The estimation of frequency in the multichannel sinusoidal model

Andrew J. Grant*, Barry G. Quinn

Department of Mathematics and Statistics, Macquarie University, NSW 2109, Australia

Abstract

The problems of estimating the frequency parameter in a univariate noisy sinusoidal model, and deriving the asymptotic properties of the estimator, are well understood. In the multivariate context, there is very little known, apart from papers in the array processing literature, where the common frequency parameter is generally not estimated accurately. We present a method for estimating this frequency and show that the resulting estimator is strongly consistent and follows a central limit theorem. The performance of the estimator is demonstrated using the results of simulation studies.

Keywords: Central limit theorem, Multichannel frequency estimation, Spectral density matrix, Strong consistency.
2010 MSC: Primary 62H12, Secondary 62F12, 62M15

1. Introduction

There is a long history of fitting sinusoids to univariate time series in order to estimate the frequency of periodic components. In this paper, we consider a multivariate generalisation of the problem, where we wish to estimate ω in the model

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\alpha} \cos(\omega t) + \boldsymbol{\beta} \sin(\omega t) + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\omega \in (0, \pi)$ is scalar, $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{x}_t$ are $d \times 1$ and $\{\boldsymbol{\varepsilon}_t\}$ is a d-dimensional stationary centered stochastic process. This has been called the multichannel sinusoidal model by [20]. Each element of the periodic component oscillates at the same fixed frequency, ω, but has potentially different amplitudes and phases. The jth element of $\mathbf{x}_t$ may be written as

$$X_{t,j} = \mu_j + \rho_j \cos\left(\omega t + \phi_j\right) + \varepsilon_{t,j},$$

where ρ_j is the amplitude and ϕ_j is the phase of the jth sinusoid, and $\varepsilon_{t,j}$ is the jth element of $\boldsymbol{\varepsilon}_t$. The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (1) are related to the amplitudes and phases by $\alpha_j = \rho_j \cos\phi_j$ and $\beta_j = -\rho_j \sin\phi_j$, $j \in \{1, \ldots, d\}$, where α_j and β_j are the jth elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, and $\boldsymbol{\alpha}^\top \boldsymbol{\alpha} + \boldsymbol{\beta}^\top \boldsymbol{\beta} = \boldsymbol{\rho}^\top \boldsymbol{\rho}$, where $\boldsymbol{\rho} = [\rho_1 : \ldots : \rho_d]^\top$.

The estimation of ω when $\{\mathbf{x}_t\}$ is univariate, that is when $d = 1$, has been widely studied (see, for example, [14, 15] for an overview). Nonlinear least squares estimation consists of two stages. For fixed ω, (1) is linear in $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and these parameters can be estimated using ordinary linear regression. Substituting these estimators back into the sum of squares function results in a function of ω only, which is then minimised. The minimiser of this function is asymptotically equivalent to the maximiser of the periodogram, given by

$$I_{T,\mathbf{x}}(\omega) = \frac{2}{T} \left| \sum_{t=0}^{T-1} \mathbf{x}_t e^{-i\omega t} \right|^2,$$

where T is the sample size. Let $\widehat{\omega}$ be the periodogram maximiser and ω_0 be the true frequency. If $\{\boldsymbol{\varepsilon}_t\}$ is Gaussian and white, then $T(\widehat{\omega} - \omega_0) \to 0$ almost surely and the distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance $24\sigma^2/\rho^2$. This result was first given by [24] and then later proved by [23]. Moreover,

*Corresponding author. Email address: `andrew.grant@mrc-bsu.cam.ac.uk`

[8] showed that if $\{\boldsymbol{\varepsilon}_t\}$ is not white, but coloured, with spectral density $f_\varepsilon(\omega)$, and not necessarily Gaussian, then $T(\widehat{\omega}-\omega_0)\to 0$ almost surely and the distribution of $T^{3/2}(\widehat{\omega}-\omega_0)$ converges to the normal distribution with mean zero and variance $48\pi f_\varepsilon(\omega_0)/\rho^2$ assuming only very weak extra conditions on $\{\boldsymbol{\varepsilon}_t\}$.

In this paper we derive estimators for ω in the multichannel model and establish their strong consistency and central limit theorems, generalising the results of [23] and [8]. Unlike the univariate case, efficient estimation of frequency will require estimation of the spectral density matrix of $\{\boldsymbol{\varepsilon}_t\}$. We propose estimating the spectral density matrix by fitting a long-order autoregression and show how to incorporate this into the frequency estimation procedure. The results of a simulation study are presented which demonstrate the performance of the estimation procedure in practice.

2. The estimators and their asymptotic properties

We estimate the parameters in (1) by maximising the log-likelihood as though $\{\boldsymbol{\varepsilon}_t\}$ were Gaussian and white. We shall call this the log-likelihood, even though it will truly only be the log-likelihood under those assumptions. The results that we present hold under the more general assumptions. Following a similar approach to the univariate case, we first estimate $\boldsymbol{\mu}$, $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\Sigma}=E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right)$ for fixed ω. We then substitute these estimators back into the log-likelihood which will then be a function of ω only.

Let

$$\mathbf{X}=\left[\begin{array}{ccccc}\mathbf{x}_0 & : & \ldots & : & \mathbf{x}_{T-1}\end{array}\right],\quad \boldsymbol{\theta}=\left[\begin{array}{ccc}\boldsymbol{\mu} & \boldsymbol{\alpha} & \boldsymbol{\beta}\end{array}\right]$$

and $\mathbf{M}_T(\omega)$ be the $T\times 3$ matrix with $(t+1)$th row

$$\left[\begin{array}{ccc}1 & \cos(\omega t) & \sin(\omega t)\end{array}\right],\quad t\in\{0,\ldots,T-1\}.$$

The log-likelihood equals

$$l(\boldsymbol{\theta},\boldsymbol{\Sigma},\omega)=-\frac{Td}{2}\ln(2\pi)-\frac{T}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}\operatorname{tr}\left[\left\{\mathbf{X}-\boldsymbol{\theta}\mathbf{M}_T^\top(\omega)\right\}^\top\boldsymbol{\Sigma}^{-1}\left\{\mathbf{X}-\boldsymbol{\theta}\mathbf{M}_T^\top(\omega)\right\}\right],$$

where $|\cdot|$ denotes determinant. For fixed ω, the maximiser of $l(\boldsymbol{\theta},\boldsymbol{\Sigma},\omega)$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{X}\mathbf{M}_T(\omega)\left\{\mathbf{M}_T^\top(\omega)\mathbf{M}_T(\omega)\right\}^{-1},$$

assuming the inverse exists. But

$$\mathbf{M}_T^\top(\omega)\mathbf{M}_T(\omega)=T\operatorname{diag}\left(1,\frac{1}{2},\frac{1}{2}\right)+O(1),$$

where $O(\cdot)$ indicates the order as $T\to\infty$, since, for $\omega\neq 0,\pi$,

$$\sum_{t=0}^{T-1}\cos^2(\omega t)=\frac{T}{2}+O(1),\quad \sum_{t=0}^{T-1}\sin^2(\omega t)=\frac{T}{2}+O(1),$$

$$\sum_{t=0}^{T-1}\cos(\omega t)=O(1),\quad \sum_{t=0}^{T-1}\sin(\omega t)=O(1),\quad \sum_{t=0}^{T-1}\cos(\omega t)\sin(\omega t)=O(1).$$

Hence the maximisers of $l(\boldsymbol{\theta},\boldsymbol{\Sigma},\omega)$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ have the same asymptotic properties as

$$\widehat{\boldsymbol{\theta}}_T(\omega)=\left[\begin{array}{ccc}\overline{\mathbf{x}} & 2T^{-1}\sum_{t=0}^{T-1}\cos(\omega t)\mathbf{x}_t & 2T^{-1}\sum_{t=0}^{T-1}\sin(\omega t)\mathbf{x}_t\end{array}\right],$$
$$\widehat{\boldsymbol{\Sigma}}_T(\omega)=T^{-1}\left\{\mathbf{V}_T-\mathbf{c}_T(\omega)\mathbf{c}_T^\top(\omega)-\mathbf{s}_T(\omega)\mathbf{s}_T^\top(\omega)\right\},$$

respectively, where $\overline{\mathbf{x}}=T^{-1}\sum_{t=0}^{T-1}\mathbf{x}_t$,

$$\mathbf{V}_T=\sum_{t=0}^{T-1}(\mathbf{x}_t-\overline{\mathbf{x}})(\mathbf{x}_t-\overline{\mathbf{x}})^\top,\quad \mathbf{c}_T(\omega)=\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1}\cos(\omega t)\mathbf{x}_t,\quad \mathbf{s}_T(\omega)=\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1}\sin(\omega t)\mathbf{x}_t.$$

Substituting $\widehat{\boldsymbol{\theta}}_T(\omega)$ and $\widehat{\boldsymbol{\Sigma}}_T(\omega)$ into $l(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \omega)$, we have

$$l\left(\widehat{\boldsymbol{\theta}}_T(\omega), \widehat{\boldsymbol{\Sigma}}_T(\omega), \omega\right) = -\frac{Td}{2}\{1 + \ln(2\pi)\} - \frac{T}{2}\ln\left|\widehat{\boldsymbol{\Sigma}}_T(\omega)\right|.$$

The maximum likelihood estimator of ω is thus found by minimising $\left|\widehat{\boldsymbol{\Sigma}}_T(\omega)\right|$. In order to minimise this, and to derive the asymptotic properties of the minimiser, the following lemma will be useful. The proof of the lemma, as well as the proofs of all the theorems in this paper, are presented in the Appendix.

Lemma 1. *If $\boldsymbol{a}$ and $\boldsymbol{b}$ are d-dimensional vectors then*

$$\left|\boldsymbol{I}_d - \boldsymbol{a}\boldsymbol{a}^\top - \boldsymbol{b}\boldsymbol{b}^\top\right| = 1 - \left(\boldsymbol{a}^\top\boldsymbol{a} + \boldsymbol{b}^\top\boldsymbol{b}\right) + \left(\boldsymbol{a}^\top\boldsymbol{a}\right)\left(\boldsymbol{b}^\top\boldsymbol{b}\right) - \left(\boldsymbol{a}^\top\boldsymbol{b}\right)^2$$

and

$$\left|\boldsymbol{I}_d + \boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{b}\boldsymbol{b}^\top\right| = 1 + \left(\boldsymbol{a}^\top\boldsymbol{a} + \boldsymbol{b}^\top\boldsymbol{b}\right) + \left(\boldsymbol{a}^\top\boldsymbol{a}\right)\left(\boldsymbol{b}^\top\boldsymbol{b}\right) - \left(\boldsymbol{a}^\top\boldsymbol{b}\right)^2.$$

Now, the determinant of $\widehat{\boldsymbol{\Sigma}}_T(\omega)$ is

$$T^{-d}\left|\mathbf{V}_T - \mathbf{c}_T(\omega)\mathbf{c}_T^\top(\omega) - \mathbf{s}_T(\omega)\mathbf{s}_T^\top(\omega)\right| = T^{-d}\left|\mathbf{V}_T^{1/2}\right|\left|\mathbf{I}_d - \mathbf{V}_T^{-1/2}\mathbf{c}_T(\omega)\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1/2} - \mathbf{V}_T^{-1/2}\mathbf{s}_T(\omega)\mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1/2}\right|\left|\mathbf{V}_T^{1/2}\right|.$$

From Lemma 1, putting $\mathbf{a} = \mathbf{V}_T^{-1/2}\mathbf{c}_T(\omega)$ and $\mathbf{b} = \mathbf{V}_T^{-1/2}\mathbf{s}_T(\omega)$, we obtain

$$\begin{aligned}&\left|\mathbf{I}_d - \mathbf{V}_T^{-1/2}\mathbf{c}_T(\omega)\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1/2} - \mathbf{V}_T^{-1/2}\mathbf{s}_T(\omega)\mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1/2}\right| \\ &= 1 - \mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{c}_T(\omega) - \mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega) + \left\{\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{c}_T(\omega)\right\}\left\{\mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega)\right\} - \left\{\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega)\right\}^2.\end{aligned}$$

In order to compute the maximum likelihood estimator of ω we must maximise

$$J_T(\omega) = \mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{c}_T(\omega) + \mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega) - \left\{\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{c}_T(\omega)\right\}\left\{\mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega)\right\} + \left\{\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1}\mathbf{s}_T(\omega)\right\}^2.$$

Of course, when evaluating the asymptotic properties of the estimator, we do not assume that $\{\boldsymbol{\varepsilon}_t\}$ is Gaussian and white. Let

$$\mathbf{F}_{\varepsilon}(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\boldsymbol{\Gamma}_{\varepsilon}(j)e^{-ij\omega},$$

where $\boldsymbol{\Gamma}_{\varepsilon}(j) = E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+j}^\top\right)$, be the spectral density matrix of $\{\boldsymbol{\varepsilon}_t\}$. Also let

$$\widehat{\omega} = \underset{\omega}{\operatorname{argmax}}\, J_T(\omega),$$

and denote by ω_0, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0 = \begin{bmatrix}\boldsymbol{\mu}_0 & \boldsymbol{\alpha}_0 & \boldsymbol{\beta}_0\end{bmatrix}$ the true values of ω, $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}$, respectively. We assume that $\mathbf{F}_{\varepsilon}(\omega)$ is continuous at ω_0. Theorem 1 shows that $T(\widehat{\omega} - \omega_0)$ converges almost surely to zero and Theorem 2 establishes the central limit theorem.

Theorem 1. $T(\widehat{\omega} - \omega_0) \to 0$ *almost surely as* $T \to \infty$.

Theorem 2. *The distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$, as $T \to \infty$, converges to the normal distribution with mean zero and variance*

$$48\pi\frac{\boldsymbol{\alpha}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{F}_{\varepsilon}(\omega_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{F}_{\varepsilon}(\omega_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0}{\left(\boldsymbol{\alpha}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right)^2}.$$

The estimator given above is computed under Gaussian white assumptions, and so a more general estimator may have better asymptotic variance when these assumptions are violated. It should be noted that there is no gain in doing this when $d = 1$, but we shall show below that there is improvement when $d > 1$.

We can re-write $J_T(\omega)$ as

$$1 - \left|\mathbf{V}_T^{-1}\right|\left|\mathbf{V}_T - \mathbf{f}_T(\omega)\mathbf{f}_T^*(\omega)\right|,$$

where

$$\mathbf{f}_T(\omega) = \mathbf{c}_T(\omega) - i\mathbf{s}_T(\omega) = \left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} e^{-i\omega t}\mathbf{x}_t,$$

and $*$ denotes the complex conjugate transpose. Theorem 1 is proved by showing that the sufficiency condition of [26] is met, which follows from the fact that, for $a \neq 0$,

$$\lim_{T\to\infty}\left\{J_T(\omega_0) - J_T(\omega_0 + a/T)\right\} > 0.$$

The result hinges on the fact that $\mathbf{V}_T$ is positive definite and that

$$T^{-1}\mathbf{f}_T(\omega_0 + a/T)\mathbf{f}_T^*(\omega_0 + a/T) \to \frac{1}{2}\left(\boldsymbol{\alpha}_0^\top\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\boldsymbol{\beta}_0\right)\frac{\sin^2(a/2)}{(a/2)^2}$$

almost surely as $T \to \infty$, where, for all $a \neq 0$,

$$\frac{\sin^2(a/2)}{(a/2)^2} < 1.$$

In light of this, we may expect to obtain a consistent estimator by maximising a simpler function of the form

$$\widetilde{J}_{T,\mathbf{\Omega}}(\omega) = \mathbf{c}_T^\top(\omega)\mathbf{\Omega}\mathbf{c}_T(\omega) + \mathbf{s}_T^\top(\omega)\mathbf{\Omega}\mathbf{s}_T(\omega) = \mathbf{f}_T^*(\omega)\mathbf{\Omega}\mathbf{f}_T(\omega),$$

where $\mathbf{\Omega}$ is a suitable positive definite symmetric matrix. Let

$$\widetilde{\omega} = \operatorname*{argmax}_{\omega} \widetilde{J}_{T,\mathbf{\Omega}}(\omega).$$

Theorem 3 shows that $T(\widetilde{\omega} - \omega_0)$ converges almost surely to zero and Theorem 4 establishes the central limit theorem.

Theorem 3. $T(\widetilde{\omega} - \omega_0) \to 0$ *almost surely as* $T \to \infty$.

Theorem 4. *The distribution of* $T^{3/2}(\widetilde{\omega} - \omega_0)$*, as* $T \to \infty$*, converges to the normal distribution with mean zero and variance*

$$48\pi\frac{\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\boldsymbol{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\boldsymbol{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\beta}_0}{\left(\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\boldsymbol{\beta}_0\right)^2}. \tag{2}$$

Letting $\boldsymbol{\eta}_0 = \boldsymbol{\alpha}_0 - i\boldsymbol{\beta}_0$, (2) may be rewritten as

$$48\pi\frac{\boldsymbol{\eta}_0^*\mathbf{\Omega}\mathbf{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\eta}_0}{\left(\boldsymbol{\eta}_0^*\mathbf{\Omega}\boldsymbol{\eta}_0\right)^2}.$$

An application of the Kantorovich matrix inequality (see, for example, [6], Section 8.4.3) shows that

$$\frac{\boldsymbol{\eta}_0^*\mathbf{\Omega}\mathbf{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\eta}_0}{\left(\boldsymbol{\eta}_0^*\mathbf{\Omega}\boldsymbol{\eta}_0\right)^2} \geqslant \frac{1}{\boldsymbol{\eta}_0^*\mathbf{F}_\varepsilon^{-1}(\omega_0)\boldsymbol{\eta}_0}$$

with equality if and only if $\mathbf{\Omega} = c\mathbf{F}_\varepsilon^{-1}(\omega_0)$ for some constant c. The estimator of ω in this class with the smallest asymptotic variance may therefore be obtained by maximising $\widetilde{J}_{T,\mathbf{\Omega}}(\omega)$ with $\mathbf{\Omega}$ equal to $\mathbf{F}_\varepsilon^{-1}(\omega_0)$. In practice, $\mathbf{F}_\varepsilon(\omega_0)$ will not be known, and will need to be estimated. The multivariate case is thus very different from the univariate, where estimation of the spectral density is not needed.

In practice, we propose estimating ω in two stages. In the first stage, we set $\mathbf{\Omega} = \mathbf{I}_d$ and maximise $\widetilde{J}_{T,\mathbf{\Omega}}(\omega)$. This provides a consistent estimator of ω. Although it is not asymptotically relatively efficient, it is of the correct order of efficiency. In the second stage, we estimate the spectral density matrix of the residuals obtained by removing the corresponding sinusoid by regression, and then maximise $\widetilde{J}_{T,\mathbf{\Omega}}(\omega)$ with $\mathbf{\Omega}$ set to the inverse of the estimator. As long as we estimate the spectral density matrix of the residuals consistently, we will obtain an asymptotically efficient estimator of ω in the second stage.

There are many methods for estimating the spectral density of a multivariate stationary process. Nonparametric methods apply smoothing to periodogram matrices (see, for example, [12]). In the following section we propose estimating the spectral density matrix parametrically by fitting long-order autoregressions. We then incorporate this approach into the two step procedure which we present formally in Section 4.

3. Autoregressive approximation

Although $\{\boldsymbol{\varepsilon}_t\}$ may not be truly autoregressive, approximation of stationary processes using long-order autoregressions has a long history [4, 5]. Moreover, accurate nonparametric estimation of the spectral density matrix typically requires very large sample sizes. We thus fit

$$\boldsymbol{\varepsilon}_t + \sum_{j=1}^{p} \boldsymbol{\delta}_j \boldsymbol{\varepsilon}_{t-j} = \mathbf{u}_t \tag{3}$$

to $\{\boldsymbol{\varepsilon}_t\}$, where $\boldsymbol{\delta}_j$, $j \in \{1, \ldots, p\}$, are $d \times d$ and the residual process, $\{\mathbf{u}_t\}$, is d-dimensional. The assumptions we make on $\{\mathbf{u}_t\}$ are

$$E(\mathbf{u}_t \mid \mathcal{F}_{t-1}) = 0, \quad E\left(\mathbf{u}_t \mathbf{u}_t^\top \mid \mathcal{F}_{t-1}\right) = \mathbf{G}, \tag{4}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{\mathbf{u}_t, \mathbf{u}_{t-1}, \ldots\}$. If $\{\boldsymbol{\varepsilon}_t\}$ satisfies (3) then

$$\mathbf{F}_{\boldsymbol{\varepsilon}}(\omega) = \frac{1}{2\pi}\left(\mathbf{I}_d + \sum_{j=1}^{p} \boldsymbol{\delta}_j e^{-ij\omega}\right)^{-1} \mathbf{G} \left\{\left(\mathbf{I}_d + \sum_{j=1}^{p} \boldsymbol{\delta}_j e^{-ij\omega}\right)^{*}\right\}^{-1} \tag{5}$$

(see, for example, [18], Section 2.3). The autoregressive order, p, will in general be unknown and will need to be estimated. For now, we will assume that p is fixed, and discuss its estimation at the end of this section.

Let

$$\widetilde{\boldsymbol{\delta}} = \operatorname{vec}\left[\begin{array}{ccccc} \boldsymbol{\delta}_1 & : & \ldots & : & \boldsymbol{\delta}_p \end{array}\right], \quad \widetilde{\mathbf{G}} = \operatorname{vec}\mathbf{G}.$$

The log-likelihood equals

$$l(\mathbf{\Theta}) = -\frac{Td}{2}\ln(2\pi) - \frac{T}{2}\ln|\mathbf{G}| - \frac{1}{2}\sum_{t=0}^{T-1} \widetilde{\boldsymbol{u}}_t^\top \mathbf{G}^{-1} \widetilde{\boldsymbol{u}}_t,$$

where

$$\mathbf{\Theta} = \left[\begin{array}{cccccc} \boldsymbol{\mu}^\top & \widetilde{\boldsymbol{\delta}}^\top & \widetilde{\mathbf{G}}^\top & \boldsymbol{\alpha}^\top & \boldsymbol{\beta}^\top & \omega \end{array}\right]^\top, \quad \widetilde{\boldsymbol{u}}_t = \mathbf{d}(z)\{\mathbf{x}_t - \boldsymbol{\mu} - \boldsymbol{\alpha}\cos(\omega t) - \boldsymbol{\beta}\sin(\omega t)\}, \quad \mathbf{d}(z) = \mathbf{I}_d + \sum_{j=1}^{p} \boldsymbol{\delta}_j z^j$$

and $z : a_t \to a_{t-1}$. Let

$$\widehat{\omega} = \underset{\omega}{\operatorname{argmax}} \max_{\boldsymbol{\mu}, \widetilde{\boldsymbol{\delta}}, \widetilde{\mathbf{G}}, \boldsymbol{\alpha}, \boldsymbol{\beta}} l(\mathbf{\Theta}).$$

Theorem 5 establishes the central limit theorem for $\widehat{\omega}$.

Theorem 5. *If $\{\boldsymbol{\varepsilon}_t\}$ satisfies (3) and (4), the distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$, as $T \to \infty$, converges to the normal distribution with mean zero and variance*

$$\frac{48\pi}{\boldsymbol{\eta}_0^* \boldsymbol{F}_{\boldsymbol{\varepsilon}}^{-1}(\omega_0)\boldsymbol{\eta}_0}.$$

In light of this, an asymptotically relatively efficient estimator, at least in the Gaussian case, of ω may be obtained by fitting a long-order autoregression to $\{\boldsymbol{\varepsilon}_t\}$ and using the inverse of the estimated spectral density matrix in place of $\boldsymbol{\Omega}$ in maximising $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$. Methods for fitting autoregressions to multivariate time series are well known (see, for example, [7]). A computationally efficient method is the Whittle recursion [25]. The recursion can easily incorporate the estimation of the autoregressive order, p, using, for example, an information criterion such as the automatic information criterion, AIC [1], the Bayesian information criterion, BIC [2, 21] or the Hannan–Quinn information criterion, HQIC [11, 13]. Both BIC and HQIC give strongly consistent estimators of p. Furthermore, the spectral density matrix is consistently estimated, even when the underlying process is not autoregressive [9].

4. The algorithm

The full procedure for estimating ω, incorporating the autoregressive approximation of $\{\boldsymbol{\varepsilon}_t\}$, is as follows.

1. Put $\boldsymbol{\Omega} = \mathbf{I}_d$ and let
$$\widetilde{\omega} = \underset{\omega}{\operatorname{argmax}}\, \widetilde{J}_{T,\boldsymbol{\Omega}}(\omega).$$

2. Fit an autoregression to
$$\mathbf{x}_t - \overline{\mathbf{x}} - \widehat{\boldsymbol{\alpha}}\cos(\widetilde{\omega}t) - \widehat{\boldsymbol{\beta}}\sin(\widetilde{\omega}t),$$
where
$$\widehat{\boldsymbol{\alpha}} = \frac{2}{T}\sum_{t=0}^{T-1}\cos(\widetilde{\omega}t)\,\mathbf{x}_t, \quad \widehat{\boldsymbol{\beta}} = \frac{2}{T}\sum_{t=0}^{T-1}\sin(\widetilde{\omega}t)\,\mathbf{x}_t$$
and the autoregressive order is estimated using a strongly consistent information criterion. Denote the order estimate by $\widehat{p}$, the autoregressive parameter estimates by $\widehat{\boldsymbol{\delta}}_1, \ldots, \widehat{\boldsymbol{\delta}}_{\widehat{p}}$ and the residual covariance matrix estimate by $\widehat{\mathbf{G}}$.

3. Put
$$\boldsymbol{\Omega} = 2\pi\left(\mathbf{I}_d + \sum_{j=1}^{\widehat{p}}\widehat{\boldsymbol{\delta}}_j e^{-ij\widetilde{\omega}}\right)^{*}\widehat{\mathbf{G}}^{-1}\left(\mathbf{I}_d + \sum_{j=1}^{\widehat{p}}\widehat{\boldsymbol{\delta}}_j e^{-ij\widetilde{\omega}}\right)$$
and let
$$\widetilde{\omega} = \underset{\omega}{\operatorname{argmax}}\, \widetilde{J}_{T,\boldsymbol{\Omega}}(\omega).$$

It remains to maximise $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$ for a given $\boldsymbol{\Omega}$. This can be done, for example, using the Gauss–Newton algorithm. Given a current estimate of ω, denoted $\widetilde{\omega}$, the Gauss–Newton algorithm updates the estimate by
$$\widetilde{\omega} + \frac{\operatorname{Re}\left\{\mathbf{f}_T^{*}(\widetilde{\omega})\,\boldsymbol{\Omega}\frac{\partial}{\partial\omega}\mathbf{f}_T(\widetilde{\omega})\right\}}{\frac{\partial}{\partial\omega}\mathbf{f}_T^{*}(\widetilde{\omega})\,\boldsymbol{\Omega}\frac{\partial}{\partial\omega}\mathbf{f}_T(\widetilde{\omega})},$$
where
$$\frac{\partial}{\partial\omega}\mathbf{f}_T(\omega) = -i\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} t e^{-i\omega t}\mathbf{x}_t,$$
and repeats until convergence. In order to initialise the algorithm, we could take the maximiser of $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$ computed at the Fourier frequencies. These values can be easily computed using the fast Fourier transform. In the univariate case, using the maximiser of the periodogram over the Fourier frequencies as an initial value does not guarantee that the Gauss–Newton algorithm will converge to the true frequency [19]. However, [17] have shown that the Gauss–Newton algorithm will converge if the initial estimator is computed using the periodogram of the time series zero-padded to four times its length. While it remains to show theoretically whether this result applies to the multivariate case, simulations suggest that a zero-padding factor of four is appropriate here. For example, for the simulation study presented in Section 5, zero-padding the time series to four times their length generally produced the same results as zero-padding them to eight times their length. We therefore adopt this approach when finding an initial value for the Gauss–Newton algorithm.

5. Simulations

Sets of time series with 10,000 replications were simulated from (1) with

$$\boldsymbol{\mu} = 0, \quad \boldsymbol{\alpha} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^{\top}, \quad \boldsymbol{\beta} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^{\top}.$$

Note that the amplitudes of the sinusoids are 1. The sample sizes were $T = 100, 250, 500$ and $1,000$ and the frequency was $\omega = \pi/5$. The noise process was either generated from white noise, where $\boldsymbol{\varepsilon}_t = \mathbf{u}_t$, the autoregressive model

$$\boldsymbol{\varepsilon}_t + \boldsymbol{\delta}_1 \boldsymbol{\varepsilon}_{t-1} = \mathbf{u}_t$$

with

$$\boldsymbol{\delta}_1 = \begin{bmatrix} 0.7 & 0.3 \\ -0.3 & 0.7 \end{bmatrix},$$

or the moving average model

$$\boldsymbol{\varepsilon}_t = \mathbf{u}_t + \boldsymbol{\delta}_2 \mathbf{u}_{t-1}$$

with

$$\boldsymbol{\delta}_2 = \begin{bmatrix} 0.8 & 0.1 \\ -0.1 & 0.8 \end{bmatrix}.$$

The moving average model was chosen in order to test the algorithm when the noise is not autoregressive. The residuals, $\{\mathbf{u}_t\}$, were simulated from the multivariate normal distribution with mean zero and covariance matrix either

$$\mathbf{G}_1 = \mathbf{I}_2, \quad \mathbf{G}_2 = \begin{bmatrix} 0.25 & 0 \\ 0 & 2 \end{bmatrix}, \quad \text{or} \quad \mathbf{G}_3 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Fig. 1 shows the spectral densities for each component of the $\{\boldsymbol{\varepsilon}_t\}$ processes for the autoregressive and moving average cases with the residual covariance matrix equal to $\mathbf{G}_1$. Also shown is the coherency, which equals

$$\frac{\left| f_{\varepsilon,12}(\omega) \right|^2}{f_{\varepsilon,11}(\omega)\, f_{\varepsilon,22}(\omega)},$$

where $f_{\varepsilon,ij}(\omega)$ is the (i, j)th element of $\mathbf{F}_{\varepsilon}(\omega)$.

The estimation procedure of Section 4 was applied using the Gauss–Newton algorithm to maximise $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$ with a tolerance for convergence of 10^{-6}. In each case, the frequency parameter was estimated twice. In the first estimation, the order of the stationary component was known (note that, in the moving average case, the true parameter values were used). In the second estimation, the stationary component was estimated using autoregressive approximation with BIC to estimate the autoregressive orders. The variances of the estimators were approximated using (2) and substituting $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_p$ and $\mathbf{G}$ by their respective parameter estimates given by stage 2 of the estimation procedure. Tables 1–3 show the means and standard deviations of the estimates, the means of the standard errors (that is, the square root of the estimated variances) and the coverage (that is, the proportion of times that the 95% confidence interval contained the true frequency parameter value). The 95% confidence intervals were constructed using the estimated standard errors.

A second set of simulations was generated to evaluate the performance of the estimator as the signal to noise ratio varies. Sets of time series were simulated using the same models as before, but with the covariance matrix of $\{\mathbf{u}_t\}$ set to $g\mathbf{I}_2$, $g \in \{0.1, \ldots, 4\}$. The Gauss–Newton algorithm was used to maximise $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$ with a tolerance for convergence of 10^{-6}, and the autoregressive orders were estimated using BIC.

Figs. 2–4 show the logarithm of the mean square errors of the resulting estimates of ω. The plots show that the mean square errors are close to the theoretical asymptotic variances up to a point at which the signal-to-noise ratio becomes too small, that is, when g becomes too large. This is known as the threshold effect [16] and shows that, as the amplitude of the sinusoid is reduced, $\widetilde{J}_{T,\boldsymbol{\Omega}}(\omega)$ is more likely to be maximised away from the true frequency. The plots show where the threshold effect begins to occur for the various scenarios considered. For example, when $T = 100$ and $\{\boldsymbol{\varepsilon}_t\}$ is white noise, Fig. 2 shows the threshold effect occurring around $g = 1.5$. It occurs sooner in the autoregressive

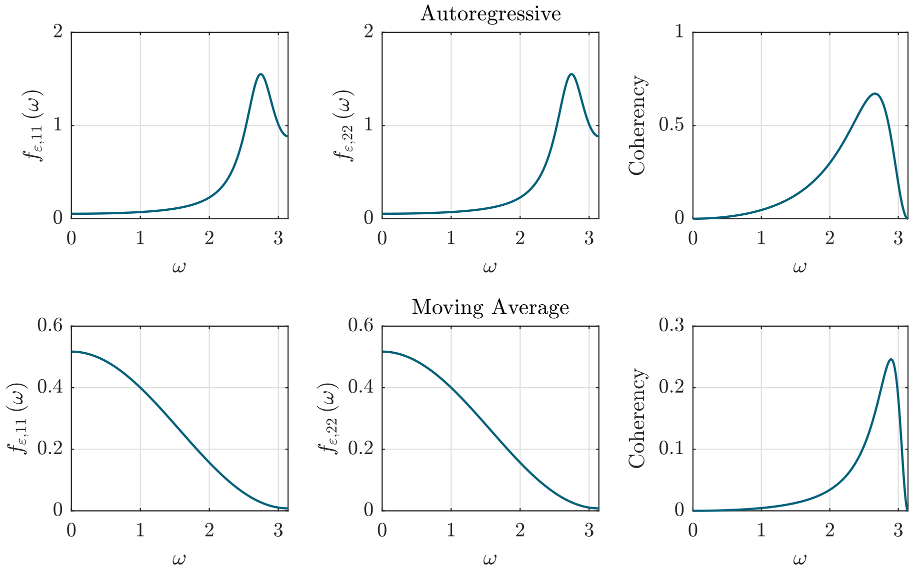

Fig. 1: The spectral densities of each component, as well as their coherency, for the autoregressive and moving average processes described in Section 5.

and moving average cases than in the white noise cases. In all cases, there was no threshold effect when $T = 1,000$ up to $g = 4$.

The second set of simulations was repeated for the cases where the covariance matrix of $\{\mathbf{u}_t\}$ was set to $g\mathbf{G}_2$ and $g\mathbf{G}_3$, $g \in \{0.1, \ldots, 4\}$. Plots of the logarithms of the mean square errors for these simulations are shown in the supplementary material.

Table 1
Results of simulations when the covariance matrix of the noise process was $\mathbf{G}_1$. Shown are the mean estimates (Mean), the standard deviation of the estimates $\times 10^2$ (S.D), the mean estimated standard deviation $\times 10^2$ (Mean S.E.) and the proportion of 95% confidence intervals which contain the true frequency parameter (Coverage). Note that the true frequency parameter is $pi/5 = 0.6283$

		Known order				BIC estimated order			
Model	T	Mean	S.D.	Mean S.E.	Coverage	Mean	S.D.	Mean S.E.	Coverage
WN	100	0.629	0.357	0.332	0.926	0.629	0.357	0.332	0.926
	250	0.628	0.088	0.086	0.940	0.628	0.088	0.086	0.940
	500	0.628	0.031	0.031	0.945	0.628	0.031	0.031	0.945
	1000	0.628	0.011	0.011	0.953	0.628	0.011	0.011	0.953
AR(1)	100	0.651	22.444	0.353	0.939	0.659	25.890	0.352	0.935
	250	0.628	0.054	0.054	0.939	0.628	0.054	0.054	0.939
	500	0.628	0.019	0.019	0.945	0.628	0.019	0.019	0.945
	1000	0.628	0.007	0.007	0.949	0.628	0.007	0.007	0.949
MA(1)	100	0.641	15.118	0.566	0.888	0.646	17.589	0.580	0.885
	250	0.628	0.162	0.148	0.924	0.628	0.162	0.154	0.933
	500	0.628	0.057	0.053	0.928	0.628	0.057	0.055	0.941
	1000	0.628	0.020	0.019	0.934	0.628	0.020	0.020	0.947

6. Discussion

In this paper, we have presented a method for estimating a single frequency in the multichannel sinusoidal model and have shown that the resulting estimator is strongly consistent and follows a central limit theorem. The multichan-

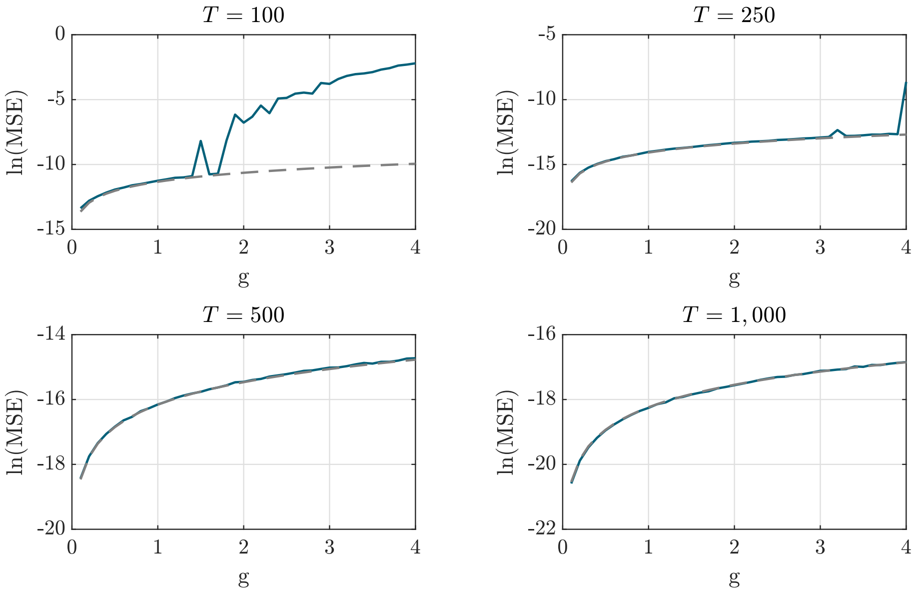

Fig. 2: Logarithm of the mean square errors (MSE) of the estimates of ω when $\{\varepsilon_t\}$ is white noise. The theoretical variance is indicated by the dashed line.

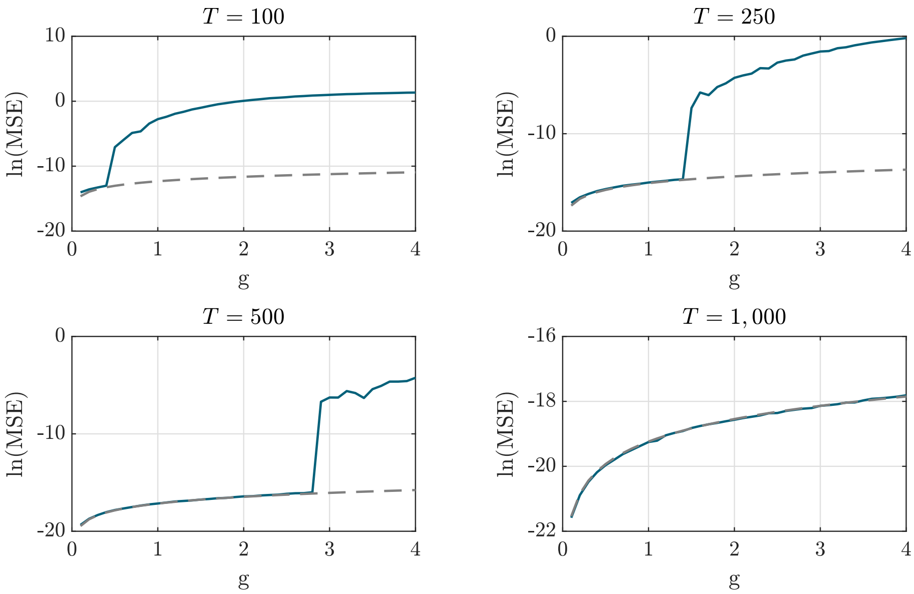

Fig. 3: Logarithm of the mean square errors (MSE) of the estimates of ω when $\{\varepsilon_t\}$ is autoregressive. The theoretical variance is indicated by the dashed line.

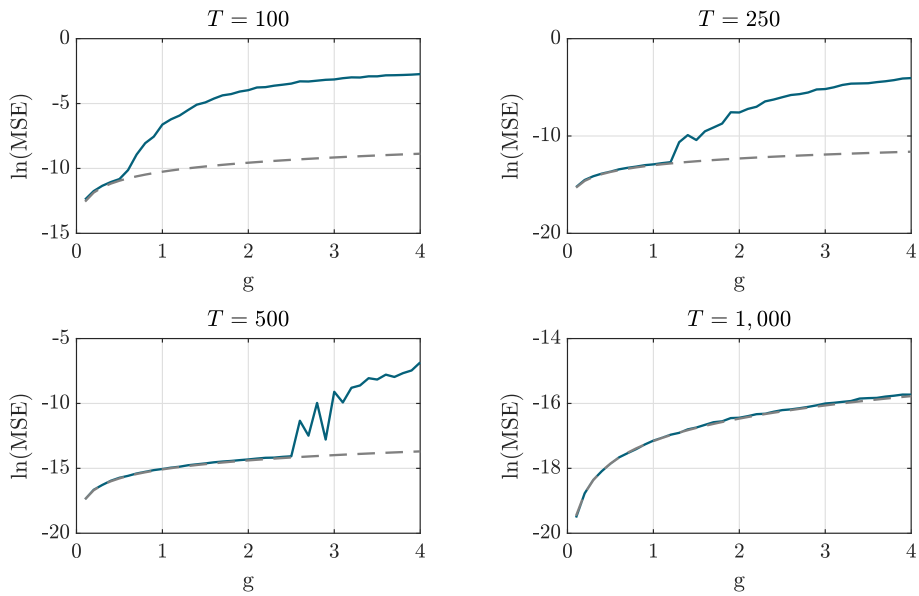

Fig. 4: Logarithm of the mean square errors (MSE) of the estimates of ω when $\{\varepsilon_t\}$ is a moving average process. The theoretical variance is indicated by the dashed line.

nel case differs notably from the univariate in that the spectral density of the stationary component must be estimated in order to obtain an efficient estimator.

Not considered here is the case where there is more than one sinusoid in the periodic component. When the time series is univariate, the frequencies may be estimated one at a time, with each estimated sinusoid removed from the time series by regression before estimating the next. The multichannel model with more than one sinusoid was considered by [20]. It was assumed that the true frequencies were Fourier frequencies, that is, in the set $2\pi j/T$, $j \in \{1, \ldots, \lfloor (T-1)/2 \rfloor\}$, where $\lfloor k \rfloor$ denotes the largest integer less than or equal to k, and also that the stochastic component was Gaussian and white. It was noted that in the multichannel case, unlike the univariate, the frequencies cannot be estimated and removed sequentially. It is possible, for example, that the frequency that maximises the likelihood when there is a single sinusoid is not in the subset of frequencies which maximises the likelihood when there are two frequencies.

Acknowledgments

We thank two reviewers and an Associate Editor for helpful comments that improved the manuscript. Andrew J. Grant has received support through an Australian Government Research Training Program Scholarship.

Appendix A. Proofs

In what follows, where convergence is indicated, it will mean convergence in the almost sure sense, unless otherwise stated. Where $O(\cdot)$ notation is used, it will indicate the order in the almost sure sense as $T \to \infty$.

Proof of Lemma 1. Letting $\mathbf{V} = \begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix}$,

$$\left|\mathbf{I}_d - \mathbf{a}\mathbf{a}^\top - \mathbf{b}\mathbf{b}^\top\right| = \left|\mathbf{I}_d - \mathbf{V}\mathbf{V}^\top\right|.$$

By the partitioned matrix identities,

$$\begin{vmatrix} \mathbf{I}_d & \mathbf{V} \\ \mathbf{V}^\top & \mathbf{I}_2 \end{vmatrix} = \left|\mathbf{I}_d - \mathbf{V}\mathbf{V}^\top\right| = \left|\mathbf{I}_2 - \mathbf{V}^\top\mathbf{V}\right|.$$

Table 2
Results of simulations when the covariance matrix of the noise process was $\mathbf{G}_2$. Shown are the mean estimates (Mean), the standard deviation of the estimates $\times 10^2$ (S.D), the mean estimated standard deviation $\times 10^2$ (Mean S.E.) and the proportion of 95% confidence intervals which contain the true frequency parameter (Coverage). Note that the true frequency parameter is $pi/5 = 0.6283$

		Known order				BIC estimated order			
Model	T	Mean	S.D.	Mean S.E.	Coverage	Mean	S.D.	Mean S.E.	Coverage
WN	100	0.629	0.236	0.225	0.927	0.629	0.236	0.225	0.927
	250	0.628	0.059	0.058	0.943	0.628	0.059	0.058	0.943
	500	0.628	0.021	0.021	0.946	0.628	0.021	0.021	0.946
	1000	0.628	0.007	0.007	0.954	0.628	0.007	0.007	0.954
AR(1)	100	0.820	64.832	0.426	0.882	0.845	67.615	0.350	0.863
	250	0.629	2.483	0.040	0.937	0.629	2.484	0.036	0.937
	500	0.628	0.011	0.011	0.943	0.628	0.011	0.011	0.943
	1000	0.628	0.004	0.004	0.953	0.628	0.004	0.004	0.953
MA(1)	100	0.633	9.334	0.388	0.885	0.630	4.779	0.435	0.929
	250	0.628	0.113	0.104	0.927	0.628	0.113	0.108	0.939
	500	0.628	0.039	0.037	0.934	0.628	0.039	0.039	0.947
	1000	0.628	0.014	0.013	0.937	0.628	0.014	0.014	0.948

But

$$\left|\mathbf{I}_2 - \mathbf{V}^\top\mathbf{V}\right| = \begin{vmatrix} 1-\mathbf{a}^\top\mathbf{a} & -\mathbf{a}^\top\mathbf{b} \\ -\mathbf{b}^\top\mathbf{a} & 1-\mathbf{b}^\top\mathbf{b} \end{vmatrix} = 1 - \left(\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}\right) + \left(\mathbf{a}^\top\mathbf{a}\right)\left(\mathbf{b}^\top\mathbf{b}\right) - \left(\mathbf{a}^\top\mathbf{b}\right)^2,$$

proving the first part of the lemma. Similarly,

$$\begin{vmatrix} \mathbf{I}_d & \mathbf{V} \\ -\mathbf{V}^\top & \mathbf{I}_2 \end{vmatrix} = \left|\mathbf{I}_d + \mathbf{V}\mathbf{V}^\top\right| = \left|\mathbf{I}_2 + \mathbf{V}^\top\mathbf{V}\right|$$

and the second part of the lemma follows. □

Proof of Theorem 1. Let

$$\mathbf{w}_T(\omega) = \left(\frac{2}{T}\right)^{1/2} \sum_{t=0}^{T-1} e^{-i\omega t}\boldsymbol{\varepsilon}_t, \quad \widetilde{\boldsymbol{\theta}}_0 = \begin{bmatrix} \boldsymbol{\alpha}_0 & \boldsymbol{\beta}_0 \end{bmatrix}, \quad \mathbf{m}_t(\omega) = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \end{bmatrix}^\top.$$

Then

$$\begin{aligned} \mathbf{f}_T(\omega) &= \mathbf{w}_T(\omega) + \left(\frac{2}{T}\right)^{1/2} \boldsymbol{\mu}_0 \sum_{t=0}^{T-1} e^{-i\omega t} + \left(\frac{2}{T}\right)^{1/2} \widetilde{\boldsymbol{\theta}}_0 \sum_{t=0}^{T-1} \mathbf{m}_t(\omega_0) e^{-i\omega t} \\ &= \mathbf{w}_T(\omega) + (2T)^{-1/2} \widetilde{\boldsymbol{\theta}}_0 \sum_{t=0}^{T-1} \begin{bmatrix} e^{-i(\omega-\omega_0)t} \\ \frac{1}{i} e^{-i(\omega-\omega_0)t} \end{bmatrix} + O(1) \\ &= \mathbf{w}_T(\omega) + \left(\frac{T}{2}\right)^{1/2} \boldsymbol{\eta}_0 h_T(\omega-\omega_0) + O(1), \end{aligned}$$

where

$$h_T(x) = T^{-1} \sum_{t=0}^{T-1} e^{-ixt} = T^{-1} \frac{e^{-ixT}-1}{e^{-ix}-1} = T^{-1} e^{-ix(T-1)/2} \frac{\sin(xT/2)}{\sin(x/2)},$$

noting that $h_T(0) = 1$. Consider the case where $\omega = \omega_0 + a/T$ for some $a \geqslant 0$. Then

$$T^{-1}\mathbf{f}_T(\omega_0 + a/T)\mathbf{f}_T^*(\omega_0 + a/T) = \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top \left|h_T(a/T)\right|^2 + O\left(T^{-1}\ln T\right),$$

Table 3
Results of simulations when the covariance matrix of the noise process was $\mathbf{G}_3$. Shown are the mean estimates (Mean), the standard deviation of the estimates $\times 10^2$ (S.D), the mean estimated standard deviation $\times 10^2$ (Mean S.E.) and the proportion of 95% confidence intervals which contain the true frequency parameter (Coverage). Note that the true frequency parameter is $pi/5 = 0.6283$

		Known order				BIC estimated order			
Model	T	Mean	S.D.	Mean S.E.	Coverage	Mean	S.D.	Mean S.E.	Coverage
WN	100	0.629	0.443	0.407	0.922	0.629	0.443	0.407	0.922
	250	0.628	0.109	0.106	0.940	0.628	0.109	0.106	0.940
	500	0.628	0.038	0.038	0.949	0.628	0.038	0.038	0.949
	1000	0.628	0.013	0.013	0.947	0.628	0.013	0.013	0.947
AR(1)	100	0.657	25.395	0.384	0.944	0.657	25.395	0.384	0.944
	250	0.628	0.064	0.064	0.941	0.628	0.064	0.064	0.941
	500	0.628	0.022	0.022	0.948	0.628	0.022	0.022	0.948
	1000	0.628	0.008	0.008	0.951	0.628	0.008	0.008	0.951
MA(1)	100	0.787	51.013	0.662	0.732	0.704	35.864	0.684	0.815
	250	0.632	8.372	0.180	0.912	0.629	0.606	0.187	0.924
	500	0.628	0.070	0.064	0.927	0.628	0.070	0.067	0.937
	1000	0.628	0.024	0.023	0.935	0.628	0.024	0.024	0.946

since $\boldsymbol{\eta}_0\boldsymbol{\eta}_0^* = \widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top$ and $\mathbf{w}_T(\omega) = O\left\{(\ln T)^{1/2}\right\}$. Now, as $T \to \infty$,

$$|h_T(a/T)|^2 = \left|e^{-ix(T-1)/2}\frac{\sin(a/2)}{T\sin(a/2T)}\right|^2 \to \frac{\sin^2(a/2)}{(a/2)^2},$$

and so

$$T^{-1}\mathbf{f}_T(\omega_0 + a/T)\mathbf{f}_T^*(\omega_0 + a/T) \to \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\frac{\sin^2(a/2)}{(a/2)^2}$$

as $T \to \infty$. Also,

$$\overline{\mathbf{x}} = \boldsymbol{\mu}_0 + T^{-1}\widetilde{\boldsymbol{\theta}}_0\sum_{t=0}^{T-1}\mathbf{m}_t(\omega) + T^{-1}\sum_{t=0}^{T-1}\boldsymbol{\varepsilon}_t,$$

and so

$$\mathbf{x}_t - \overline{\mathbf{x}} = \widetilde{\boldsymbol{\theta}}_0\mathbf{m}_t(\omega) + \boldsymbol{\varepsilon}_t + O\left\{T^{-1}(\ln\ln T)^{1/2}\right\}$$

since

$$T^{-1}\sum_{t=0}^{T-1}\boldsymbol{\varepsilon}_t = O\left\{T^{-1}(\ln\ln T)^{1/2}\right\},$$

from [22]. Thus, as $T \to \infty$,

$$T^{-1}\mathbf{V}_T = T^{-1}\sum_{t=0}^{T-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top + T^{-1}\widetilde{\boldsymbol{\theta}}_0\left\{\sum_{t=0}^{T-1}\mathbf{m}_t(\omega)\mathbf{m}_t^\top(\omega)\right\}\widetilde{\boldsymbol{\theta}}_0^\top + O\left\{T^{-1/2}(\ln\ln T)^{1/2}\right\} \to \Sigma_0 + \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top \qquad \text{(A.1)}$$

Now,

$$\begin{aligned} J_T(\omega) &= 1 - \left|\mathbf{I}_d - \mathbf{V}_T^{-1/2}\mathbf{c}_T(\omega)\mathbf{c}_T^\top(\omega)\mathbf{V}_T^{-1/2} - \mathbf{V}_T^{-1/2}\mathbf{s}_T(\omega)\mathbf{s}_T^\top(\omega)\mathbf{V}_T^{-1/2}\right| \\ &= 1 - \left|\mathbf{V}_T^{-1}\right|\left|\mathbf{V}_T - \mathbf{f}_T(\omega)\mathbf{f}_T^*(\omega)\right|. \end{aligned}$$

Therefore

$$J_T(\omega_0 + a/T) \to 1 - \left|\boldsymbol{\Sigma}_0 + \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\right|^{-1}\left|\boldsymbol{\Sigma}_0 + \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top - \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\frac{\sin^2(a/2)}{(a/2)^2}\right|.$$

From Lemma 1, putting $\begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix} = 2^{-1/2}\boldsymbol{\Sigma}_0^{-1/2}\widetilde{\boldsymbol{\theta}}_0$, we have

$$\left|\boldsymbol{\Sigma}_0 + \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\right| = |\boldsymbol{\Sigma}_0|\left|\mathbf{I}_d + \frac{1}{2}\boldsymbol{\Sigma}_0^{-1/2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1/2}\right| = |\boldsymbol{\Sigma}_0|\left\{1 + \operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right) + \frac{1}{4}\left|\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right|\right\}$$

and

$$\left|\boldsymbol{\Sigma}_0 + \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top - \frac{1}{2}\widetilde{\boldsymbol{\theta}}_0\widetilde{\boldsymbol{\theta}}_0^\top\frac{\sin^2(a/2)}{(a/2)^2}\right| = |\boldsymbol{\Sigma}_0|\left\{1 + \frac{c}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right) + \frac{c^2}{4}\left|\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right|\right\},$$

where

$$0 \leqslant c = 1 - \frac{\sin^2(a/2)}{(a/2)^2} \leqslant 1.$$

Thus, as $T \to \infty$,

$$J_T(\omega_0 + a/T) \to 1 - \frac{1 + \frac{c}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right) + \frac{c^2}{4}\left|\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right|}{1 + \operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right) + \frac{1}{4}\left|\widetilde{\boldsymbol{\theta}}_0^\top\boldsymbol{\Sigma}_0^{-1}\widetilde{\boldsymbol{\theta}}_0\right|},$$

and so the almost sure limit of $J_T(\omega_0 + a/T)$ is 1 if and only if $a = 0$. That is, if $\kappa > 0$,

$$\liminf_{T\to\infty}\inf_{|\omega-\omega_0|>\kappa/T}\{J_T(\omega_0) - J_T(\omega)\} > 0$$

and it follows from Lemma 1 of [26] that $T(\widehat{\omega} - \omega_0) \to 0$ as $T \to \infty$. □

Proof of Theorem 2. Let

$$K_T(\omega) = \ln\left|\widehat{\boldsymbol{\Sigma}}_T(\omega)\right|.$$

From the mean value theorem,

$$0 = \frac{d}{d\omega}K_T(\widehat{\omega}) = \frac{d}{d\omega}K_T(\omega_0) + \frac{d^2}{d\omega^2}K_T(\omega^*)(\widehat{\omega} - \omega_0),$$

where ω^* is a point on the line segment between ω_0 and $\widehat{\omega}$. Since $T(\widehat{\omega} - \omega_0) \to 0$, it follows that $T^{3/2}(\widehat{\omega} - \omega_0)$ has the same asymptotic distribution as

$$-\frac{T^{-1/2}\frac{d}{d\omega}K_T(\omega_0)}{T^{-2}\frac{d^2}{d\omega^2}K_T(\omega_0)}.$$

The first and second derivatives of $K_T(\omega)$ are

$$\frac{d}{d\omega}K_T(\omega) = \operatorname{tr}\left\{\widehat{\boldsymbol{\Sigma}}_T^{-1}(\omega)\frac{d}{d\omega}\widehat{\boldsymbol{\Sigma}}_T(\omega)\right\},$$
$$\frac{d^2}{d\omega^2}K_T(\omega) = \operatorname{tr}\left[-\widehat{\boldsymbol{\Sigma}}_T^{-1}(\omega)\left\{\frac{d}{d\omega}\widehat{\boldsymbol{\Sigma}}_T(\omega)\right\}\widehat{\boldsymbol{\Sigma}}_T^{-1}(\omega)\left\{\frac{d}{d\omega}\widehat{\boldsymbol{\Sigma}}_T(\omega)\right\} + \widehat{\boldsymbol{\Sigma}}_T^{-1}(\omega)\left\{\frac{d^2}{d\omega^2}\widehat{\boldsymbol{\Sigma}}_T(\omega)\right\}\right].$$

The first and second derivatives of $\widehat{\boldsymbol{\Sigma}}_T(\omega)$ are

$$\frac{d}{d\omega}\widehat{\boldsymbol{\Sigma}}_T(\omega) = -T^{-1}\begin{bmatrix} \frac{d}{d\omega}\mathbf{c}_T(\omega) & \frac{d}{d\omega}\mathbf{s}_T(\omega) \end{bmatrix}\begin{bmatrix} \mathbf{c}_T^\top(\omega) \\ \mathbf{s}_T^\top(\omega) \end{bmatrix} - T^{-1}\begin{bmatrix} \mathbf{c}_T(\omega) & \mathbf{s}_T(\omega) \end{bmatrix}\begin{bmatrix} \frac{d}{d\omega}\mathbf{c}_T^\top(\omega) \\ \frac{d}{d\omega}\mathbf{s}_T^\top(\omega) \end{bmatrix},$$
$$\frac{d^2}{d\omega^2}\widehat{\boldsymbol{\Sigma}}_T(\omega) = -T^{-1}\begin{bmatrix} \frac{d^2}{d\omega^2}\mathbf{c}_T(\omega) & \frac{d^2}{d\omega^2}\mathbf{s}_T(\omega) \end{bmatrix}\begin{bmatrix} \mathbf{c}_T^\top(\omega) \\ \mathbf{s}_T^\top(\omega) \end{bmatrix} - T^{-1}\begin{bmatrix} \mathbf{c}_T(\omega) & \mathbf{s}_T(\omega) \end{bmatrix}\begin{bmatrix} \frac{d^2}{d\omega^2}\mathbf{c}_T^\top(\omega) \\ \frac{d^2}{d\omega^2}\mathbf{s}_T^\top(\omega) \end{bmatrix}$$
$$- 2T^{-1}\begin{bmatrix} \frac{d}{d\omega}\mathbf{c}_T(\omega) & \frac{d}{d\omega}\mathbf{s}_T(\omega) \end{bmatrix}\begin{bmatrix} \frac{d}{d\omega}\mathbf{c}_T^\top(\omega) \\ \frac{d}{d\omega}\mathbf{s}_T^\top(\omega) \end{bmatrix},$$

where

$$\frac{d}{d\omega}\mathbf{c}_T(\omega) = -\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} t\sin(\omega t)\,\mathbf{x}_t, \quad \frac{d^2}{d\omega^2}\mathbf{c}_T(\omega) = -\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} t^2\cos(\omega t)\,\mathbf{x}_t,$$

$$\frac{d}{d\omega}\mathbf{s}_T(\omega) = \left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} t\cos(\omega t)\,\mathbf{x}_t, \quad \frac{d^2}{d\omega^2}\mathbf{s}_T(\omega) = -\left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1} t^2\sin(\omega t)\,\mathbf{x}_t.$$

For $j \in \{0, 1, 2\}$, let

$$\mathbf{y}_{Tj}(\omega) = \left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1}\frac{d^j}{d\omega^j}\cos(\omega t)\,\boldsymbol{\varepsilon}_t, \quad \mathbf{z}_{Tj}(\omega) = \left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1}\frac{d^j}{d\omega^j}\sin(\omega t)\,\boldsymbol{\varepsilon}_t.$$

Then, evaluating $\mathbf{c}_T(\omega)$, $\mathbf{s}_T(\omega)$ and their derivatives at the true parameter values, we obtain

$$\mathbf{c}_T(\omega_0) = \left(\frac{T}{2}\right)^{1/2}\boldsymbol{\alpha}_0 + \mathbf{y}_{T0}(\omega_0) + O\left(T^{-1/2}\right), \qquad \mathbf{s}_T(\omega_0) = \left(\frac{T}{2}\right)^{1/2}\boldsymbol{\beta}_0 + \mathbf{z}_{T0}(\omega_0) + O\left(T^{-1/2}\right),$$

$$\frac{d}{d\omega}\mathbf{c}_T(\omega_0) = -\left(\frac{T^3}{8}\right)^{1/2}\boldsymbol{\beta}_0 + \mathbf{y}_{T1}(\omega_0) + O\left(T^{1/2}\right), \quad \frac{d}{d\omega}\mathbf{s}_T(\omega_0) = \left(\frac{T^3}{8}\right)^{1/2}\boldsymbol{\alpha}_0 + \mathbf{z}_{T1}(\omega_0) + O\left(T^{1/2}\right),$$

$$\frac{d^2}{d\omega^2}\mathbf{c}_T(\omega_0) = -\left(\frac{T^5}{18}\right)^{1/2}\boldsymbol{\alpha}_0 + \mathbf{y}_{T2}(\omega_0) + O\left(T^{3/2}\right), \quad \frac{d^2}{d\omega^2}\mathbf{s}_T(\omega_0) = -\left(\frac{T^5}{18}\right)^{1/2}\boldsymbol{\beta}_0 + \mathbf{z}_{T2}(\omega_0) + O\left(T^{3/2}\right).$$

Thus

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_T(\omega_0) &= \boldsymbol{\Sigma}_0 + O\left\{T^{-1/2}(\ln\ln T)^{1/2}\right\},\\
\frac{d}{d\omega}\widehat{\boldsymbol{\Sigma}}_T(\omega_0) &= \left(\frac{T}{8}\right)^{1/2}\boldsymbol{\beta}_0\mathbf{y}_{T0}^\top(\omega_0) - (2T)^{-1/2}\,\mathbf{y}_{T1}(\omega_0)\,\boldsymbol{\alpha}_0^\top - \left(\frac{T}{8}\right)^{1/2}\boldsymbol{\alpha}_0\mathbf{z}_{T0}^\top(\omega_0)\\
&\quad - (2T)^{-1/2}\,\mathbf{z}_{T1}(\omega_0)\boldsymbol{\beta}_0^\top + \left(\frac{T}{8}\right)^{1/2}\mathbf{y}_{T0}(\omega_0)\boldsymbol{\beta}_0^\top - (2T)^{-1/2}\,\boldsymbol{\alpha}_0\mathbf{y}_{T1}^\top(\omega_0)\\
&\quad - \left(\frac{T}{8}\right)^{1/2}\mathbf{z}_{T0}(\omega_0)\,\boldsymbol{\alpha}_0^\top - (2T)^{-1/2}\boldsymbol{\beta}_0\mathbf{z}_{T1}^\top(\omega_0) + O(\ln\ln T)\\
&= O\left\{T^{1/2}(\ln\ln T)^{1/2}\right\},\\
\frac{d^2}{d\omega^2}\widehat{\boldsymbol{\Sigma}}_T(\omega_0) &= \frac{1}{12}T^2\left(\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^\top + \boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top\right) + O\left\{T^{3/2}(\ln\ln T)^{1/2}\right\},
\end{aligned}$$

from (A.1) and since (see [10])

$$\mathbf{y}_{Tj}(\omega_0) = O\left\{T^j(\ln\ln T)^{1/2}\right\}, \quad \mathbf{z}_{Tj}(\omega_0) = O\left\{T^j(\ln\ln T)^{1/2}\right\}.$$

Therefore, as $T \to \infty$,

$$\begin{aligned}
T^{-2}\frac{d^2}{d\omega^2}K_T(\omega_0) &= \frac{1}{12}\operatorname{tr}\left[\left\{\boldsymbol{\Sigma}_0 + O\left(T^{-1/2}(\ln\ln T)^{1/2}\right)\right\}^{-1}\left(\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^\top + \boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top\right)\right] + O\left\{T^{-1/2}(\ln\ln T)^{1/2}\right\}\\
&\to \frac{1}{12}\left(\boldsymbol{\alpha}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right).
\end{aligned}$$

Also, $T^{-1/2}\frac{d}{d\omega}K_T(\omega_0)$ has the same asymptotic distribution as

$$-2^{-1/2}\boldsymbol{\alpha}_0^\top\boldsymbol{\Sigma}_0^{-1}\left\{\mathbf{z}_{T0}(\omega_0) + 2T^{-1}\mathbf{y}_{T1}(\omega_0)\right\} + 2^{-1/2}\boldsymbol{\beta}_0^\top\boldsymbol{\Sigma}_0^{-1}\left\{\mathbf{y}_{T0}(\omega_0) - 2T^{-1}\mathbf{z}_{T1}(\omega_0)\right\}.$$

Now,

$$\mathbf{z}_{T0}(\omega_0) + 2T^{-1}\mathbf{y}_{T1}(\omega_0) = \left(\frac{2}{T}\right)^{1/2}\sum_{t=0}^{T-1}\left(1 - \frac{2t}{T}\right)\sin(\omega_0 t)\,\boldsymbol{\varepsilon}_t, \tag{A.2}$$

and

$$\mathbf{y}_{T0}(\omega_0) - 2T^{-1}\mathbf{z}_{T1}(\omega_0) = \left(\frac{2}{T}\right)^{1/2} \sum_{t=0}^{T-1} \left(1 - \frac{2t}{T}\right) \cos(\omega_0 t)\, \boldsymbol{\varepsilon}_t. \tag{A.3}$$

These are both asymptotically normal with mean zero and covariance matrix $2\pi \mathbf{F}_{\varepsilon}(\omega_0)/3$. To see this, let $\zeta_t = c^\top \boldsymbol{\varepsilon}_t$, where c is a $d \times 1$ vector of constants, and consider

$$y = T^{-1/2} \sum_{t=0}^{T-1} e^{i\omega_0 t} \zeta_t, \quad z = T^{-3/2} \sum_{t=0}^{T-1} t e^{i\omega_0 t} \zeta_t.$$

From Theorem 4 of [15], both the real and imaginary components of $\begin{bmatrix} y & z \end{bmatrix}^\top$ are asymptotically normal with mean zero and covariance matrix

$$\pi f_\zeta(\omega_0) \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix},$$

where

$$f_\zeta(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_\zeta(j) e^{-ij\omega} = c^\top \left\{ \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_\varepsilon(j) e^{-ij\omega} \right\} c = c^\top \mathbf{F}_{\varepsilon}(\omega) c$$

and $\gamma_\zeta(j)$ and $\boldsymbol{\Gamma}_\varepsilon(j)$ are the autocovariance functions of $\{\zeta_t\}$ and $\{\boldsymbol{\varepsilon}_t\}$, respectively. Thus the real and imaginary components of $2^{1/2}y - 2^{3/2}z$ are both asymptotically normal with mean zero and variance

$$\pi f_\zeta(\omega_0) \begin{bmatrix} 2^{1/2} & -2^{3/2} \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \begin{bmatrix} 2^{1/2} \\ -2^{3/2} \end{bmatrix} = \frac{2\pi}{3} f_\zeta(\omega_0) = c^\top \left\{ \frac{2\pi}{3} \mathbf{F}_{\varepsilon}(\omega) \right\} c.$$

The asymptotic distributions of (A.2) and (A.3) follow by applying the Cramér-Wold device. Therefore $T^{-1/2} \frac{d}{d\omega} K(\omega_0)$ is asymptotically normal with mean zero and variance

$$\frac{\pi}{3} \left(\boldsymbol{\alpha}_0^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{F}_\varepsilon(\omega_0) \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{F}_\varepsilon(\omega_0) \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 \right).$$

It follows that $T^{3/2}(\widehat{\omega} - \omega_0)$ is asymptotically normal with mean zero and variance

$$48\pi \frac{\boldsymbol{\alpha}_0^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{F}_\varepsilon(\omega_0) \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{F}_\varepsilon(\omega_0) \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0}{\left(\boldsymbol{\alpha}_0^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 \right)^2}.$$

□

Proof of Theorem 3. From the proof of Theorem 1, as $T \to \infty$,

$$T^{-1} \widetilde{J}_{T,\boldsymbol{\Omega}}(\omega_0 + a/T) = T^{-1} \mathbf{f}_T^*(\omega_0 + a/T) \boldsymbol{\Omega} \mathbf{f}_T(\omega_0 + a/T) \to \frac{1}{2} \widetilde{\boldsymbol{\theta}}_0^\top \boldsymbol{\Omega} \widetilde{\boldsymbol{\theta}}_0 \frac{\sin^2(a/2)}{(a/2)^2} \leqslant \frac{1}{2} \widetilde{\boldsymbol{\theta}}_0^\top \boldsymbol{\Omega} \widetilde{\boldsymbol{\theta}}_0,$$

with equality if and only if $a = 0$. Thus, as $T \to \infty$,

$$T^{-1} \left\{ \widetilde{J}_T(\omega_0) - \widetilde{J}_T(\omega_0 + a/T) \right\} \to \frac{1}{2} \widetilde{\boldsymbol{\theta}}_0^\top \boldsymbol{\Omega} \widetilde{\boldsymbol{\theta}}_0 \left\{ 1 - \frac{\sin^2(a/2)}{(a/2)^2} \right\}$$

and for $\kappa > 0$

$$\lim_{T \to \infty} T^{-1} \left\{ \widetilde{J}_T(\omega_0) - \widetilde{J}_T(\omega_0 + \kappa/T) \right\} > 0.$$

It follows from Lemma 1 of [26] that $T(\widetilde{\omega} - \omega_0) \to 0$ as $T \to \infty$. □

Proof of Theorem 4. From the mean value theorem,

$$0 = \frac{d}{d\omega}\widetilde{J}_{T,\mathbf{\Omega}}(\widetilde{\omega}) = \frac{d}{d\omega}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0) + \frac{d^2}{d\omega^2}\widetilde{J}_{T,\mathbf{\Omega}}(\omega^*)(\widetilde{\omega}-\omega_0)$$

where ω^* is a point on the line segment between ω_0 and $\widetilde{\omega}$. Since $T(\widetilde{\omega}-\omega_0) \to 0$, it follows that $T^{3/2}(\widetilde{\omega}-\omega_0)$ has the same asymptotic distribution as

$$-\frac{T^{-3/2}\frac{d}{d\omega}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0)}{T^{-3}\frac{d^2}{d\omega^2}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0)}.$$

The first and second derivatives of $\widetilde{J}_{T,\mathbf{\Omega}}(\omega)$ are

$$\frac{d}{d\omega}\widetilde{J}_{T,\mathbf{\Omega}}(\omega) = 2\left[\begin{array}{cc}\frac{d}{d\omega}\mathbf{c}_T^\top(\omega) & \frac{d}{d\omega}\mathbf{s}_T^\top(\omega)\end{array}\right]\mathbf{\Omega}\left[\begin{array}{c}\mathbf{c}_T(\omega)\\ \mathbf{s}_T(\omega)\end{array}\right],$$

$$\frac{d^2}{d\omega^2}\widetilde{J}_{T,\mathbf{\Omega}}(\omega) = 2\left[\begin{array}{cc}\frac{d^2}{d\omega^2}\mathbf{c}_T^\top(\omega) & \frac{d^2}{d\omega^2}\mathbf{s}_T^\top(\omega)\end{array}\right]\mathbf{\Omega}\left[\begin{array}{c}\mathbf{c}_T(\omega)\\ \mathbf{s}_T(\omega)\end{array}\right] + 2\left[\begin{array}{cc}\frac{d}{d\omega}\mathbf{c}_T^\top(\omega) & \frac{d}{d\omega}\mathbf{s}_T^\top(\omega)\end{array}\right]\mathbf{\Omega}\left[\begin{array}{c}\frac{d}{d\omega}\mathbf{c}_T(\omega)\\ \frac{d}{d\omega}\mathbf{s}_T(\omega)\end{array}\right].$$

Thus, using the results of the proof of Theorem 2,

$$\frac{d^2}{d\omega^2}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0) = -\frac{1}{12}T^3\left(\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\boldsymbol{\beta}_0\right) + O\left\{T^{5/2}(\ln\ln T)^{1/2}\right\}$$

and so, as $T \to \infty$,

$$-T^{-3}\frac{d^2}{d\omega^2}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0) \to \frac{1}{12}\left(\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\boldsymbol{\beta}_0\right).$$

Also

$$\begin{aligned}\frac{d}{d\omega}\widetilde{J}_{T,\mathbf{\Omega}}(\omega_0) = &-\left(\frac{T^3}{2}\right)^{1/2}\boldsymbol{\beta}_0^\top\mathbf{\Omega}\mathbf{y}_{T0}(\omega_0) - (2T)^{1/2}\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\mathbf{y}_{T1}(\omega_0)\\ &+\left(\frac{T^3}{2}\right)^{1/2}\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\mathbf{z}_{T0}(\omega_0) + (2T)^{1/2}\boldsymbol{\beta}_0^\top\mathbf{\Omega}\mathbf{z}_{T1}(\omega_0) + O(T\ln\ln T),\end{aligned}$$

and hence $T^{-3/2}\frac{d}{d\omega}\widetilde{J}_T(\omega_0)$ has the same asymptotic distribution as

$$2^{-1/2}\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\left\{\mathbf{z}_{T0}(\omega_0) + 2T^{-1}\mathbf{y}_{T1}(\omega_0)\right\} - 2^{-1/2}\boldsymbol{\beta}_0^\top\mathbf{\Omega}\left\{\mathbf{y}_{T0} - 2T^{-1}\mathbf{z}_{T1}(\omega_0)\right\},$$

which, as shown in the proof of Theorem 2, is asymptotically normal with mean zero and variance

$$\frac{\pi}{3}\left\{\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\mathbf{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}F_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\beta}_0\right\}.$$

Therefore $T^{3/2}(\widetilde{\omega}-\omega_0)$ is asymptotically normal with mean zero and variance

$$48\pi\frac{\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\mathbf{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\mathbf{F}_\varepsilon(\omega_0)\mathbf{\Omega}\boldsymbol{\beta}_0}{\left(\boldsymbol{\alpha}_0^\top\mathbf{\Omega}\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\mathbf{\Omega}\boldsymbol{\beta}_0\right)^2}.$$

□

Proof of Theorem 5. Let $\mathbf{\Theta}_0 = \left[\begin{array}{cccccc}\boldsymbol{\mu}_0^\top & \widetilde{\boldsymbol{\delta}}_0^\top & \widetilde{\mathbf{G}}_0^\top & \boldsymbol{\alpha}_0^\top & \boldsymbol{\beta}_0^\top & \omega_0\end{array}\right]$ denote the true value of $\mathbf{\Theta}$ and

$$\mathcal{I} = \lim_{T\to\infty} -\mathbf{N}_T^{-1}\frac{\partial^2 l(\mathbf{\Theta}_0)}{\partial\mathbf{\Theta}_0\partial\mathbf{\Theta}_0^\top}\mathbf{N}_T^{-1},$$

where

$$\mathbf{N}_T = \left[\begin{array}{cc}T^{1/2}I_{3d+(p+1)d^2} & 0\\ 0 & T^{3/2}\end{array}\right].$$

The $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and ω components of $\mathcal{I}$ are

$$J = \frac{1}{2}\begin{bmatrix} \boldsymbol{\Delta}_1 & \boldsymbol{\Delta}_2 & \frac{1}{2}\left(-\boldsymbol{\Delta}_2\boldsymbol{\alpha}_0 + \Delta_1\boldsymbol{\beta}_0\right) \\ -\boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_1 & \frac{1}{2}\left(-\boldsymbol{\Delta}_2\boldsymbol{\beta}_0 - \Delta_1\boldsymbol{\alpha}\right) \\ \frac{1}{2}\left(\boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_2 + \boldsymbol{\beta}_0^\top\Delta_1\right) & \frac{1}{2}\left(\boldsymbol{\beta}_0^\top\boldsymbol{\Delta}_2 - \boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_1\right) & \frac{1}{3}\left(\boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_1\boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0^\top\boldsymbol{\Delta}_1\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^\top\boldsymbol{\Delta}_2\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_2\boldsymbol{\beta}_0\right) \end{bmatrix},$$

where

$$\boldsymbol{\Delta}_1 = \boldsymbol{\delta}_R^\top\mathbf{G}_0^{-1}\boldsymbol{\delta}_R + \boldsymbol{\delta}_I^\top\mathbf{G}_0^{-1}\boldsymbol{\delta}_I, \quad \boldsymbol{\Delta}_2 = -\boldsymbol{\delta}_I^\top\mathbf{G}_0^{-1}\boldsymbol{\delta}_R + \boldsymbol{\delta}_R^\top\mathbf{G}_0^{-1}\boldsymbol{\delta}_I,$$

$$\boldsymbol{\delta}_R = \mathrm{Re}\left(\mathbf{I}_d + \sum_{j=1}^{p}\boldsymbol{\delta}_{0j}e^{-ij\omega}\right), \quad \boldsymbol{\delta}_I = \mathrm{Im}\left(\mathbf{I}_d + \sum_{j=1}^{p}\boldsymbol{\delta}_{0j}e^{-ij\omega}\right).$$

The non-diagonal blocks of the $\boldsymbol{\mu}$, $\widetilde{\boldsymbol{\delta}}$ and $\widetilde{\mathbf{G}}$ components of $\mathcal{I}$ are zero. Let

$$\boldsymbol{\tau} = \begin{bmatrix} \boldsymbol{\Delta}_1 & \boldsymbol{\Delta}_2 \\ -\boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_1 \end{bmatrix}, \quad \boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{\beta}_0 \\ -\boldsymbol{\alpha}_0 \end{bmatrix}.$$

Then, since $\boldsymbol{\tau}$ is symmetric,

$$J = \frac{1}{2}\begin{bmatrix} \boldsymbol{\tau} & \frac{1}{2}\boldsymbol{\tau\varphi} \\ \frac{1}{2}\boldsymbol{\varphi}^\top\boldsymbol{\tau} & \frac{1}{3}\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi} \end{bmatrix}$$

and so, from the matrix inversion lemma,

$$J^{-1} = 2\begin{bmatrix} \boldsymbol{\tau}^{-1} + \frac{1}{4}\boldsymbol{\tau}^{-1}(\boldsymbol{\tau\varphi})\mathbf{H}(\boldsymbol{\varphi}^\top\boldsymbol{\tau})\boldsymbol{\tau}^{-1} & -\frac{1}{2}\boldsymbol{\tau}^{-1}(\boldsymbol{\tau\varphi})\mathbf{H} \\ -\frac{1}{2}\mathbf{H}(\boldsymbol{\varphi}^\top\boldsymbol{\tau})\boldsymbol{\tau}^{-1} & \mathbf{H} \end{bmatrix},$$

where

$$\mathbf{H}^{-1} = \left\{\frac{1}{3}\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi} - \frac{1}{4}\left(\boldsymbol{\varphi}^\top\boldsymbol{\tau}\right)\boldsymbol{\tau}^{-1}(\boldsymbol{\tau\varphi})\right\} = \frac{1}{12}\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi}.$$

That is,

$$J^{-1} = \begin{bmatrix} 2\boldsymbol{\tau}^{-1} + \frac{6}{\boldsymbol{\theta}^\top\boldsymbol{\tau\theta}}\boldsymbol{\varphi\varphi}^\top & -\frac{12}{\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi}}\boldsymbol{\varphi} \\ -\frac{12}{\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi}}\boldsymbol{\varphi}^\top & \frac{24}{\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi}} \end{bmatrix}.$$

The first derivatives of $l(\boldsymbol{\Theta})$ with respect to $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and ω at $\boldsymbol{\Theta}_0$ are

$$\frac{\partial l(\boldsymbol{\Theta}_0)}{\partial\boldsymbol{\alpha}} = \sum_{t=0}^{T-1}\left[\sum_{j=0}^{p}\boldsymbol{\delta}_{0j}\cos\{\omega_0(t-j)\}\right]\mathbf{G}_0^{-1}\mathbf{u}_t,$$

$$\frac{\partial l(\boldsymbol{\Theta}_0)}{\partial\boldsymbol{\beta}} = \sum_{t=0}^{T-1}\left[\sum_{j=0}^{p}\boldsymbol{\delta}_{0j}\sin\{\omega_0(t-j)\}\right]\mathbf{G}_0^{-1}\mathbf{u}_t,$$

$$\frac{\partial l(\boldsymbol{\Theta}_0)}{\partial\omega} = -\sum_{t=0}^{T-1}\sum_{j=0}^{p}\boldsymbol{\delta}_{0j}\left[\boldsymbol{\alpha}_0(t-j)\sin\{\omega_0(t-j)\} - \boldsymbol{\beta}_0(t-j)\cos\{\omega_0(t-j)\}\right]\mathbf{G}_0^{-1}\mathbf{u}_t.$$

Thus, since $E(\mathbf{u}_t\mathbf{u}_t^\top) = \mathbf{G}_0$,

$$E\left\{\frac{\partial l(\boldsymbol{\Theta}_0)}{\partial\widetilde{\boldsymbol{\theta}}}\frac{\partial l(\boldsymbol{\Theta}_0)}{\partial\widetilde{\boldsymbol{\theta}}^\top}\right\} = J.$$

It follows from the martingale central limit theorem [3] that $T^{3/2}(\widehat{\omega} - \omega_0)$ is asymptotically normal with mean zero and variance $24/\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi}$. From (5),

$$\boldsymbol{\varphi}^\top\boldsymbol{\tau\varphi} = \boldsymbol{\beta}_0^\top\boldsymbol{\Delta}_1\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_2\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^\top\boldsymbol{\Delta}_2\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_0^\top\boldsymbol{\Delta}_1\boldsymbol{\alpha}_0 = \frac{1}{2\pi}\boldsymbol{\eta}_0^*\mathbf{F}_{\varepsilon}^{-1}(\omega_0)\boldsymbol{\eta}_0.$$

□

Appendix B. Supplementary material

Supplementary material related to this article can be found online.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (1974) 716–723.
[2] H. Akaike, A Bayesian analysis of the minimum AIC procedure, Ann. Inst. Statist. Math. 30 (1978) 9–14.
[3] P. Billingsley, The Lindeberg-Lévy theorem for martingales, Proc. Amer. Math. Soc. 12 (1961) 788–792.
[4] J. Durbin, Efficient estimation of parameters in moving-average models, Biometrika 46 (1959) 306–316.
[5] J. Durbin, The fitting of time-series models, Int. Stat. Rev. 28 (1960) 233–244.
[6] J. E. Gentle, Matrix Algebra : Theory, Computations and Applications in Statistics, Springer International Publishing, Cham, 2nd edition, 2017.
[7] E. J. Hannan, Multiple Time Series, Wiley, New York, 1970.
[8] E. J. Hannan, The estimation of frequency, J. Appl. Probab. 10 (1973) 510–519.
[9] E. J. Hannan, L. Kavalieris, Multivariate linear time series models, Advances in Applied Probability 16 (1984) 492–561.
[10] E. J. Hannan, M. Mackisack, A law of the iterated logarithm for an estimate of frequency, Stochastic Process. Appl. 22 (1986) 103–109.
[11] E. J. Hannan, B. G. Quinn, The determination of the order of an autoregression, J. R. Statist. Soc. B 41 (1979) 190–195.
[12] M. B. Priestley, Spectral Analysis and Time Series, Academic Press, London, 1981.
[13] B. G. Quinn, Order determination for a multivariate autoregression, J. R. Statist. Soc. B 42 (1980) 182–185.
[14] B. G. Quinn, The estimation of frequency, in: T. S. Rao, S. S. Rao, C. Rao (Eds.), Time Series Analysis: Methods and Applications, volume 30 of *Handbook of Statistics*, Elsevier, 2012, pp. 585–621.
[15] B. G. Quinn, E. J. Hannan, The Estimation and Tracking of Frequency, Cambridge University Press, Cambridge, 2001.
[16] B. G. Quinn, P. J. Kootsookos, Threshold behavior of the maximum likelihood estimator of frequency, IEEE Trans. Signal Process. 42 (1994) 3291–3294.
[17] B. G. Quinn, R. G. McKilliam, I. V. L. Clarkson, Maximizing the periodogram, in: 2008 IEEE Global Telecommunications Conference, pp. 1–5.
[18] G. C. Reinsel, Elements of Multivariate Time Series Analysis, Springer-Verlag, New York, 1993.
[19] J. A. Rice, M. Rosenblatt, On frequency estimation, Biometrika 75 (1988) 477–484.
[20] H. Sakai, The determination of the number of terms in a multichannel sinusoidal regression, J. Time Series Anal. 14 (1993) 621–628.
[21] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[22] W. F. Stout, The Hartman-Wintner law of the iterated logarithm for martingales, Ann. Math. Statist. 41 (1970) 2158–2160.
[23] A. M. Walker, On the estimation of a harmonic component in a time series with stationary independent residuals, Biometrika 58 (1971) 21–36.
[24] P. Whittle, The simultaneous estimation of a time series harmonic components and covariance structure, Trabajos Estadist. 3 (1952) 43–57.
[25] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, Biometrika 50 (1963) 129–134.
[26] C.-F. Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.