted several simulations with the aim of a direct comparison between the AR matrix estimator and the Banded and Tapered estimator of McMurry and Politis(2010) which is denoted by BT-ER because the relevant tuning parameter is chosen via the empirical rule (ER) of Politis(2003). The empirical rule (ER) works based on a plot of the sample autocorrelations, and deciding if these are negligible after some lag.

Implementation of the AR matrix estimator $\hat{\Gamma}_n^{AR}$ requires choosing the order p of the fitted AR model which can readily be done via minimizing AIC (denoted AR-AIC). However, choosing p can also be done by an investigation of the plot of the sample *partial* autocorrelations which would vanish for lags bigger than p under a true AR(p) model. This intuitive notion was made rigorous in Section 10.3 of McElroy and Politis(2020) in a fashion analogous to the aforementioned empirical rule of Politis(2003). We decided to implement this PACF empirical rule in implementing the AR matrix estimator, and denoted this method as AR-PACF.

Our simulation results show that the AR matrix estimator $\hat{\Gamma}_n^{AR}$ shows substantial improvements when the time series is indeed an AR process (especially with a large autoregressive coefficient). Unsurprisingly, the Banded and Tapered estimator BT-ER performs better in the setting of an MA model. In addition, the PACF empirical rule tends to choose a more accurate order than the traditional AIC approach under a true, finite-order autoregressive setting.

We consider four different kinds of models: AR(1), MA(1), AR(p) and a non-linear model.

- **AR(1)** The data was simulated from the AR(1) process $X_t = \phi X_{t-1} + \epsilon_t$, where the ϵ_t were iid $N(0, 1-\phi^2)$, for $\phi = \pm 0.1, \pm 0.5, \pm 0.9$; the error variance was chosen to make $Var[X_t] = 1$ for all simulations.

- **MA(1)** The data was simulated from the MA(1) process $X_t = \epsilon_t + \theta \epsilon_{t-1}$, for $\theta = \pm 0.1, \pm 0.5, \pm 0.9$; the error variance was chosen to make $Var[X_t] = 1$ for all simulations.

- **AR(p)** We generate a large AR(p) model, where $p = 3$. The data is simulated from $X_t = 0.9X_{t-1} - 0.5X_{t-2} + 0.3X_{t-3} + \epsilon_t$; errors are iid normal, and the variance was chosen to make $Var[X_t] = 1$ for all simulations.

- **Non-linear** The data was drawn from the model $X_t = 0.5|X_{t-1}| + \epsilon_t$, where ϵ_t were iid normal and variance chosen such that $Var[X_t] = 1$. The autocovariance function for the absolute AR model does not have a simple closed form, so we approximate the true Γ_n with its empirical version using a very large amount of data(n = 100000). while this provides estimates of the loss, it induces some additional, and difficult to quantify, uncertainty into the results.

For all of the models, we generate $n = 64, 256$ and 1024 data points, and the performance is recomputed $N = 1000$ new replications. Tables 1–3 contain the simulation results; in general, the AR matrix estimator $\hat{\Gamma}_n^{AR}$ has a better performance than the banded and tapered covariance matrix estimator $\hat{\Gamma}_n^{BT}$ under a true AR setting no matter what method we use to choose p, especially for large autoregressive parameters. Furthermore, under an AR setting, it appears we have roughly root-n consistency for AR-AIC and AR-PACF, and somehow slower convergence rate for BT-ER. On the other hand, under a moving average setting, the BT-ER outperforms the other methods and appears to have roughly root-n consistency.

When the model itself has very small autocorrelations, like $AR(1)$ with $\phi = \pm 0.1$ and $MA(1)$ with $\theta = \pm 0.1$, the empirical rule could end up choosing banding parameter $\ell = 0$. Same thing happens for the AR-PACF approach with partial autocorrelations. However, since we want to fit an $AR(p)$ model by the Yule-Walker method, we set the minimum lag to be at least 1. This explains an apparent mismatch in results of models like $AR(1)$ with $\phi = \pm 0.1$ and $MA(1)$ with $\theta = \pm 0.1$; see Remark 5 for more discussion.

Model	Parameter	AR-AIC	AR-PACF	BT-ER
AR(1)	$\phi = -0.9$	11.2427(12.8360)	10.86762(12.8229)	12.1903(8.4601)
AR(1)	$\phi = -0.5$	1.3238(1.7791)	1.1669(1.3266)	1.4615(1.0081)
AR(1)	$\phi = -0.1$	0.6134(0.9141)	0.3009(0.2776)	0.3710(0.1802)
AR(1)	$\phi = 0.1$	0.5522(0.8004)	0.2750(0.2816)	0.3697(0.1698)
AR(1)	$\phi = 0.5$	1.2896(1.1929)	1.0106(0.8770)	1.4123(0.6585)
AR(1)	$\phi = 0.9$	10.4621(5.6156)	10.2503(5.4467)	11.4615(3.7975)
MA(1)	$\theta = -0.9$	1.7137(1.3544)	1.2989(0.8668)	0.5112(0.4735)
MA(1)	$\theta = -0.5$	1.1506(1.0776)	0.8349(0.7787)	0.5889(0.4779)
MA(1)	$\theta = -0.1$	0.5698(0.8240)	0.3096(0.4069)	0.3701(0.2162)
MA(1)	$\theta = 0.1$	0.5790(0.8362)	0.2921(0.3759)	0.3745(0.2212)
MA(1)	$\theta = 0.5$	1.1111(1.1581)	0.7837(0.6540)	0.6238(0.4911)
MA(1)	$\theta = 0.9$	1.6840(1.3963)	1.3142(0.9062)	0.5698(0.5412)
AR(3)	$\phi = (0.9, -0.7, 0.5)$	2.9240(1.9473)	2.8520(1.8740)	2.9811(1.0867)
Absolute AR(1)	$\phi = 0.5$	0.8998(1.0738)	0.4839(0.5878)	0.6629(0.2399)

Table 1: Average operator norm loss(standard deviation) when $n = 64$

Model	Parameter	AR-AIC	AR-PACF	BT-ER
AR(1)	$\phi = -0.9$	8.4698(7.5174)	7.9419(6.9436)	10.2071(9.8205)
AR(1)	$\phi = -0.5$	0.7700(1.0235)	0.5601(0.5416)	0.8533(0.7818)
AR(1)	$\phi = -0.1$	0.2962(0.3214)	0.1425(0.1595)	0.2839(0.0690)
AR(1)	$\phi = 0.1$	0.3079(0.3632)	0.1334(0.1040)	0.2792(0.0725)
AR(1)	$\phi = 0.5$	0.6919(0.7054)	0.5248(0.4151)	0.8388(0.5686)
AR(1)	$\phi = 0.9$	7.9710(6.7533)	7.5502(5.7301)	9.0394(6.5020)
MA(1)	$\theta = -0.9$	1.3019(0.8387)	0.9307(0.4954)	0.2615(0.2885)
MA(1)	$\theta = -0.5$	0.6823(0.5816)	0.5896(0.2780)	0.2330(0.2178)
MA(1)	$\theta = -0.1$	0.3079(0.3707)	0.1462(0.1323)	0.2745(0.0638)
MA(1)	$\theta = 0.1$	0.2995(0.3151)	0.1404(0.1143)	0.2745(0.0839)
MA(1)	$\theta = 0.5$	0.6147(0.4646)	0.5690(0.2660)	0.2268(0.1980)
MA(1)	$\theta = 0.9$	1.2462(0.8274)	0.9025(0.4672)	0.2594(0.2557)
AR(3)	$\phi = (0.9, -0.7, 0.5)$	1.8086(1.5764)	1.6126(1.2869)	1.9341(1.1459)
Absolute AR(1)	$\phi = 0.5$	0.5257(0.4191)	0.3444(0.2092)	0.5896(0.1936)

Table 2: Average operator norm loss(standard deviation) when $n = 256$

For the purpose of comparing the different methods of choosing the AR order, we provide Table 4 showing the average AR order p chosen by the two methods AR-AIC and AR-PACF. We complement this by providing the average order q of a fitted

Model	Parameter	AR-AIC	AR-PACF	BT-ER
AR(1)	$\phi = -0.9$	4.4237(3.7529)	4.0719(3.3924)	6.6491(7.0058)
AR(1)	$\phi = -0.5$	0.3466(0.3113)	0.2564(0.2094)	0.4200(0.2188)
AR(1)	$\phi = -0.1$	0.1338(0.1390)	0.0732(0.0547)	0.1879(0.0767)
AR(1)	$\phi = 0.1$	0.1296(0.1186)	0.0857(0.2656)	0.1916(0.0797)
AR(1)	$\phi = 0.5$	0.4026(0.3529)	0.2814(0.2030)	0.4566(0.2644)
AR(1)	$\phi = 0.9$	4.3243(3.6388)	4.0664(3.2851)	6.0410(4.3676)
MA(1)	$\theta = -0.9$	0.8054(0.3990)	0.6159(0.2979)	0.1099(0.0854)
MA(1)	$\theta = -0.5$	0.3224(0.1848)	0.3124(0.1162)	0.0993(0.0719)
MA(1)	$\theta = -0.1$	0.1362(0.1337)	0.1148(0.0523)	0.1739(0.0751)
MA(1)	$\theta = 0.1$	0.1507(0.1619)	0.1290(0.0551)	0.1828(0.0772)
MA(1)	$\theta = 0.5$	0.3512(0.2212)	0.3182(0.1211)	0.1079(0.0956)
MA(1)	$\theta = 0.9$	0.8012(0.3434)	0.6292(0.2195)	0.1250(0.1142)
AR(3)	$\phi = (0.9, -0.7, 0.5)$	0.9841(0.7022)	0.8613(0.5854)	1.1590(0.6936)
Absolute AR(1)	$\phi = 0.5$	0.3606(0.4460)	0.1716(0.1941)	0.3562(0.1355)

Table 3: Average operator norm loss(standard deviation) when $n = 1024$

MA(q) model using the empirical rule of Politis(2003) to choose q; see also Section 10.1 of McElroy and Politis(2020).

Results show that when the model comes from a true AR(p) model, the PACF empirical rule chooses a more accurate order, i.e., closer to the true p, as compared to AIC. In an MA setting, the AIC typically chooses a higher order for the approximating AR model as compared to the PACF empirical rule. The tendency of the AIC towards (slight) over-fitting is well-known, and is not necessarily considered to be an issue since, in general, data may arise from an AR(∞) model.

Remark 5. *Denote the partial autocorrelation at lag k by α_k, and the autocorrelation by $\varrho_k = \gamma_k/\gamma_0$. The lag-1 autocorrelation and partial autocorrelation of an AR(1) model with $\phi = \pm 0.1$ are $\varrho_1 = \alpha_1 = \pm 0.1$, and that of an MA(1) with $\theta = \pm 0.1$ are $\varrho_1 = \alpha_1 \approx \pm 0.1$. The threshold for the two empirical rules (AR-PACF and BT-ER) in the context of Table 4 is $c = 0.106$ which is approximately 0.1. Since the estimates of ϱ_1 and α_1 are approximately normal with mean 0.1 (focusing momentarily on the case of a positive AR/MA coefficient), it follows that the estimates of ϱ_1 and α_1 would be under the threshold about half of the time. Hence, the two empirical rules would choose p and ℓ to be zero about half of the time (and 1 the other half). However, for practical purposes in the simulation we set that the minimum chosen order p should be 1, which may explain the differences seen in Table 4.*

Model	Parameter	p by AR-AIC	p by AR-PACF	q by BT-ER
AR(1)	$\phi = -0.9$	1.90	1.02	23.96
AR(1)	$\phi = -0.5$	2.00	1.07	2.85
AR(1)	$\phi = -0.1$	1.73	1.05	0.42
AR(1)	$\phi = 0.1$	1.75	1.10	0.36
AR(1)	$\phi = 0.5$	2.25	1.03	2.80
AR(1)	$\phi = 0.9$	1.98	1.02	22.47
MA(1)	$\theta = -0.9$	16.29	8.33	1.03
MA(1)	$\theta = -0.5$	4.58	2.34	1.03
MA(1)	$\theta = -0.1$	1.90	1.06	0.43
MA(1)	$\theta = 0.1$	1.99	1.08	0.38
MA(1)	$\theta = 0.5$	4.86	2.37	1.04
MA(1)	$\theta = 0.9$	15.91	8.52	1.10
AR(3)	$\phi = (0.9, -0.5, 0.3)$	3.711	3.025	5.87
Absolute AR(1)	$\phi = 0.5$	1.728	1.069	0.358

Table 4: Average order chosen when $n = 1024$; p is the order of fitted AR model, while q is the order of implicitly fitted MA model.

5.2 Discussion

The AR-based spectral density estimator $\hat{f}^{AR}(\lambda)$ is not new. However, its consistency under a possibly nonlinear setting was not established to date; the paper at hand fills this gap. This is an important result since $\hat{f}^{AR}(\lambda)$ has been widely used in applications, especially when the spectral density exhibits strong peaks/poles; see e.g. Kay (1988).

The AR-based autocovariance matrix estimator $\hat{\Gamma}_n^{AR}$ was proposed by McMurry and Politis (2015, Rejoinder) who provided simulations to compare its empirical performance to different banded/tapered estimators $\hat{\Gamma}_n^{BT}$. Interestingly, the consistency of $\hat{\Gamma}_n^{AR}$ was unknown even under the assumption of a linear time series. We were able to show the consistency, and quantify the rate of convergence, of $\hat{\Gamma}_n^{AR}$ without assuming linearity.

Our results only give an upper bound on the error in estimation, but they concur with the above-mentioned empirical findings, namely that $\hat{\Gamma}_n^{AR}$ is better than $\hat{\Gamma}_n^{BT}$ when the underlying time series satisfies an $AR(p)$ model, but not necessarily otherwise; see also our Remark 4.

In constructing either $\hat{f}^{AR}(\lambda)$ or $\hat{\Gamma}_n^{AR}$, the practitioner must choose the order p of the fitted AR(p) model. Minimizing the AIC criterion is a popular way to select the order p when fitting AR models in practice. It can be useful here as well; in fact, minimizing AIC was what McMurry and Politis (2015, Rejoinder) used in their simulations. We have also implemented a second method, the PACF empirical rule, yielding good simulation results for choosing p. Nevertheless, we can also envision an

entirely different method for choosing p that specifically optimizes the convergence of our AR-based estimators; this can be the subject of future work.

The extension to the multivariate setting is an open problem; it may be feasible in the future but it is beyond the scope of the paper at hand. To mention one difficulty, in the multivariate scenario even the notion of Γ_p is not straightforward, i.e., it is replaced by a collection of matrices; the latter can be concatenated in a single matrix framework that unfortunately loses some of the helpful features, e.g. the Toeplitz property. Nevertheless, the potential to multivariate setting is still there; this can also be the subject of future work.

Acknowledgement. Many thanks are due to Stathis Paparoditis for a discussion on the AR-approximation of nonlinear series that spans more than ten years; see e.g. Berg et al. (2010). We are also grateful to two anonymous reviewers for their helpful suggestions. This research was partially supported by NSF grant DMS 19-14556.

Data Availability Statement. There are no data related to this article.

Appendices

Theorem 1

Proof. By Lemma 2.2 of Kreiss, Paparoditis and Politis(2011), there exists some p_o and $C > 0$ (both depending on $f(\lambda)$ only), such that for all $p \geq p_o$:

$$\sum_{k=0}^{p}(1+k)^r|\phi_k^{AR} - \phi_k| \leq C \cdot \sum_{k=p+1}^{\infty}(1+k)^r|\phi_k|$$

Since

$$\begin{aligned}\sum_{k=0}^{p}(1+k)^r|\phi_k^{AR} - \phi_k| &\geq \sum_{k=0}^{p}|\phi_k^{AR} - \phi_k| \\ &\geq \sqrt{\sum_{k=0}^{p}(\phi_k^{AR} - \phi_k)^2} \\ &= ||\underline{\phi}^{AR}(p) - \underline{\phi}(p)||_2\end{aligned}$$

So

$$||\underline{\phi}^{AR}(p) - \underline{\phi}(p)||_2 \leq C \cdot \sum_{k=p+1}^{\infty}(1+k)^r|\phi_k|$$

The right hand side $\sum_{k=p+1}^{\infty}(1+k)^r|\phi_k|$ is finite following directly from Lemma 2.2 in Kreiss et al(2011). □

Lemma 1

Proof. By definition of eigenvalue of Toeplitz matrix:

$$\hat{\lambda}_p = \max_x \frac{x'\hat{\Gamma}_p x}{x'x}$$

$$\hat{\lambda}_1 = \min_x \frac{x'\hat{\Gamma}_p x}{x'x}$$

Let $\hat{f}_p(\lambda) = \frac{1}{2\pi}\sum_{k=1-p}^{p-1}\hat{\gamma}_k e^{-\mathrm{ik}\lambda}$. Since

$$\begin{aligned} x'\hat{\Gamma}_p x &= \sum_{k=0}^{p-1}\sum_{j=0}^{p-1}\hat{\gamma}_{k-j}x'_k x_j \\ &= \sum_{k=0}^{p-1}\sum_{j=0}^{p-1}[\int_0^{2\pi}\hat{f}_p(\lambda)e^{\mathrm{i(k-j)}\lambda}d\lambda]x'_k x_j \\ &= \int_0^{2\pi}|\sum_{k=0}^{p-1}x_k e^{ik\lambda}|^2\hat{f}_p(\lambda)d\lambda \end{aligned}$$

and similarly

$$\begin{aligned} x'x &= \sum_{k=0}^{p-1}|x_j|^2 \\ &= \frac{1}{2\pi}\int_0^{2\pi}|\sum_{k=0}^{p-1}x_k e^{\mathrm{ik}\lambda}|^2 d\lambda \end{aligned}$$

besides $\hat{f}_p(\lambda)$ is continuous, so it has maximum and minimum on $[0, 2\pi]$. Then we have

$$2\pi\min_{\lambda\in[0,2\pi]}\hat{f}_p(\lambda) \le \frac{\frac{1}{2\pi}\int_0^{2\pi}|\sum_{k=0}^{p-1}x_k e^{\mathrm{ik}\lambda}|^2\hat{f}_p(\lambda)d\lambda}{\frac{1}{2\pi}\int_0^{2\pi}|\sum_{k=0}^{p-1}x_k e^{\mathrm{ik}\lambda}|^2 d\lambda} = \frac{x'\hat{\Gamma}_p x}{x'x} \le 2\pi\max_{\lambda\in[0,2\pi]}\hat{f}_p(\lambda)$$

which means

$$2\pi\min_{\lambda\in[0,2\pi]}\hat{f}_p(\lambda) \le \hat{\lambda}_1 \le \hat{\lambda}_p \le 2\pi\max_{\lambda\in[0,2\pi]}\hat{f}_p(\lambda)$$

Now given all the assumptions in the theorem, with a small modification of Therorem 4 of Xiao and Wu(2011)(choose kernel function to be rectangular kernel), and considering the remark, we have

$$\lim_{n\to\infty} P(\max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - \mathbb{E}\hat{f}_p(\lambda)| \leq \frac{6}{\pi} c_\alpha \sqrt{\frac{p\log p}{n}}) = 1$$

On the other hand we have

$$|\mathbb{E}\hat{f}_p(\lambda) - f(\lambda)| \leq \frac{2}{n}\sum_{k=1}^{p} k|\gamma_k| + 2\sum_{k=p+1}^{\infty} |\gamma_k|$$

Since

$$\max_{\lambda\in[0,2\pi]} \hat{f}_p(\lambda) \leq \max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - \mathbb{E}\hat{f}_p(\lambda)| + \max_{\lambda\in[0,2\pi]} |\mathbb{E}\hat{f}_p(\lambda) - f(\lambda)| + \max_{\lambda\in[0,2\pi]} f(\lambda)$$

The bound for $\hat{\lambda}_p$ follows. Similarly, the bound for $\hat{\lambda}_1$ follows due to

$$\begin{aligned}\min_{\lambda\in[0,2\pi]} \hat{f}_p(\lambda) &\geq \min_{\lambda\in[0,2\pi]} f(\lambda) - \max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - f(\lambda)| \\ &\geq \min_{\lambda\in[0,2\pi]} f(\lambda) - \max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - \mathbb{E}\hat{f}_p(\lambda)| - \max_{\lambda\in[0,2\pi]} |\mathbb{E}\hat{f}_p(\lambda) - f(\lambda)|\end{aligned}$$

□

Corollary 1

Proof. Let $f_p(\lambda) = \frac{1}{2\pi}\sum_{k=1-p}^{p-1} \gamma_k e^{-ik\lambda}$. We have shown in the proof of lemma 1 that

$$\rho(\hat{\Gamma}_p - \Gamma_p) \leq 2\pi \max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - f_p(\lambda)|$$

On the other hand,

$$\max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - f_p(\lambda)| \leq \max_{\lambda\in[0,2\pi]} |\hat{f}_p(\lambda) - \mathbb{E}\hat{f}_p(\lambda)| + \max_{\lambda\in[0,2\pi]} |\mathbb{E}\hat{f}_p(\lambda) - f_p(\lambda)|$$

Since

$$|\mathbb{E}\hat{f}_p(\lambda) - f_p(\lambda)| \leq \frac{2}{n}\sum_{k=1}^{p} k|\gamma_k|$$

The result follows using similar reasoning as proof of Lemma 1. □

Theorem 2

Proof.

$$\begin{aligned}||\hat{\underline{\phi}}(p) - \underline{\phi}^{AR}(p)||_2 &= ||\hat{\Gamma}_p^{-1}\hat{\underline{\gamma}}(p) - \Gamma_p^{-1}\underline{\gamma}(p)||_2 \\ &\leq ||\hat{\Gamma}_p^{-1}(\hat{\underline{\gamma}}(p) - \underline{\gamma}(p))||_2 + ||(\hat{\Gamma}_p^{-1} - \Gamma_p^{-1})\underline{\gamma}(p)||_2 \\ &\leq \rho(\hat{\Gamma}_p^{-1})||\hat{\underline{\gamma}}(p) - \underline{\gamma}(p)||_2 + \rho(\hat{\Gamma}_p^{-1})\rho(\Gamma_p^{-1})\rho(\hat{\Gamma}_p - \Gamma_p)||\underline{\gamma}(p)||_2\end{aligned}$$

Since

$$\rho(\hat{\Gamma}_p^{-1}) = \frac{1}{\hat{\lambda}_1}$$

and

$$||\underline{\gamma}(p)||_2 \leq \sum_{l=0}^{\infty} |\gamma_k| < \infty$$

and

$$\rho(\Gamma_p^{-1}) \leq \frac{1}{2\pi F_1}$$

and

$$||\hat{\underline{\gamma}}(p) - \underline{\gamma}(p)||_2 \leq \sum_{k=1}^{p} |\hat{\gamma}_k - \gamma_k| \preceq ||\hat{\Gamma}_p - \Gamma_p||_1 \leq \sqrt{p}\rho(\hat{\Gamma}_p - \Gamma_p)$$

where notation $A \preceq B$ means $A \leq B + o(B)$. Therefore,

$$\begin{aligned}||\hat{\underline{\phi}}(p) - \underline{\phi}^{AR}(p)||_2 &\leq \rho(\hat{\Gamma}_p^{-1})\rho(\hat{\Gamma}_p - \Gamma_p)(||\underline{\gamma}(p)||_2\rho(\Gamma_p^{-1}) + \sqrt{p}) \\ &\leq \rho(\hat{\Gamma}_p^{-1})\rho(\hat{\Gamma}_p - \Gamma_p)(\sum_{l=0}^{\infty} |\gamma_k|\frac{1}{2\pi F_1} + \sqrt{p})\end{aligned}$$

As for $\frac{1}{\hat{\lambda}_1}$, there exists N such that when $n > N$,

$$2\pi F_1 - 12c_\alpha\sqrt{\frac{p\log p}{n}} - 2\pi b_n \geq \pi F_1$$

and the result follows when we apply Corollary 1. □

Corollary 3

Proof.

$$\begin{aligned}||\hat{\underline{\phi}}(p) - \underline{\phi}^{AR}(p)||_1 &= ||\hat{\Gamma}_p^{-1}\hat{\underline{\gamma}}(p) - \Gamma_p^{-1}\underline{\gamma}(p)||_1 \\ &\leq ||\Gamma_p^{-1}(\hat{\underline{\gamma}}(p) - \underline{\gamma}(p))||_1 + ||(\hat{\Gamma}_p^{-1} - \Gamma_p^{-1})\hat{\underline{\gamma}}(p)||_1 \\ &\leq ||\Gamma_p^{-1}||_1||\hat{\underline{\gamma}}(p) - \underline{\gamma}(p)||_1 + ||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1||\hat{\underline{\gamma}}(p)||_1\end{aligned}$$

Since

$$||\underline{\gamma}(p)||_1 \leq \sum_{l=0}^{\infty} |\gamma_k| < \infty$$

and

$$||\Gamma_p^{-1}||_1 \leq M$$

and

$$||\underline{\hat{\gamma}}(p) - \underline{\gamma}(p)||_2 \leq \sum_{k=1}^{p} |\hat{\gamma}_k - \gamma_k| \preceq ||\hat{\Gamma}_p - \Gamma_p||_1 \leq \sqrt{p}\rho(\hat{\Gamma}_p - \Gamma_p)$$

Therefore, by the proof of theorem 2 of McMurry and Politis(2010), the inverse difference is of the same order as the matrix difference, which leads to

$$||\underline{\hat{\phi}}(p) - \underline{\phi}^{AR}(p)||_1 \leq C^* p^{\frac{1}{2}} \rho(\hat{\Gamma}_p - \Gamma_p).$$

□

Theorem 3

Proof. To start with, recall the assumption that $\beta > r + \frac{1}{2}$, and that the short range dependence assumption $\Theta_\alpha(m) = O(m^{-\beta})$ implies that $\sum_{i=m}^{\infty} |\gamma_i| = O(m^{-\beta})$. We will now show that the latter further implies (a) that the sum $\sum_{k=0}^{\infty}(1+k)^r|\gamma_k|$ is finite, and (b) that $p^{\frac{1}{2}} \sum_{k=p}^{\infty}(1+k)^r|\phi_k| \to 0$.

Note that $\sum_{i=m}^{\infty} |\gamma_i| = O(m^{-\beta})$ implies that there exist $C > 0$ and M large enough such that when $m > M$,

$$\sum_{k=m}^{\infty} |\gamma_k| < Cm^{-\beta}.$$

Now we prove that $\lim_{m\to\infty} m^{\frac{1}{2}} \sum_{k=m}^{\infty}(1+k)^r|\gamma_k| = 0$. Let $n > m$, and apply Abel's summation by parts theorem as follows

$$\begin{aligned}
&m^{\frac{1}{2}} \sum_{k=m}^{n}(1+k)^r|\gamma_k| \\
=&m^{\frac{1}{2}}(1+m)^r \sum_{k=m}^{\infty} |\gamma_k| - m^{\frac{1}{2}}(1+n)^r \sum_{k=n+1}^{\infty} |\gamma_k| + m^{\frac{1}{2}} \sum_{k=m+1}^{n} (\sum_{i=k}^{\infty} |\gamma_i|((1+k)^r - k^r)) \\
\leq&m^{\frac{1}{2}}(1+m)^r \cdot C \cdot m^{-\beta} + m^{\frac{1}{2}}(1+n)^r \cdot C \cdot (1+n)^{-\beta} + m^{\frac{1}{2}} \sum_{k=m+1}^{n} (\sum_{i=k}^{\infty} |\gamma_i|((1+k)^r - k^r)) \\
=&T_1 + T_2 + T_3.
\end{aligned}$$

Note that T_1 and T_2 both converge to zero as $m, n \to \infty$ since $\beta > r + \frac{1}{2}$.
For T_3, by Lagrange mean value theorem, there exists $\xi_k \in (k, k+1)$ such that

$$\begin{aligned} T_3 &= m^{\frac{1}{2}} \sum_{k=m+1}^{n} (\sum_{i=k}^{\infty} |\gamma_i|((1+k)^r - k^r)) \\ &= m^{\frac{1}{2}} \sum_{k=m+1}^{n} (\sum_{i=k}^{\infty} |\gamma_i| \cdot r\xi_k^{r-1}) \\ &\leq Cm^{\frac{1}{2}} \cdot r \cdot 2^{r-1} \sum_{k=m+1}^{n} \frac{1}{k^{\beta+1-r}}. \end{aligned}$$

Since $\beta > r + \frac{1}{2}$, we have that T_3 goes to zero as well when $m, n \to \infty$. Therefore, $V_m = m^{\frac{1}{2}} \sum_{k=m}^{\infty} (1+k)^r |\gamma_k|$ has limit 0, implying that the sum $\sum_{k=0}^{\infty} (1+k)^r |\gamma_k|$ is finite. Finally, by Theorem 2.3 of Baxter(1962), the fact that V_m converges to zero implies that $m^{\frac{1}{2}} \sum_{k=m}^{\infty} (1+k)^r |\phi_k|$ also converges to zero. This is important in order to derive the convergence (19) from the bound (20); we now proceed in showing the latter.

Having verified that $\sum_{k=0}^{\infty} (1+k)^r |\gamma_k| < \infty$ (which is needed in order to apply Theorem 1), we proceed as follows:

$$\begin{aligned} |\hat{f}^{AR}(\lambda) - f(\lambda)| &= |\frac{\hat{\sigma}^2}{2\pi|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2} - \frac{\sigma^2}{2\pi|\Phi(e^{\mathrm{i}\lambda})|^2}| \\ &= \frac{|\hat{\sigma}^2|\Phi(e^{\mathrm{i}\lambda})|^2 - \sigma^2|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2|}{2\pi|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2|\Phi(e^{\mathrm{i}\lambda})|^2} \\ &\leq \frac{|\hat{\sigma}^2 - \sigma^2||\Phi(e^{\mathrm{i}\lambda})|^2 + \sigma^2||\hat{\Phi}(e^{\mathrm{i}\lambda}|^2 - |\Phi(e^{\mathrm{i}\lambda})|^2)|}{2\pi|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2|\Phi(e^{\mathrm{i}\lambda})|^2} \\ &= \frac{|\hat{\sigma}^2 - \sigma^2|}{2\pi|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2} + \frac{||\hat{\Phi}(e^{\mathrm{i}\lambda})|^2 - |\Phi(e^{\mathrm{i}\lambda})|^2|}{|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2 f(\lambda)} \\ &\leq \frac{|\hat{\sigma}^2 - \sigma^2|}{2\pi|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2} + \frac{|\hat{\Phi}(e^{\mathrm{i}\lambda}) - \Phi(e^{\mathrm{i}\lambda})|(|\hat{\Phi}(e^{\mathrm{i}\lambda})| + |\Phi(e^{\mathrm{i}\lambda})|)}{|\hat{\Phi}(e^{\mathrm{i}\lambda})|^2 f(\lambda)} \end{aligned}$$

Now since $f(\lambda)$ is bounded, $(\hat{\Phi}(e^{\mathrm{i}\lambda}))^{-1}$ is bounded below and $|\hat{\Phi}(e^{\mathrm{i}\lambda})| + |\hat{\Phi}(e^{\mathrm{i}\lambda})|$ is also bounded, it suffices to show that the two difference both go to zero in probability. Equivalently,

$$\sup_{\lambda} |\hat{f}^{AR}(\lambda) - f(\lambda)| \leq c_1|\hat{\sigma}^2 - \sigma^2| + c_2 \sup_{\lambda} |\hat{\Phi}(e^{\mathrm{i}\lambda}) - \Phi(e^{\mathrm{i}\lambda})|$$

where c_1 and c_2 are two constants relying only on $f(\lambda)$. For the second term on the right hand side,

$$\begin{aligned}\sup_{\lambda}|\hat{\Phi}(e^{\mathrm{i}\lambda})-\Phi(e^{\mathrm{i}\lambda})| &\leq \sup_{\lambda}\sum_{k=1}^{k}|(\hat{\phi}_k-\phi_k)e^{\mathrm{ik}\lambda}|+\sum_{k=p+1}^{\infty}|\phi_k e^{\mathrm{ik}\lambda}| \\ &\leq \sqrt{p}||\underline{\hat{\phi}}(p)-\underline{\phi}(p)||_2+\sum_{k=p+1}^{\infty}|\phi_k|.\end{aligned}$$

Under the assumption that $p=O(n^{\gamma})$, by Theorem 2 it follows that $\hat{\Phi}(e^{\mathrm{i}\lambda})$ converges in probability to $\Phi(e^{i\lambda})$.

As for σ^2, note that

$$\begin{aligned}|\hat{\sigma}^2-\sigma^2| &= |(\hat{\gamma}_0-\underline{\hat{\phi}}(p)'\underline{\hat{\gamma}}(p)-(\gamma_0-\underline{\phi}(p)'\underline{\gamma}(p)-\sum_{i=p+1}^{\infty}\phi_i\gamma_i)| \\ &\leq |\hat{\gamma}_0-\gamma_0|+||\underline{\hat{\phi}}(p)-\underline{\phi}(p)||_2||\underline{\gamma}(p)||_2+||\underline{\hat{\gamma}}(p))-\underline{\gamma}(p)||_2||\underline{\hat{\phi}}(p)||_2+\sum_{i=p+1}^{\infty}|\phi_i\gamma_i| \\ &= S_1+S_2+S_3+S_4\end{aligned}$$

Since $\hat{\gamma}_0$ is $\sqrt{n}$-consistent towards γ_0, it follows that S_1 converges to 0 in probability . Since $||\underline{\gamma}(p))||_2<\infty$, S_2 converges to 0 in probability with the rate of $O_p(\sqrt{\frac{p^2\log p}{n}})$. By Theorem 2, $||\underline{\hat{\phi}}(p)||_2$ is bounded in probability, S_3 converges to 0 in probability with the rate of $O_p(\sqrt{\frac{p^2\log p}{n}})$. Finally, since $\{\phi_i\}_{i=1}^{\infty}$and $\{\gamma_i\}_{i=1}^{\infty}$ are both absolutely summable, S_4 converges to 0 as p tends to infinity.

Now that we know $|\hat{\sigma}^2-\sigma^2|$ converges to zero and $|\hat{\Phi}(e^{\mathrm{i}\lambda})-\Phi(e^{\mathrm{i}\lambda})|$ also converges to zero. The bound (20) is now obtained by simple calculations and applying our previous theorems.

Under the additional assumption on the inverse of the autocovariance matrix, we have

$$\begin{aligned}|\hat{\Phi}(e^{\mathrm{i}\lambda})-\Phi(e^{\mathrm{i}\lambda})| &\leq \sum_{k=1}^{k}|(\hat{\phi}_k-\phi_k)e^{\mathrm{ik}\lambda}|+\sum_{k=p+1}^{\infty}|\phi_k e^{\mathrm{ik}\lambda}| \\ &\leq ||\underline{\hat{\phi}}(p)-\underline{\phi}(p)||_1+\sum_{k=p+1}^{\infty}|\phi_k|.\end{aligned}$$

as well as

$$|\hat{\sigma}^2 - \sigma^2| = |(\hat{\gamma}_0 - \underline{\hat{\phi}}(p)'\underline{\hat{\gamma}}(p) - (\gamma_0 - \underline{\phi}(p)'\underline{\gamma}(p) - \sum_{i=p+1}^{\infty} \phi_i\gamma_i)|$$

$$\leq |\hat{\gamma}_0 - \gamma_0| + ||\underline{\hat{\phi}}(p) - \underline{\phi}(p)||_1||\underline{\gamma}(p)||_1 + ||\underline{\hat{\gamma}}(p)) - \underline{\gamma}(p)||_1||\underline{\hat{\phi}}(p)||_1 + \sum_{i=p+1}^{\infty} |\phi_i\gamma_i|$$

Combining the reasoning above with Corollary 3, eq. (21) follows. □

Theorem 4

Proof. In the proof of lemma 1, we have already shown that the largest eigenvalue of a Toeplitz matrix is bounded by the maximum of its "spectral density". Now the "spectral density" of $\hat{\Gamma}_n^{AR} - \Gamma_n$ is

$$\begin{aligned}\hat{f}_f(\lambda) &= \frac{1}{2\pi}\sum_{k=-n}^{n}(\hat{\gamma}_k^{AR} - \gamma_k)e^{\mathrm{i}\lambda\mathrm{k}} \\ &= \hat{f}^{AR}(\lambda) - f(\lambda) + \sum_{k=n}^{\infty}\gamma_k e^{\mathrm{i}\lambda\mathrm{k}} - \sum_{k=n}^{\infty}\hat{\gamma}_k^{AR}e^{\mathrm{i}\lambda\mathrm{k}}\end{aligned}$$

Therefore we have

$$\begin{aligned}\rho(\hat{\Gamma}_n^{AR} - \Gamma_n) &\leq \max_{\lambda\in[0,2\pi]}|\hat{f}_f(\lambda)| \\ &\leq \max_{\lambda\in[0,2\pi]}|\hat{f}^{AR}(\lambda) - f(\lambda)| + \sum_{k=n}^{\infty}|\gamma_k| + \sum_{k=n}^{\infty}|\hat{\gamma}_k| \\ &= T_1 + T_2 + T_3\end{aligned}$$

T_1 converges to zero in probability because of Theorem 3; T_2 converges to zero as well. As for T_3, since with probability one all (estimated) roots will be different, we can use equation (3.3.10) of Brockwell and Davis(1991) and Lemma 2.3 of Kreiss, Paparoditis and Politis(2011) to yield:

$$
\begin{aligned}
|T_3| &= \sum_{k=n}^{\infty} |\hat{\gamma}_k^{AR}| \\
&\leq \sum_{k=n}^{\infty} |C_\beta p (\frac{p}{p+1})^{-k}| \\
&\leq C_\beta p(p+1)(\frac{p}{p+1})^{n+1} \\
&= C_\beta p^2 (\frac{p}{p+1})^{p\frac{n}{p}} \\
&\leq C_\beta p^2 (\frac{1}{2})^{\frac{n}{p}}.
\end{aligned}
$$

So as long as $p = O(n^\gamma)$, T_3 goes to zero (and fast). In all, we have that the operator norm of the difference converges to zero in probability.

□

References

Baxter G. 1962 An Asymptotic Result for the Finite Predictor. *Mathematica Scandinavica*, 10, 137-144

Berg A, Paparoditis E and Politis DN. 2010. A bootstrap test for time series. linearity, *J. Statist. Plann. Infer.*, 140(12), 3841-3857

Berk KN. 1974. Consistent Autoregressive Spectral Estimates. *Annals of Statistics*, 59, 817-858

Brockwell PJ and Davis RA. 1991. *Time Series: Theory and Methods, 2nd Ed.*. Springer, New York

Gray RM. 2006. Toeplitz and Circulant Matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3), 155-239

Hannan EJ.1960. *Time Series Analysis*. Methuen, London

Kay, SM. 1988. *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, NJ

Kreiss JP, Paparoditis E and Politis DN. 2011. On the Range of Validity of the Autoregressive Sieve Bootstrap. *Annals of Statistics*, 39(4), 2103-2130

McElroy TS and Politis DN. 2020. *Time Series: A First Course with Bootstrap Starter.* CRC Press, Boca Raton.

McMurry TL and Politis DN. 2011. Banded and tapered estimates of autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31, 471-482

McMurry TL and Politis DN. 2015. High-dimensional autocovariance matrices and optimal linear prediction. *Electronic Journal of Statistics*, 9, 753-788; 2015. Rejoinder, 9, 811-822

McMurry TL and Politis DN. 2018. Estimating MA parameters through factorization of the autocovariance matrix and an MA-sieve bootstrap. *Journal of Time Series Analysis*, 39(3), 433-446

Paparoditis E and Politis DN. 2012 Nonlinear spectral density estimation: thresholding the correlogram. *Journal of Time Series Analysis*, 33(3), 386-397

Parzen E. 1957. On Consistent Estimates of the Spectrum of a Stationary Time Series, *Annals of Mathematical Statistics*, 28(2), 329-348.

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

Politis DN.2015. *Model-Free Prediction and Regression: a Transformation-based Approach to Inference*. Springer, New York

Politis DN and Romano JP. 1995. Bias-Corrected Nonparametric Spectral Estimation *Journal of Time Series Analysis*, 16(1), 67-104

Schlogl A. 2006. A comparison of Multivariate Autoregressive Estimators. *Signal Processing*, 86, 2426-2429

Shao X and Wu WB. 2007. Asymptotic spectral theory for nonlinear time series. *Annals of Statistics*, 35, 1773-1801

Wu WB and Pourahmadi M. 2009. Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 19(4), 1755-1768

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

Xiao H and Wu WB. 2011. Covariance matrix estimation for stationary time series. *Annals of Statistics*, 40, 466-493

Xiao H and Wu WB. 2013. Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications*, 123, 2899–2920