**JIRSS (2011)**
**Vol. 10, No. 2, pp 141-166**

# Regularized Autoregressive Multiple Frequency Estimation

**Bei Chen, Yulia R. Gel**

Department of Statistics and Actuarial Science, University of Waterloo, Canada.

**Abstract.** The paper addresses a problem of tracking multiple number of frequencies using Regularized Autoregressive (RAR) approximation. The RAR procedure allows to decrease approximation bias, comparing to other AR-based frequency detection methods, while still providing competitive variance of sample estimates. We show that the RAR estimates of multiple periodicities are consistent in probability and illustrate dynamics of RAR in respect to sample size and signal-to-noise ration by simulations.

**Keywords.** Autoregressive approximation, frequency tracking, least squares method, periodic time series, regularization.

**MSC:** 62F12, 93E10, 60G35, 37M10, 94A12.

## 1 Introduction

The problem of tracking unknown frequencies is widely encountered in a variety of applications, ranging from speech recognition in electrical engineering to search for pulsars in astronomy. Although the topic has been explored for many years (see, for example, Prony, 1795; Pisarenko, 1973; Hannan and Huang, 1993), it continues to attract considerable attention in statistical and engineering literature (Chen et al., 2000; Song and Li, 2006; Duan et al., 2010; Elasmi-Ksibi et al., 2010; Liu et

Bei Chen(bchen@uwaterloo.ca), Yulia R. Gel (✉)(ygl@math.uwaterloo.ca)
Received: April, 2011; September, 2011

al., 2011). (For detailed review and historical perspective see Brillinger, 1987, Quinn and Hannan, 2001). Among various methods in the existing literature, the autoregressive (AR) frequency estimation is one of the most popular approaches due to its computational ease and theoretical convenience (Truft and Kumaresan, 1982; Mackisack and Poskitt, 1989, 1990; Hannan and Quinn, 2001). However, it is well known that the AR-based frequency estimates are asymptotically biased when the order $k$ of an approximating AR model is fixed (Stoica et al., 1987). A simple remedy is to allow the AR order $k \to \infty$ as sample size $n \to \infty$. This, however, may lead to deficiency in estimating a covariance matrix and also implies that a new model order needs to be re-selected upon the arrival of new observations and all the earlier estimated AR parameters need to be re-calculated.

In order to avoid such shortcomings, Chen and Gel (2010) introduce an alternative approach, so-called regularized AR (RAR) approximation, to detect hidden frequencies. The idea of RAR is to regularize a sample autocovariance matrix by a ridge operator of a nuclear type, which allows to fit a "longer" AR model than the one suggested by Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), hence, reducing approximation bias, and then to recursively estimate AR parameters using the Regularized Least Square (RLS) method (Gel and Fomin, 2001, and Gel and Barabanov, 2007). Note that the proposed RAR procedure falls under the "large $k$ – small $n$" framework but in a time series context. Indeed, the order of the RAR approximation can be potentially very close to $n$, so the regularization technique is particularly crucial to avoid deficiency in model identification. With the help of the nuclear ridge regularizer, RAR allows to estimate AR parameters with different level of accuracy, while the number of estimated parameters grows with the sample size. Therefore, the repeated model selection and parameter estimation are avoided as the sample size increases, which makes the RAR procedure especially attractive for online modeling when the observed sample size is unknown a-priori.

In this paper we generalize the results of Mackisack and Poskitt (1989, 1990) and Chen and Gel (2010) to a case of tracking multiple frequencies. In particular, we show that the RAR estimates of multiple frequencies are strongly consistent and asymptotically normally distributed. We also illustrate performance of RAR by numerical experiments and a case study on the sunspot data. The paper is organized as follows. In the next section, we review the RAR frequency estimation procedure. In section 3, asymptotic properties of the RAR estimates are derived. In section 4, we present an extensive simulation studies on estimation of fixed multiple frequencies. The paper is concluded by discussion in section 5.

# 2 Regularized AR frequency estimation

Consider a mixed-spectrum process $\{Y_t, t \in \mathbb{Z}\}$

$$Y_t = X_t + \epsilon_t \text{ and } X_t = \sum_{j=1}^{q} \rho_j \cos(\omega_j t + \phi_j), \tag{1}$$

where $\rho_j$ and $\omega_j$ are constants with $\rho_j > 0$ and $0 < \omega_1 < \ldots < \omega_q < \pi$; $\phi_j$ are independently identically distributed (i.i.d.) random variables uniformly distributed on $[0, 2\pi)$; $\{\epsilon_t\}$ are i.i.d. random variables with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 < \infty$. Assume that $q \geq 1$ is known, and $\{\epsilon_t\}$ is independent of $\{\phi_j\}$ and hence of $\{X_t\}$. Given observations $\{Y_1, \ldots, Y_n\}$, our goal is to estimate the frequencies

$$\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)'. \tag{2}$$

First, let us review the RAR approach for estimating $\boldsymbol{\omega}$. Consider an AR($k$) model

$$a(B)Y_t = \nu_{k,t} \tag{3}$$

where $B$ is a backward shift operator ($BY_t = Y_{t-1}$) and $a(z) = 1 + a_1 z + \ldots + a_k z^k$ is a polynomial of degree $k$. The AR model (3) can be written in a state-space form:

$$Y_t = \boldsymbol{\Phi}'_{k,t-1}\boldsymbol{\tau}_k + \nu_{k,t}, \tag{4}$$

where $\boldsymbol{\Phi}_{k,t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k})'$ and $\boldsymbol{\tau}_k = -(a_1, a_2, \ldots, a_k)'$. The RAR frequency estimation procedure consists of the following three steps:

- **Step 1**: Approximate $\{Y_1, \ldots, Y_n\}$ by a "long" AR($k$) process whose order $k \to \infty$ when $n \to \infty$ and $k$ may substantially exceed the model order suggested by AIC and BIC, i.e. $k \gg \log n$.

- **Step 2**: Estimate the vector of unknown AR parameters $\boldsymbol{\tau}_k$ by the iterative RLS method

$$\begin{aligned}\hat{\boldsymbol{\tau}}_{k,n+1} &= \hat{\boldsymbol{\tau}}_{k,n} + \boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n}(1 + \boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1})^{-1}(Y_{n+1} - \boldsymbol{\Phi}'_{k,n}\hat{\boldsymbol{\tau}}_{k,n})\\ \boldsymbol{\gamma}^{\varepsilon}_{k,n+1} &= \boldsymbol{\gamma}^{\varepsilon}_{k,n} - \boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1}(1 + \boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1})^{-1}\boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}\end{aligned} \tag{5}$$

with initial conditions $\hat{\boldsymbol{\tau}}_{k,0} = 0$ and $\boldsymbol{\gamma}^{\varepsilon}_{k,0} = (\varepsilon \boldsymbol{\Lambda_k})^{-1}$. The matrix $\boldsymbol{\gamma}^{\varepsilon}_{k,n}$ is inverse to the sample information matrix $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$, i.e. $\boldsymbol{\gamma}^{\varepsilon}_{k,n} = (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}$, where

$$\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} = \hat{\boldsymbol{R}}_{k,n} + \varepsilon \boldsymbol{\Lambda}_k \quad \text{with} \quad \hat{\boldsymbol{R}}_{k,n} = \sum_{t=1}^{n} \boldsymbol{\Phi}_{k,t} \boldsymbol{\Phi}'_{k,t} \tag{6}$$

and $\boldsymbol{\Lambda}_k = \mathrm{diag}\{\mathrm{e}^{\mu_\mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$, $\mu_j > 0$, is a ridge regularizer of a nuclear form. Note that $n^{-1}\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$ is a sample estimate of the covariance matrix $\boldsymbol{R}_k = \{r_{i-j}\}_{i,j=0}^{k-1}$, where $r_h = E(Y_t Y_{t+h})$ is defined as the theoretical autocovariance function (ACVF).

- **Step 3**: Let $\{\hat{\beta}_j e^{\pm i \hat{\omega}_{k,j}}\}_{j=1}^{q}$ denote the $2q$ roots of $\hat{a}(z)$ which are closest to the unit circle, then the angular positions of these roots

$$\hat{\boldsymbol{\omega}}_k = (\hat{\omega}_{k,1}, \ldots, \hat{\omega}_{k,q})' \tag{7}$$

are the estimates of hidden frequencies. Equivalently, $\boldsymbol{\omega}$ can be estimated by locating the minimum of the transfer function

$$\hat{f}_k(\theta) = \left|\hat{a}(e^{i\theta})\right|^2 = \left|\sum_{j=0}^{k} \hat{a}_j (e^{ij\theta})\right|^2. \tag{8}$$

Note that the model order $k$ can be a priori selected to be equal to (or even to exceed) a potential upper bound of all practically fittable AR models, given the current sample size $n$. Since we employ a nuclear form of ridge regularization $\boldsymbol{\Lambda}_k$, the AR parameters are obtained with different precision, while the number of accurately identified parameters smoothly grows with the sample size. Hence, RAR can be viewed as a smoothed version of model selection.

## 3 Asymptotic properties of RAR estimates

In this section, we extend the results of Mackisack and Poskitt (1989 and 1990) and prove the strong consistency and the asymptotic normality of the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$. First, our goal is to show that the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$ converge almost surely (a.s.) to the vector of unknown frequencies $\boldsymbol{\omega}$. The proof of this result is based on strict consistency of the RLS estimates of autoregressive parameters. Note that by Theorem 1 of Stoica et al. (1987), we have

$$\boldsymbol{\tau}_k = -\frac{2}{k}\left(\sum_{j=1}^{q}\cos(\omega_j), \sum_{j=1}^{q}\cos(2\omega_j), \ldots, \sum_{j=1}^{q}\cos(k\omega_j)\right)' + O\left(\frac{1}{k^2}\right). \quad (9)$$

**Theorem 3.1.** *Let $Y_t$ be generated by (1). Let $\boldsymbol{q}_k = (q_1, \ldots, q_k)' \in \mathbb{R}^k$ denote a $k \times 1$ vector satisfying $\|\boldsymbol{q}_k\| \leq O(k^{1/2})$. Assume that $\mu(n) = o(\log n)$. If $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, then*

*(1)* $\left|\boldsymbol{q}_k'(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k)\right| \to 0$ *a.s.*

*(2)* $\sup_{\theta\in(0,\pi)} \left||\hat{a}(e^{i\theta})|^2 - |a(e^{i\theta})|^2\right| = o(1)$ *a.s.*

(The proof of Theorem 3.1 is given in the Appendix.)

Based on the consistency of the RAR parameter estimates $\hat{\boldsymbol{\tau}}_k$ in Theorem 3.1, we derive the following almost sure convergence result of $\hat{\boldsymbol{\omega}}_k$.

**Theorem 3.2.** *Under the assumptions of Theorem 3.1, $\hat{\boldsymbol{\omega}}_k \to \boldsymbol{\omega}$ almost surely as $n \to \infty$, $k \to \infty$ such that $k^2/n \to 0$.*

(The proof of Theorem 3.2 is given in the Appendix.)

Second, we verify the asymptotic normality of the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$ and start from deriving the asymptotic distribution of $\hat{\boldsymbol{\tau}}_k$.

**Theorem 3.3.** *Under the conditions of Theorem 3.1 and $E(\epsilon_t^4) = \kappa\sigma^4 < \infty$, if $k, n \to \infty$ and $k^2/n \to 0$,*

$$\sqrt{T}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) \to N(0, \boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1}),$$

*where*

$$\boldsymbol{M}_k = \begin{pmatrix} a_1 & a_2 & \ldots & a_k & 0 \\ a_2 & a_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{k-1} & a_k & \ldots & 0 & 0 \\ a_k & 0 & \ldots & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & a_0 & \ldots & 0 & 0 \\ 0 & a_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & a_{k-2} & \ldots & a_0 & 0 \\ 0 & a_{k-1} & \ldots & a_1 & a_0 \end{pmatrix}. \quad (10)$$

*Here $\boldsymbol{\Sigma}_k = \{\sigma_{ij}^{\varepsilon}\}_{i,j=0}^{k}$, where*

$$\sigma_{ij}^{\varepsilon} = \begin{cases} \delta_{i,j}\sigma^4 + \sigma^2\sum_{s=1}^{q} 2\rho_s^2\cos(\omega_s i)\cos(\omega_s j), & i,j \neq 0, \\ (\kappa-1)\sigma^4 + \sigma^2\sum_{s=1}^{q} 2\rho_s^2, & i,j = 0. \end{cases} \quad (11)$$

(See the Appendix for the proof of Theorem 3.3.)

Using the result of Theorem 3.3, we derive the asymptotic normality of $\hat{\boldsymbol{\omega}}_k$. Let $a^*(z)$ be a polynomial of degree $k$, i.e. $a^*(z) = 1 + a_1^* z + \ldots + a_k^* z^k$, and $\boldsymbol{\tau}_k^* = -(a_1^*, \ldots, a_k^*)$, such that

$$\boldsymbol{\tau}_k^* = \boldsymbol{R}_k^+ \boldsymbol{r}_k, \tag{12}$$

where $\boldsymbol{R}_k^+$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{R}_k$. The results of Stoica et al.(1989) imply that

$$a^*(z) = B^*(z)A(z), \tag{13}$$

where $A(z) = \prod_{s=1}^q (1 - 2\cos\omega_s z + z^2)$ and $B^*(z)$ is a monic polynomial of degree $(k - 2q)$ uniquely defined by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B^*(e^{i\omega})\right|^2 \left|A(e^{i\omega})\right|^2 d\omega = \min_{\{B\}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B(e^{i\omega})\right|^2 \left|A(e^{i\omega})\right|^2 d\omega. \tag{14}$$

Note that $A(z)$ has a pair of roots located on the unit circle at $e^{\pm i\omega_s}$, $s = 1, \ldots, q$. The remaining roots of $a^*(z)$, which are the roots of $B^*(z)$, are located outside the unit circle. For the large value of $k$, the roots of $B^*(z)$ may be located very close to the unit circle, which eventually causes spurious frequency estimates. (We discuss trimming algorithm of such spurious roots in the next section.) The following theorem states the result on asymptotic normality of $\hat{\boldsymbol{\omega}}_k$.

**Theorem 3.4.** *Under the conditions of Theorem 3.1 and if $k^2 \geqslant cn^{1-\delta}$, for $0 < \delta < 2/3$, such that $k^2/n \to 0$, then*

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) \to N(0, \boldsymbol{F}\boldsymbol{G}\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1}\boldsymbol{G}'\boldsymbol{F}')$$

*in distribution, where*

$$\boldsymbol{F} = \begin{pmatrix} \frac{\psi_1}{(\theta_1^2+\psi_1^2)} & 0 & \ldots & 0 & -\frac{\theta_1}{(\theta_1^2+\psi_1^2)} & 0 & \ldots & 0 \\ 0 & \frac{\psi_2}{(\theta_2^2+\psi_2^2)} & \ldots & 0 & 0 & -\frac{\theta_2}{(\theta_2^2+\psi_2^2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \frac{\psi_q}{(\theta_q^2+\psi_q^2)} & 0 & 0 & \ldots & -\frac{\theta_q}{(\theta_q^2+\psi_q^2)} \end{pmatrix}, \tag{15}$$

$$\begin{aligned} \boldsymbol{G} &= (\boldsymbol{h}_1', \ldots, \boldsymbol{h}_q', \boldsymbol{g}_1', \ldots, \boldsymbol{g}_q')', \quad \textit{for} \quad s = 1, \ldots, q, \\ \theta_s &= (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\boldsymbol{\tau}_k^*, \\ \psi_s &= (\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\boldsymbol{\tau}_k^*, \\ \boldsymbol{h}_s &= (\cos\omega_s, \cos 2\omega_s, \ldots, \cos k\omega_s)', \\ \boldsymbol{g}_s &= (\sin\omega_s, \sin 2\omega_s, \ldots, \sin k\omega_s)'. \end{aligned} \tag{16}$$

(The proof of Theorem 3.4 is given in the Appendix.)

Hence, we conclude that the RAR estimates $\hat{\omega}_k$ of multiple frequencies converge almost surely and are asymptotically normally distributed. However, in practice, RLS allows to avoid the model order selection step, and to fit a much longer AR model compared to the one chosen by AIC.

**Remark 3.1.** Note that any asymptotic variance of the frequency estimate based on autoregressive approximation of order $k$, with or without regularization, does depend on $k$. See the asymptotic variance in the regularized case (RAR) stated in Theorem 3.3 and its unregularized analogue (MP) shown in Theorem 2 and 2' of Mackisack and Poskitt (1990) (in a single frequency case).

**Remark 3.2.** In view of the results of Li et al. (1994), Lau et al. (2002) and the classical results of Bartlett (see Bartlett, 1955; Brockwell and Davis, 2006, Theorem 7.2.1 and Proposition 7.3.1), the asymptotic results of Theorems 3.1–3.4 can be extended under a more general condition when $\{\epsilon_t\}$ is a linear process of the form

$$\epsilon_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}, \tag{17}$$

where $\{\xi_t\}$ are i.i.d random variables with $E(\xi_t) = 0$, $E(\xi_t^2) = \sigma_\xi^2$ and $\{\psi_t\}$ is an absolutely summable deterministic sequence with $\sum |\psi_j| < \infty$. In this case $\{y_t\}$ is referred to as mixed-spectrum process.

**Remark 3.3.** Note that here we consider the case of a "soft" regularized estimation of multiple frequencies, i.e. $\mu(n) = o(\log n)$, which in asymptotics leads to the same restriction on the AR approximation order $k$ as in the unregularized case of a singular frequency, derived by Mackisack and Poskitt (1989 and 1990). Echoing the discussion of Lau et al. (2002) on behavior of an AR($k$) spectral estimator when both $k$ and $n \to \infty$, an interesting question is: Can we increase the rate of AR approximation $k$ and adequately balance the bias-variance issue with the help of a "stronger" regularizer? I.e., can Regularized AR($k$) estimator go beyond $k^2/n \to \infty$? For example, the potential regularizer candidates are nuclear exponential or polynomial operators with increasing diagonal values, i.e. $\mathbf{\Lambda}_k = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ (Gel and Barabanov, 2007) or $\mathbf{\Lambda}_k = \mathrm{diag}\{\mathrm{j}^{\mathrm{p}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ (Barabanov and Gel, 2005). After numerous unsuccessful attempts to derive asymptotic properties of RAR approximant of a higher order order, we have decided to leave it as a conjecture:

**Conjecture 3.1.** *Let $Y_t$ be generated by (1) and let $\mathbf{\Lambda}_k = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$. As $k \to \infty$, $n \to \infty$ such that $k/n \to \infty$, then for any $\delta \in (0,1)$, it holds in probability:*

*(1)* $V_t = n^{-\delta}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k)' \hat{\boldsymbol{R}}^{\varepsilon}_{k,n}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) \to 0,$

*(2)* $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} \geq C > 0,$

*which implies that* $n^{(1-\delta)/2}\left|(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k)\right| \to 0$ *in probability.*

While we were able to derive condition (2), we could not show condition (1) and, hence, leave it as an open problem (see the Appendix for proof of condition (2)). This conjecture is empirically supported by a number of simulations discussed in the Section 4.

# 4 Numerical examples

In this section, we demonstrate the performance of the RAR frequency estimation by simulation studies, using a "stronger" regularizer $\Lambda_k = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ and $\mu$ selected by a cross-validation procedure. As discussed in Chen and Gel (2010), spurious roots typically occur when the AR approximation order $k$ is high, which results in false frequency estimates. In order to reduce this effect and increase the accuracy of the frequency estimates, we apply the robust trimming algorithm (RTA) (Chen and Gel, 2010) in our simulation studies. Here we consider two-sinusoid processes with different combinations of amplitudes and frequencies are considered, as shown in Table 1 (Stoica et al., 1989a) and $\{\epsilon_t\}$ are i.i.d N(0,1).

| Case | $\rho_1$ | $\rho_2$ | $\omega_1$ | $\omega_2$ |
|---|---|---|---|---|
| 1 | $10\sqrt{2}$ | $10\sqrt{2}$ | $0.53\pi$ | $0.23\pi$ |
| 2 | $10\sqrt{2}$ | $10\sqrt{2}$ | $0.33\pi$ | $0.23\pi$ |
| 3 | $10\sqrt{2}$ | $10\sqrt{2}$ | $0.26\pi$ | $0.23\pi$ |
| 4 | $\sqrt{2}$ | $\sqrt{2}$ | $0.53\pi$ | $0.23\pi$ |
| 5 | $\sqrt{2}$ | $\sqrt{2}$ | $0.33\pi$ | $0.23\pi$ |
| 6 | $\sqrt{2}$ | $\sqrt{2}$ | $0.26\pi$ | $0.23\pi$ |
| 7 | $\sqrt{20}$ | $\sqrt{2}$ | $0.53\pi$ | $0.23\pi$ |
| 8 | $\sqrt{20}$ | $\sqrt{2}$ | $0.33\pi$ | $0.23\pi$ |
| 9 | $\sqrt{20}$ | $\sqrt{2}$ | $0.26\pi$ | $0.23\pi$ |

Table 1: Amplitudes and frequencies of simulation studies.

First, we investigate the variances of $\hat{\omega}_{k,1}$ and $\hat{\omega}_{k,2}$, denoted as $Var(\hat{\omega}_{k,1})$ and $Var(\hat{\omega}_{k,2})$, under different signal-to-noise ratio (SNR) in Cases 1-3. Note that SNR compares the level of a desired signal to the level of a background noise, which is defined as

$$\text{SNR}(j) = 10 \log_{10} \frac{0.5\rho^2}{j\sigma^2} \text{ dB}. \tag{18}$$

Here, we take $\rho = 10\sqrt{2}$ and $j = 9.5, 9, 8.5, 8, \ldots, 1, 0.5$. In all considered cases, $Var(\hat{\omega}_{k,1})$ and $Var(\hat{\omega}_{k,2})$ are compared to the Cramer-Rao Lower Bound (CRLB) (Stoica et al., 1989b) where

$$\text{CRBL}_1 = 24\sigma^2/(\rho_1^2 T^3) \quad \text{and} \quad \text{CRBL}_2 = 24\sigma^2/(\rho_2^2 T^3). \tag{19}$$

Since $\{\epsilon_t\}$ are assumed to be i.i.d N(0,1) in the simulated samples, $\sigma^2$ is 1.

Suppose that our sample size $n$ is 2000 measurements. Based on the first 700 observations, the cross-validation procedure (Chen and Gel, 2010; Bickel and Gel, 2011) selects an "optimal" regularizing parameter $\mu = 0.11$ and AR order $k = 80$. Figure 4 shows $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ respectively compared to CRLB while SNR increases from 0.45 to 26.02. Both $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ monotonically decrease as SNR increases and approach CRLB. Also, notice that the differences between the frequencies $\omega_1$ and $\omega_2$ in Case 1, 2 and 3 are correspondingly $0.2\pi$, $0.1\pi$ and $0.03\pi$. As the distance between frequencies decreases, the rate of convergence of $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ to CRLB also decreases.

Second, we study dynamics of $Var(\hat{\omega}_{k,1})$ and $Var(\hat{\omega}_{k,2})$ in respect to an increasing sample size, given SNR of 20dB. Due to the RAR properties, the model order $k$ and regularizing parameter $\mu$ remain the same whenever sample size changes, i.e. the previously chosen AR(80) with $\mu$ of 0.11 are employed in all cases while $T$ increases from 1000 to 5000. As shown in Figure 2-4, both $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ strictly decrease as sample size increases. Similar to Figure 4, the variances are close to CRLB when the frequencies are well-separated (Case 1, 4 and 7), while the difference becomes larger as the frequencies are closer. Note that the magnitude of variance negatively relates to the amplitude of the sinusoid and hence $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ in Figure 2 are considerably smaller than those in Figure 3 and 4.

Since RAR can be viewed as an extension of the results of Mackisack and Poskitt (1989) (from here on referred to as MP), we compare the mean square error (MSE) of RAR to that of MP under varying SNR.

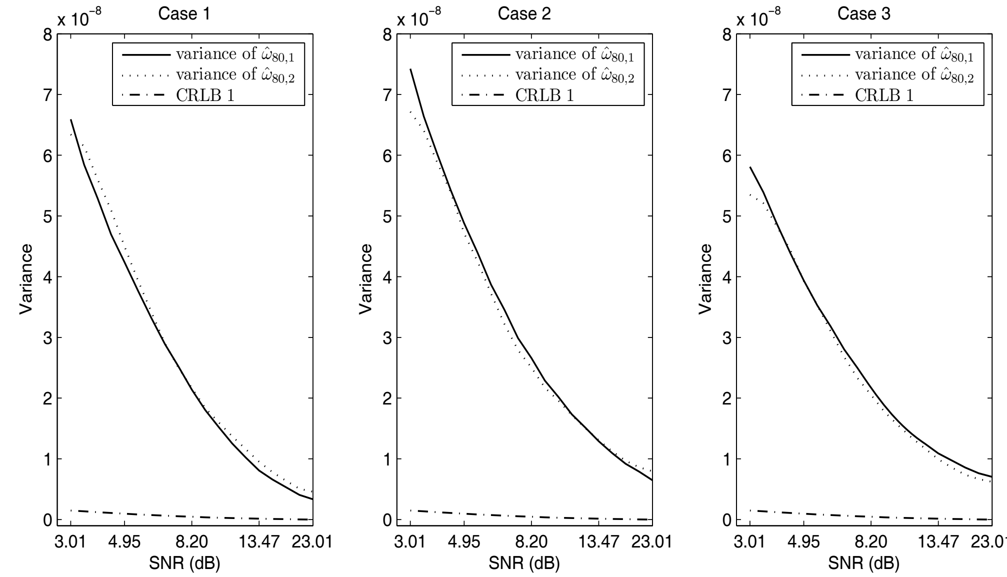


Figure 1: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying SNR.

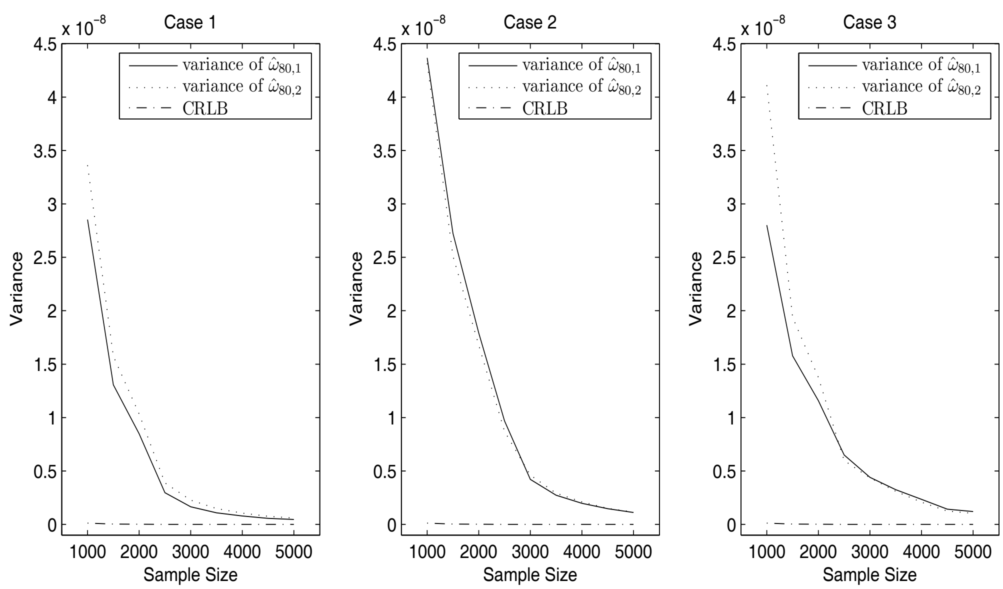


Figure 2: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \rho_2 = 10\sqrt{2}$.

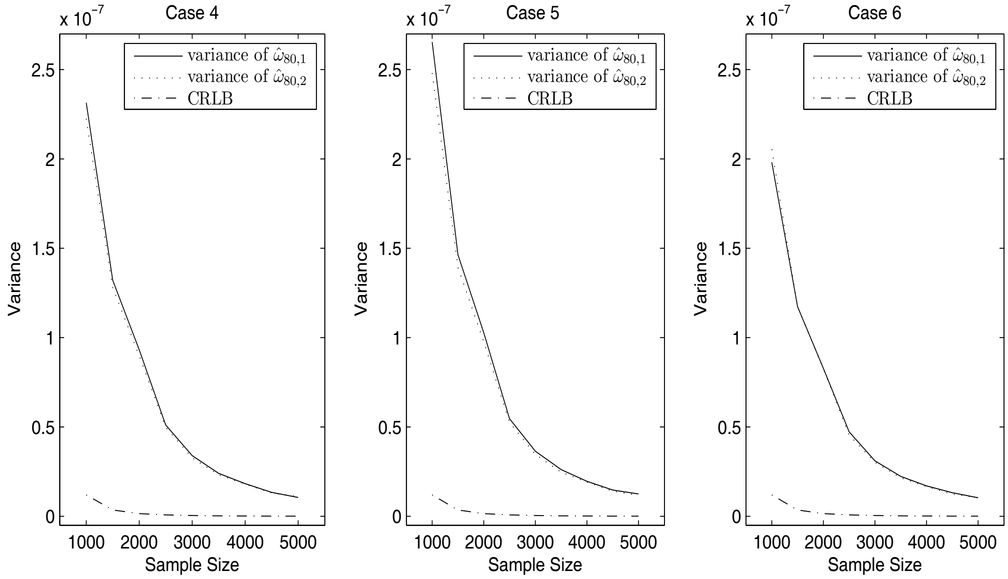


Figure 3: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \rho_2 = \sqrt{2}$.

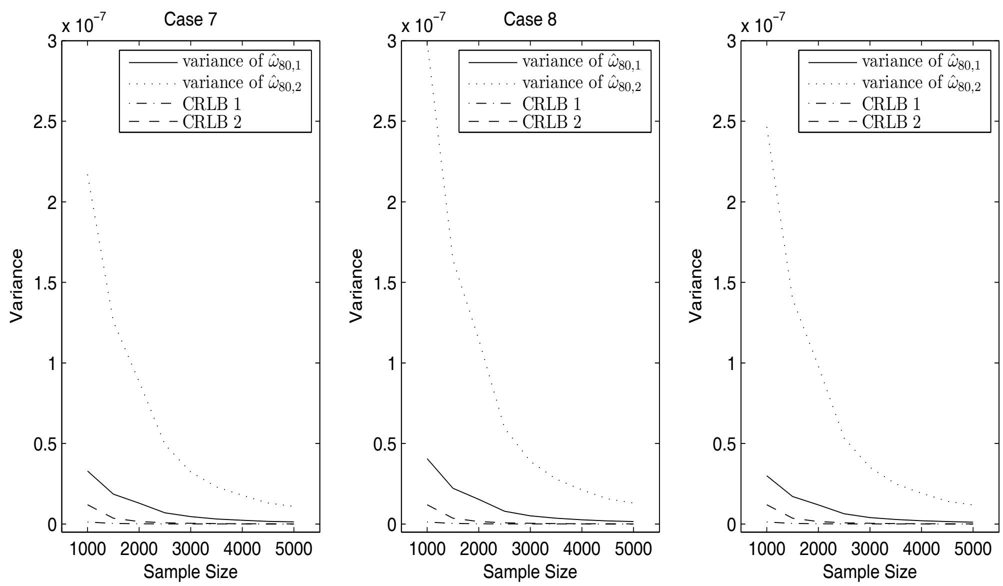


Figure 4: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \sqrt{20}$ and $\rho_2 = \sqrt{2}$.

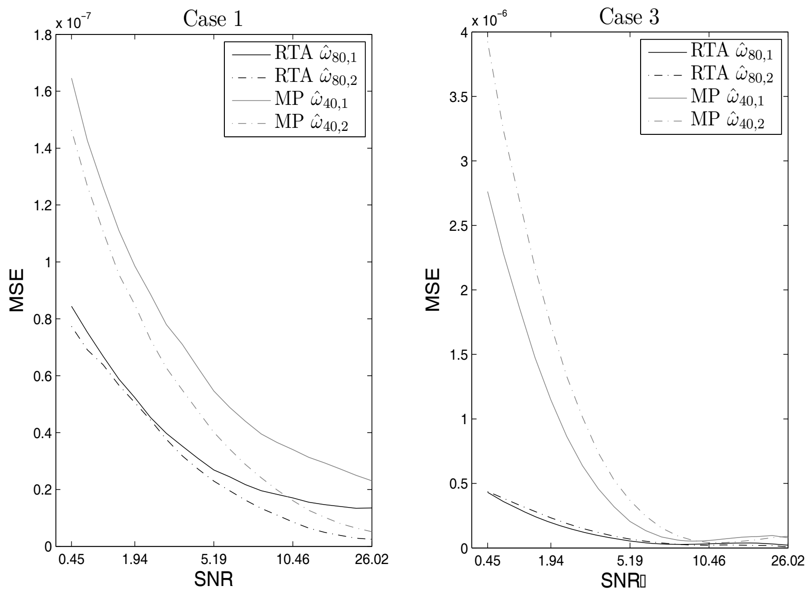


Figure 5: Comparison of MSE yielded by MP and RTA for various SNR.

Suppose that an observed sample consists of 2000 data points. MP approximates the underlying process by an AR(40) model selected by AIC. Denote a mean squared error (MSE) of the RTA and MP frequency estimates by $\mathrm{MSE}_{\mathrm{RTA}}$ and $\mathrm{MSE}_{\mathrm{MP}}$ respectively. Figure 6 illustrates the comparison of $\mathrm{MSE}_{\mathrm{RTA}}$ and $\mathrm{MSE}_{\mathrm{MP}}$ while SNR increases from 0.45 dB to 26.02 dB. From Figure 6, we find that $\mathrm{MSE}_{\mathrm{RTA}}$ is noticeably smaller than $\mathrm{MSE}_{\mathrm{MP}}$ when $\omega_1$ and $\omega_2$ are well-separated (Case 1) regardless of SNR; as well as when $\omega_1$ and $\omega_2$ are very close (Case 2) but SNR is low (SNR $<$ 8.43dB). As SNR increases, both $\mathrm{MSE}_{\mathrm{RTA}}$ and $\mathrm{MSE}_{\mathrm{MP}}$ decay exponentially and tend to converge after a certain threshold. Hence, fitting a longer AR model with robust trimming can effectively reduce MSE, especially in noisy conditions which is frequently the case for many applications.

## 5 Case Study

One of the classical examples of a periodic process is the sunspot observations. The earliest surviving record of sunspot dates from the 364

B.C., according a star catalogue by Chinese astronomer Gan De (Hockey, 1999). In order to demonstrate the proposed RAR method, we take a sample of annual sunspot observations from 1700 to 1988 (see Figure 6) and then apply the RAR procedure to estimate the hidden frequency. Using cross-validation, we select an AR(25) model with regularizing parameter $\mu = 0.1$. As a result, the RAR frequency estimate is 0.5721.

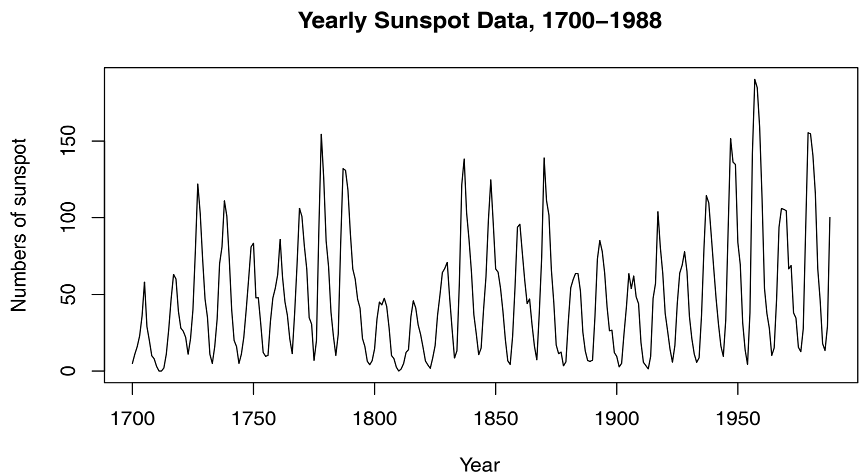


Figure 6: Yearly Sunspot data from 1700 to 1988.

Since RAR can be viewed as an extension of the results of MacKisack and Poskitt (1989) (from here on referred to as MP), we also apply the MP procedure to the sunspot observations for comparison purpose. The AIC selects an AR(9) model and consequently, the MP frequency estimate is 0.3634. In fact, it is well-known that the sunspot populations rise and fall on an irregular cycle of 11 years, i.e., the hidden frequency is equal to 0.5712. Clearly, the estimation error by RAR is about 0.43% of that by MP.

## 6 Discussion

This paper generalizes the result of Chen and Gel (2010) on regularized autoregressive (RAR) frequency estimation to a case of multiple unknown periodicities. We show that the RAR estimates of multiple frequencies are strongly consistent and asymptotically distributed. Since the idea of RAR is to approximate generalized spectral density of an observed periodic time series by a spectral density of a "long" autoregressive model whose order is substantially higher than suggested by AIC or BIC, we encounter a "large $k$-small $n$" problem but in a time series context. We approach this problem by a nuclear-type ridge regularization of a sample autocovariance matrix and choose an "optimal" regularizer with cross-validation (Chen and Gel, 2010; Bickel and Gel, 2011). Our simulation results indicate that as sample size and/or signal-to-noise ratio increases, the RAR frequency estimates approach the Cramer-Rao Lower Bound, and convergence rate is faster if frequencies are father apart. Since RAR enables us to avoid frequent re-estimation of approximating model order and parameters, the new procedure is relatively computationally inexpensive and, hence, feasible for online tracking of unknown multiple frequencies.

The proposed method can be extended for a case of colored generating noise as it is a more realistic assumption for a number of applications, e.g. astronomy and speech recognition. Another interesting future extension consists of employing banding and thresholding as regularization techniques as well as exploring bootstrap-based selection of an an "optimal" regularizer.

## Acknowledgement

The research of Yulia R. Gel was supported by grants from the National Science and Engineering Research Council of Canada. Bei Chen was partially supported by the WatRisk scholarship. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: http://www.sharcnet.ca).

## References

Bartlett, M. S. (1955), An introduction to stochastic processes. Cambridge University Press.

Barabanov, Y. and Gel, Y. R. (2005), Strong consistency of the Least-Squares method with a polynomial regularizer for infinite AR models. Automation and Remote Control, **1**(66), 92–108.

Bickel, P. and Gel, Y. R. (2011), Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. To appear in the Journal of Royal Statistical Society, Series. B.

Brillinger, D. R. (1987), Fitting cosines: some procedures and some physical examples. in: B. MacNeill, G.J. Umphrey, Applied Probability, Stochastic Processes, and Sampling Theory, Ed.I., Reidel, Boston, 75-100.

Brockwell, P. J. and Davis, R. A. (2006), Time Series: Theory and Methods. New York: Springer.

Chen, B. and Gel, Y. R. (2010), Autoregressive frequency detection using least squares. Journal of Multivariate Analysis, **101**(7), 1712-1727.

Chen, Z., Wu, K., and Dahlhaus, R. (2000), Hidden Frequency Estimation with Data Tapers. Journal of Time Series Analysis, **21**(2), 113-142.

Duan, Z., Pardo, B., and Zhang, C. (2010), Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions. IEEE Transactions on Audio, Speech, and Language Processing, **18**(8), 2121-2133.

Elasmi-Ksibi, R., Besbes, H., and López-Valcarce, R. (2010), Frequency estimation of real-valued single-tone in colored noise using multiple autocorrelation lags. Signal Processing, **90**(7), 2303-2307.

Gel, Y. R. and Fomin, V. N. (2001), Identification of an unstable ARMA equation. Mathematical Problems in Engineering, **7**, 97-112.

Gel, Y. R. and Barabanov, A. (2007), Strong consistency of the regularized least-squares estimates of infinite autoregressive models. Journal of Statistical Planning and Inference, **137**, 1260-1277.

Hannan, E. J. and Huang, D. (1993), On-Line Frequency Estimation. Journal of Time Series Analysis, **14**(2), 147-161.

Hockey, T. A. (1999), Galileo's Planet: Observing Jupiter Before Photography. London: Institute of Physics Publishing.

Lau, S. S., Sherman, P. J., and White, L. B. (2002), Asymptotic statistical properties of AR spectral estimators for processes with mixed spectra. IEEE Transactions on Information Theory, **48**(4), 909-917.

Li, T. H., Kedem, B., and Yakowitz, S. (1994), Asymptotic normality of sample autocovariances with an applications in frequency estimation. Stochastic Processes and their Applications, **52**, 329-349.

Liu, Y., Nie, Z., Zhao, Z., and Liu, Q. (2011), Generalization of iterative Fourier interpolation algorithms for single frequency estimation. Digital Signal Processing, **21**(1), 141-149.

MacKisack, M. S. and Poskitt, D. S. (1989), Autoregressive frequency estimation. Biometrika, **76**(3), 565-575.

MacKisack, M. S. and Poskitt, D. S. (1990), Some properties of autoregressive estimates for processes with mixed spectra. Journal of Time Series Analysis, **11**(4), 325-337.

Pisarenko, V. F. (1973), The retrieval of harmonic from a covariance function. Geophysical Journal International, **33**(3), 347-366.

Prony, G. R. (1795), Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l?eau et de la vapeur de l?alkool, à. différentes températures. Journal de L'École Polytechnique Floréal et Plairial, an III, **1**(22), 24-76.

Quinn, B. G. and Hannan, E. J. (2001), The Estimation and Tracking of Frequency. UK: Cambridge University Press.

Song, K. S and Li, T. H., On convergence and bias correction of a joint estimation algorithm for multiple sinusoidal frequencies. Journal of the American Statistical Association, **101**(474), 830-842.

Stoica, P., Friedlander, B., and Söderström, T. (1987), Asymptotic bias of the high-order autoregressive estimates of sinusoidal frequencies. Circuits Systems Signal Process, **3**(6), 287-298.

Stoica, P., Söderström, T., and Ti, F. N. (1989a), Overdeterminded Yule-walker Estimation of the Frequencies of Multiple Sinusoids: Accuracy Aspects. Signal Processing, **16**, 155-174.

Stoica, P., Moses, R. L., Friedlander, B., and Söderström, T. (1989b), Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements. IEEE Transactions on Acoustics, Speech, and Signal Processing, **37**(3), 378-392.

Tufts, D. W. and Kumaresan, R. (1982), Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. Proceedings of the IEEE, **70**, 975-989.

## Appendix

Let us denote covariance vectors by $\boldsymbol{r}_k = (r_1, \ldots, r_k)'$, $\boldsymbol{r}_{k,0} = (r_0, r_1, \ldots, r_k)'$. Also, denote sample ACVF by $\hat{r}_j = \frac{1}{n}\sum_{t=1}^{n-j} Y_t Y_{t+j}$, $j = 0, 1, \ldots, k$, for $k = 0, 1, ..., n-1$, which forms sample covariance vectors $\hat{\boldsymbol{r}}_k = (\hat{r}_1, \ldots, \hat{r}_k)'$ and $\hat{\boldsymbol{r}}_{k,0} = (\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_k)'$.

**Proof of Theorem 3.1.** Let $\boldsymbol{\Gamma}_k$, $\hat{\boldsymbol{\Gamma}}_{k,n}$, $\hat{\boldsymbol{\Gamma}}^{\varepsilon}_{k,n}$ and $\boldsymbol{\Lambda}^{(1)}_k$ denote the $k \times (k+1)$ matrices formed respectively from the $(k+1) \times (k+1)$ matrices $\boldsymbol{R}_k$, $\hat{\boldsymbol{R}}_{k,n}$, $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$ and $\boldsymbol{\Lambda}_k$ by deleting their first rows. Following the proof of Theorem 3 in Mackisack and Poskitt (1990), we express $(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k)$ as

$$
\begin{aligned}
(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_{k,n}) &= \gamma^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}^{\varepsilon}_k - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)' \\
&= \gamma^{\varepsilon}_{k,n}\left(\hat{\boldsymbol{\Gamma}}_{k,n} + \frac{\varepsilon \boldsymbol{\Lambda}^{(1)}_k}{n} - \boldsymbol{\Gamma}_k\right)(1 : \boldsymbol{\tau}'_k)' \\
&= \gamma^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)' + \gamma^{\varepsilon}_{k,n}\frac{\varepsilon \boldsymbol{\Lambda}^{(1)}_k}{n}(1 : \boldsymbol{\tau}'_k)'. \quad (20)
\end{aligned}
$$

Since as $n \to \infty$, $n^{-1}\varepsilon e^{\mu(n)} \to 0$ and $\hat{\boldsymbol{R}}_{k,n} \to \boldsymbol{R}_k$ a.s., by Theorem 4.1 of Houdré and Kedem (1995), we obtain

$$
\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} = \boldsymbol{R}_k + \big(\hat{\boldsymbol{R}}_{k,n} - \boldsymbol{R}_k\big) + \frac{\varepsilon \boldsymbol{\Lambda}^{(1)}_k}{n} > 0, \quad (21)
$$

and hence $\|\boldsymbol{\gamma}^{\varepsilon}_{k,n}\| \le C_1$, $C_1 \in \mathbb{R}^+$. Also, by equation (9), $\|(1 : \boldsymbol{\tau}'_k)'\| \le C_2$, $C_2 \in \mathbb{R}^+$, thus $\boldsymbol{\gamma}^{\varepsilon}_{k,n} n^{-1}\varepsilon\boldsymbol{\Lambda}^{(1)}_k(1 : \boldsymbol{\tau}'_k)' \to 0$ and consequently,

$$
(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_{k,n}) \sim \gamma^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)', \quad (22)
$$

which can be re-written as

$$
(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_{k,n}) \sim \Big(\gamma^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} - \boldsymbol{R}_k\big) + \boldsymbol{I}_k\Big)\boldsymbol{R}^{-1}_k\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k). \quad (23)
$$

Equation 21 implies that $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} \to \boldsymbol{R}_k$ a.s. again by Theorem 4.1 of Houdré and Kedem (1995). Following An el al.'s approach (1982, pp.929-930), we can show that

$$
\Big(\gamma^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} - \boldsymbol{R}_k\big) + \boldsymbol{I}_k\Big) \to \boldsymbol{I}_k, \quad \text{a.s.} \quad (24)
$$

and hence our problem is reduced to verify

$$\boldsymbol{q}_k'\boldsymbol{R}_k^{-1}\big(\hat{\boldsymbol{\Gamma}}_{k,n}-\boldsymbol{\Gamma}_k\big)(1:\boldsymbol{\tau}_k')\to 0, \quad \text{a.s.} \tag{25}$$

Since $\|\boldsymbol{R}_k^{-1}\|\le C_3$, $C_3\in\mathbb{R}^+$, we obtain

$$\|\boldsymbol{q}_k'\boldsymbol{R}_k^{-1}\| = O(\|\boldsymbol{q}_k\|) = O(k^{1/2}). \tag{26}$$

Thus, in the rest of the proof we investigate the asymptotic behavior of

$$M_{k,n} = Q_k'\big(\hat{\boldsymbol{\Gamma}}_{k,n}-\boldsymbol{\Gamma}_k\big)(1:\boldsymbol{\tau}_k'), \tag{27}$$

where $Q_k' = \boldsymbol{q}_k'\boldsymbol{R}_k^{-1} = (Q_1,\ldots,Q_k)$.

First, let us consider an element of $\big(\hat{\boldsymbol{\Gamma}}_{k,n}-\boldsymbol{\Gamma}_k\big)$, i.e, $\hat{r}_j - r_j$, $j = 0,\ldots,k$. Denote $r_j^x = E(X_tX_{t+j})$ and $\hat{r}_j^x = \frac{1}{n}\sum_{t=1}^{n-j}X_tX_{t+j}$, we have

$$\begin{aligned} r_j^x &= E\Big\{\sum_{n=1}^{q}\rho_n\cos(\omega_n t+\phi_n)\sum_{n=1}^{q}\rho_n\cos(\omega_n(t+j)+\phi_n)\Big\} \\ &= \sum_{n=1}^{q}\frac{\rho_n^2}{2}\cos(\omega_n j). \end{aligned} \tag{28}$$

and thus

$$\begin{aligned} \hat{r}_j^x - r_j^x &= \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\sum_{m=1}^{q}\rho_s\rho_m\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \\ &- \sum_{s=1}^{q}\frac{\rho_s^2}{2}\cos(\omega_s j) \\ &= \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=1}^{n-j}\cos\big(\omega_s(2t+j)+2\phi_s\big) \\ &+ \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=1}^{n-j}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big)\\
&- \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=n-j+1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big)\\
&+ \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big)\\
&- \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=n-j+1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \qquad (29)
\end{aligned}
$$

Using the complex exponential representation of cosine function, it can be shown that

$$
\left|\sum_{t=1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big)\right| \le \frac{1}{|\sin\omega_s|}, \qquad (30)
$$

and

$$
\begin{aligned}
&\left|\sum_{t=1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big)\right|\\
&\qquad \le \frac{1/2}{\big|\sin\big((\omega_s+\omega_m)/2\big)\big|} + \frac{1/2}{\big|\sin\big((\omega_s-\omega_m)/2\big)\big|} \qquad (31)
\end{aligned}
$$

for any $\phi_s$, $\phi_m$, $j$ and $s \neq m$. Since $\omega_s \in (0,\pi)$ for all $s$ and $\omega_s \neq \omega_m$ for all $s \neq m$, both $1/|\sin\omega_s|$ and $1/\big|\sin\big((\omega_s \pm \omega_m)/2\big)\big|$ can be bounded above by a constant. Therefore, equation (29) becomes

$$
\hat{r}_j^x - r_j^x = O(1/n) + O(j/n) + O(1/n) + O(j/t) = O(j/n), \qquad (32)
$$

and hence, we obtain for $j = 0,\ldots,k$:

$$
\begin{aligned}
\hat{r}_j - r_j &= \hat{r}_j^x - r_j^x + \frac{1}{n}\sum_{t=1}^{n-j}(x_{t+j}\epsilon_t + x_t\epsilon_{t+j} + \epsilon_t\epsilon_{t+j}) - E(\epsilon_t\epsilon_{t+j})\\
&= O(j/n) + \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\rho_s\cos\big(\omega_s(t+j)+\phi_s\big)\epsilon_t\\
&+ \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\rho_s\cos(\omega_s t+\phi_s)\epsilon_{t+j} + \frac{1}{n}\sum_{t=1}^{n-j}\epsilon_t\epsilon_{t+j} - \delta_{j,0}\sigma^2
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{q} 2\rho_s \cos(\omega_s t+\phi_s)\cos(\omega_s j)\epsilon_t \\
&+ \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t\epsilon_{t-j}-\delta_{j,0}\sigma^2) \\
&- \frac{1}{n}\sum_{t=1}^{j}\sum_{s=1}^{q}\rho_s\cos\big(\omega_s(t-j)+\phi_s\big)\epsilon_t \\
&- \frac{1}{n}\sum_{t=n-j+1}^{n}\sum_{s=1}^{q}\rho_s\cos\big(\omega_s(t+j)+\phi_s\big)\epsilon_t \\
&- \frac{1}{n}\sum_{t=-j+1}^{0}\epsilon_t\epsilon_{t+j}+O\big(\frac{j}{N}\big),
\end{aligned} \tag{33}
$$

which implies

$$
\begin{aligned}
\big|(\hat{r}_j-r_j)-S_{j,n}\big| \leqslant\ & \frac{1}{n}\sum_{t=1}^{j}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\big|+\frac{1}{n}\sum_{t=n-j+1}^{n}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\big| \\
&+\frac{1}{n}\sum_{t=-j+1}^{0}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\epsilon_{t+j}\big|+O\left(\frac{j}{n}\right),
\end{aligned} \tag{34}
$$

where

$$
S_{j,n}=\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=1}^{q}2\rho_s\cos(\omega_s t+\phi_s)\cos(\omega_s j)+\epsilon_{t-j}\right)\epsilon_t-\delta_{j,0}\sigma^2. \tag{35}
$$

Since $\rho_s$, $s=1,\ldots,q$, are constants and $\{\epsilon_t\}$ is assumed to be white noise with finite fourth moment, the four terms on the right-hand side of (34) are all $O(j/n)$ a.s. Therefore, for $j=0,\ldots,k$,

$$
\hat{r}_j-r_j=S_{j,n}+O(j/n). \tag{36}
$$

Replacing $\hat{r}_j-r_j$ by $S_{j,n}+O(j/n)$ in equation (27), we obtain

$$
M_{k,n}=Q_k'\big(\boldsymbol{S}_n+\boldsymbol{E}_n\big)(1:\boldsymbol{\tau}_k')', \tag{37}
$$

where the matrices $\boldsymbol{S}_n$ and $\boldsymbol{E}_n$ respectively have elements $S_{j-l,n}$ and $O\big((j-l)/n\big)$, $j=1,\ldots,k$ and $l=1,\ldots,k+1$, and

$$
\big|M_{k,n}\big|\leq\big|Q_k'\boldsymbol{S}_n(1:\boldsymbol{\tau}_k')'\big|+\big|Q_k'\boldsymbol{E}_n(1:\boldsymbol{\tau}_k')'\big|. \tag{38}
$$

As $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, by the Cauchy-Schwartz inequality, we have

$$\left|Q_k' \boldsymbol{E}_n(1:\boldsymbol{\tau}_k')'\right| \leq \left(O(k)O(k^3/n^2)\right)^{1/2} = O(k^2/n) = o(1). \tag{39}$$

Also, note that

$$\begin{aligned} \left|Q_k' \boldsymbol{S}_n(1:\boldsymbol{\tau}_k')'\right| &\leq \left|\sum_{j=1}^{k} Q_j S_{j,n}\right| + \left|Q_k' \boldsymbol{S}_n(0:\boldsymbol{\tau}_k')'\right| \\ &= \left|\sum_{j=1}^{k} Q_j S_{j,n}\right| + O(\|\boldsymbol{S}_n\|), \end{aligned} \tag{40}$$

thus it is sufficient to show that $O(\|\boldsymbol{S}_n\|) = o(1)$. By definition,

$$TS_{j,n} = \sum_{m=1}^{b} X_{j,m}, \tag{41}$$

where

$$X_{j,m} = \epsilon_m \sum_{s=1}^{q} 2\rho_s \cos(\omega_s m + \phi_s)\cos(\omega_s j) + \epsilon_m \epsilon_{m-j} - \delta_{j,0}\sigma^2. \tag{42}$$

The rest of the proof is same as that of Theorem 3 of Mackisack and Poskitt (1990) and hence omit here. □

**Proof of Theorem 3.2.** Let $\boldsymbol{\omega}_k = (\omega_{k,1}, \ldots, \omega_{k,q})'$ be the unknown frequencies based on the $k$-th order RAR approximation. Note that

$$\hat{\omega}_{k,j} - \omega_j = (\hat{\omega}_{k,j} - \omega_{k,j}) + (\omega_{k,j} - \omega_j), \tag{43}$$

for $j = 1, \ldots, q$. Using the result of Theorem 3.1 and applying similar arguments as the proofs of Theorem 1 in MacKisack and Poskitt (1989) to a multi-frequency case, we can show that when $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\hat{\omega}_{k,j} - \omega_{k,j}\right| \geq \varepsilon\right) = 0. \tag{44}$$

As shown by Stoica et al. (1987), $(\omega_{k,j} - \omega_j) = O(1/k^3)$ and the result follows. □

Denote the regularized sample ACVF by $\hat{r}_j^{\varepsilon} = \frac{1}{n}\{\sum_{t=1}^{n-j} Y_t Y_{t+j} + \delta_{j,0}\varepsilon\, e^{\mu(n)}\}$, $j = 0, 1, \ldots, k$, for $k = 0, 1, ..., n-1$, which forms the regularized sample covariance vector $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} = (\hat{r}_0^{\varepsilon}, \hat{r}_1^{\varepsilon}, \ldots, \hat{r}_k^{\varepsilon})'$. In fact, the utilization of regularizer only changes the diagonal entries of $\hat{\boldsymbol{R}}_{k,n}$, which is $\hat{r}_0$. Asymptotically, $\hat{\boldsymbol{R}}_{k,n}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{R}}_{k,n}$ and $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{r}}_{k,0}$ by the following argument: the regularizer vanishes as $n \to \infty$, i.e.,

$$\hat{r}_0^{\varepsilon} = \frac{1}{n}\sum_{t=1}^{n} Y_t^2 + \frac{\varepsilon e^{\mu(n)}}{n} \to \frac{1}{n}\sum_{t=1}^{n} Y_t^2 = \hat{r}_0 \tag{45}$$

and therefore,

$$\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} \to \hat{\boldsymbol{r}}_{k,0}, \quad \text{as } n \to \infty. \tag{46}$$

**Lemma A.1.** *Suppose that $E(\epsilon_t) = \kappa\sigma^4 < \infty$. If $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, then*

$$\sqrt{n}(\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} - \boldsymbol{r}_{k,0}) \to N(0, \boldsymbol{\Sigma}_k) \quad \textit{in distribution.}$$

**Proof of Lemma A.1.** By (45) and (46), $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{r}}_{k,0}$ as $n \to \infty$. Note that the assumption applied here on $\{\epsilon_t\}$, which assumes $\{\epsilon_t\} \sim i.i.d\ (0, \sigma^2)$ and $E(\epsilon_t^4) = \kappa\sigma^4 < \infty$, is a special case of the assumption applied in Lau et al. (2002). Hence, by Theorem 1 of Lau et al. (2002), $\sqrt{n}(\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} - \boldsymbol{r}_{k,0})$ is asymptotically normally distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}_k$, where $\boldsymbol{\Sigma}_k = \{\sigma_{ij}^{\varepsilon}\}_{i,j=0}^{k}$ and

$$\sigma_{ij}^{\varepsilon} = \lim_{n\to\infty} E\{n(\hat{r}_i^{\varepsilon} - r_i)(\hat{r}_j^{\varepsilon} - r_j)\}. \tag{47}$$

For any $j = 0, \ldots, k$, the estimation error of regularized sample ACVF estimate is given by

$$\begin{aligned}
\hat{r}_j^{\varepsilon} - r_j &= \frac{j}{n}\sum_{s=1}^{q}\frac{\rho_s^2}{2}\cos(j\omega_s) \\
&+ \frac{1}{n}\sum_{s=1}^{q}\rho_s^2\cos\big((n-1)\omega_s + 2\phi_s\big)\frac{\sin\big((n-j)\omega_s\big)}{2\sin\omega_s} \\
&+ \frac{1}{n}\sum_{t=1}^{n-j} X_{t+j}\epsilon_t + \frac{1}{n}\sum_{t=1}^{n-j} X_t\epsilon_{t+j} \\
&+ \frac{1}{n}\sum_{t=1}^{n-j}\epsilon_t\epsilon_{t+j} + \frac{\delta_{j,0}\varepsilon e^{\mu(n)}}{n} - \delta_{j,0}\sigma^2 \\
&= A_{1j} + A_{2j} + A_{3j} + A_{4j} + A_{5j} + \frac{\delta_{j,0}\varepsilon e^{\mu(n)}}{n} - \delta_{j,0}\sigma^2.
\end{aligned} \tag{48}$$

Notice that as $n \to \infty$, $k \to \infty$ such that $k^2/n \to 0$, $nA_{1i}A_{1j} \to 0$ for $i, j = 0, \dots, k$, Therefore, when $i, j \neq 0$, if $k^2/n \to 0$, we obtain

$$\begin{aligned}\sigma^{\varepsilon}_{ij} &= \lim_{n\to\infty}\Big\{E\big(nA_{3i}A_{3j}\big) + E\big(nA_{4i}A_{4j}\big) \\ &\qquad + E\big(nA_{3i}A_{4j}\big) + E\big(nA_{4i}A_{3j}\big) + E\big(nA_{5i}A_{5j}\big)\Big\} \\ &= \delta_{i,j}\sigma^4 + \sigma^2\sum_{s=1}^{q} 2\rho_s^2 \cos(\omega_s i)\cos(\omega_s j).\end{aligned}$$

When $i = j = 0$,

$$\sigma^{\varepsilon}_{00} = (\kappa - 1)\sigma^4 + \sigma^2\sum_{s=1}^{q} 2\rho_s^2, \tag{49}$$

and the result follows.

**Proof of Theorem 3.3.** Since $\sqrt{T}(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0} - \boldsymbol{r}_{k,0})$ converges in distribution as stated by Lemma A.1, it follows the result of Serfling (1980) that $\hat{\boldsymbol{r}}^{\varepsilon}_{k,0} = \boldsymbol{r}_{k,0} + O(1/\sqrt{n})$. Define the following quantities:

- $g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}) = (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}\hat{\boldsymbol{r}}_k = \hat{\boldsymbol{\tau}}_k$ and $g(\boldsymbol{r}_{k,0}) = (\boldsymbol{R}_k)^{-1}\boldsymbol{r}_k = \boldsymbol{\tau}_k$,
- $\boldsymbol{\Delta}_{k,i}$= $(k \times k)$-matrix with $\pm i^{th}$ off-diagonal elements equal to 1, and 0 otherwise,
- $\boldsymbol{\vartheta}_{k,i}$= $(k \times 1)$-vector with $\pm i^{th}$ element equal to 1, and 0 otherwise.

Note in particular that $\boldsymbol{\Delta}_{k,0}$ is the identity matrix and $\boldsymbol{\Delta}_{k,k}$ is a zero matrix. In the view of matrix derivative (see, e.g., Grandshteyn and Ryzhik, 2000; Lau et al., 2002),

$$\begin{aligned}\frac{\partial(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}}{\partial\hat{r}^{\varepsilon}_i} &= -(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}\boldsymbol{\Delta}_{k,i}(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1} \quad \text{and} \\ \frac{\partial\hat{\boldsymbol{r}}_k}{\partial\hat{r}^{\varepsilon}_i} &= \boldsymbol{\vartheta}_{k,i}, \quad i = 0, 1, \dots, k.\end{aligned} \tag{50}$$

Thus, by the chain rule,

$$\begin{aligned}\left.\frac{\partial g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0})}{\partial\hat{r}^{\varepsilon}_i}\right|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} &= \left.\left\{\left(\frac{\partial(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}}{\partial\hat{r}^{\varepsilon}_i}\right)\hat{\boldsymbol{r}}_k + (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}\left(\frac{\partial\hat{\boldsymbol{r}}_k}{\partial\hat{r}^{\varepsilon}_i}\right)\right\}\right|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} \\ &= -(\boldsymbol{R}_k)^{-1}\boldsymbol{\Delta}_{k,i}(\boldsymbol{R}_k)^{-1}\boldsymbol{r}_k + (\boldsymbol{R}_k)^{-1}\boldsymbol{\vartheta}_{k,i} \\ &= -(\boldsymbol{R}_k)^{-1}(\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i}).\end{aligned} \tag{51}$$

Applying the Taylor expansion,

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) &= \sqrt{n}\{g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}) - g(\boldsymbol{r}_{k,0})\} \\ &= \sqrt{n}\sum_{i=0}^{k} \frac{\partial g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0})}{\partial \hat{r}^{\varepsilon}_i}\bigg|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} (\hat{r}^{\varepsilon}_i - r_i) + o(1) \\ &= -\sum_{i=0}^{k} (\boldsymbol{R}_k)^{-1}(\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i})\sqrt{n}(\hat{r}^{\varepsilon}_i - r_i) + o(1) \\ &= -(\boldsymbol{R}_k)^{-1}[\boldsymbol{\tau}_k, (\boldsymbol{\Delta}_{k,1}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,1}), \ldots, (\boldsymbol{\Delta}_{k,k}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,k})] \\ &\quad \times \sqrt{n}(\hat{r}^{\varepsilon}_i - r_i) + o(1).\end{aligned} \tag{52}$$

Let $a_i = 0$ for $i < 0$ and $i > k$. Note that

$$\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i} = (a_{1+i}, a_{2+i}, \ldots, a_{k+i})' - (a_{1-i}, a_{2-i}, \ldots, a_{k-i})'. \tag{53}$$

Therefore, $[\boldsymbol{\tau}_k, (\boldsymbol{\Delta}_{k,1}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,1}), \ldots, (\boldsymbol{\Delta}_{k,k}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,k})] = \boldsymbol{M}_k$ and the result follows by Lemma A.1. □

**Proof of Theorem 3.4.** Let $\{\hat{\beta}_s e^{\pm i\hat{\omega}_{k,s}}\}_{s=1}^{q}$ denote the $2q$ roots of $\hat{a}(z)$ which are closet to the unit circle. Applying the same arguments as in Stoica et al. (1989) and taking into account the results on asymptotic consistency and normality of $\hat{\boldsymbol{\tau}}_k$ and Theorem 2, we obtain that $\hat{\omega}_{k,s}$ is close to $\omega_s$, $s = 1, \ldots, q$, and $\hat{\beta}_s$ is close to $\beta_s = 1$ for sufficiently large $n$. Hence, the following Taylor expansion holds under regularity conditions:

$$\begin{aligned}0 = \mathrm{Re}\{\hat{a}(\hat{\beta}_s e^{i\hat{\omega}_{k,s}})\} &= \mathrm{Re}\{\hat{a}(e^{i\omega_s})\} + \frac{\partial \mathrm{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\bigg|_{\beta=1,\omega=\omega_s} (\hat{\beta}_s - \beta_s) \\ &+ \frac{\partial \mathrm{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\bigg|_{\beta=1,\omega=\omega_s} (\hat{\omega}_{k,s} - \omega_s) + O(1/n),\end{aligned} \tag{54}$$

$$\begin{aligned}0 = \mathrm{Im}\{\hat{a}(\hat{\beta}_s e^{i\hat{\omega}_{k,s}})\} &= \mathrm{Im}\{\hat{a}(e^{i\omega_s})\} + \frac{\partial \mathrm{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\bigg|_{\beta=1,\omega=\omega_s} (\hat{\beta}_s - \beta_s) \\ &+ \frac{\partial \mathrm{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\bigg|_{\beta=1,\omega=\omega_s} (\hat{\omega}_{k,s} - \omega_s) + O(1/n),\end{aligned} \tag{55}$$

where

$$\begin{aligned}
\left.\frac{\partial \mathrm{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} &= (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\hat{\boldsymbol{\tau}}_k,\\
\left.\frac{\partial \mathrm{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} &= -(\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\hat{\boldsymbol{\tau}}_k,\\
\left.\frac{\partial \mathrm{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} &= (\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\hat{\boldsymbol{\tau}}_k,\\
\left.\frac{\partial \mathrm{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} &= (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\hat{\boldsymbol{\tau}}_k.
\end{aligned} \tag{56}$$

By Theorem 3.3, as $k \to \infty$ and $T \to \infty$ such that $k^2/n \to 0$, $\sqrt{n}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k)$ converges in distribution and thus it follows the result of Serfling (1980) that $(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) = O(1/\sqrt{n})$. Also, by Theorem 1 of Stoica et al. (1987), $(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2)$. Hence, we obtain

$$\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k^* = (\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) + (\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2) + O(1/\sqrt{n}). \tag{57}$$

Since $k^2/n \to 0$, the dominant term in (54) is not affected if we replace $\hat{\boldsymbol{\tau}}_k$ by $\boldsymbol{\tau}_k^*$, which is

$$\begin{aligned}
0 &= \mathrm{Re}\{\hat{a}(e^{i\omega_s})\} + \theta_s(\hat{\beta}_s - \beta_s) - \psi_s(\hat{\omega}_{k,s} - \omega_s) + O(1/n),\\
0 &= \mathrm{Im}\{\hat{a}(e^{i\omega_s})\} + \psi_s(\hat{\beta}_s - \beta_s) + \theta_s(\hat{\omega}_{k,s} - \omega_s) + O(1/n).
\end{aligned} \tag{58}$$

Since $a^*(e^{i\omega_s}) = 0$,

$$\begin{aligned}
\mathrm{Re}\{\hat{a}(e^{i\omega_s})\} &= \mathrm{Re}\{\hat{a}(e^{i\omega_s}) - a^*(e^{i\omega_s})\} = \boldsymbol{h}_s'(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k^*),\\
\mathrm{Im}\{\hat{a}(e^{i\omega_s})\} &= \mathrm{Im}\{\hat{a}(e^{i\omega_s}) - a^*(e^{i\omega_s})\} = \boldsymbol{g}_s'(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k^*).
\end{aligned} \tag{59}$$

Substituting (59) into (58), we obtain

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k^*) + O(1/n). \tag{60}$$

Equivalently,

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) + \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) + O(1/n). \tag{61}$$

By the result of Stoica et al. (1987) Theorem 1,

$$(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2), \quad \theta_s/k = -1/2 + O(1/k), \quad \text{and} \quad \psi_s/k = O(1/k). \tag{62}$$

Substituting (62) into $(\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s')(\theta_s^2 + \psi_s^2)^{-1}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*)$, we obtain

$$\frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^3). \tag{63}$$

Therefore,

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) + O(1/k^3) + O(1/n), \tag{64}$$

or equivalently,

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) = \sqrt{n}\boldsymbol{F}\boldsymbol{G}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) + O(\sqrt{n}/k^3) + O(1/\sqrt{n}). \tag{65}$$

If $k^2 \geqslant cn^{1-\delta}$, for $0 < \delta < 2/3$, then $\sqrt{n}/k^3 \to 0$, and $O(\sqrt{n}/k^3) \to 0$. Also, as $n \to \infty$, $O(1/\sqrt{n}) \to 0$, so we have

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) = \sqrt{n}\boldsymbol{F}\boldsymbol{G}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k). \tag{66}$$

By Theorem 3.3, $\sqrt{n}(\hat{\boldsymbol{\tau}}_k - \boldsymbol{\tau}_k) \to N(0, \boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1})$, and thus, if $k^{3/2} \geqslant cn^{1-\delta}$, for $0 < \delta < 2/3$ such that $k^2/n \to 0$,

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) \to N(0, \boldsymbol{F}\boldsymbol{G}\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1}\boldsymbol{G}'\boldsymbol{F}'). \quad \square \tag{67}$$