Electronic Journal of Statistics
Vol. 14 (2020) 605–631
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1672

Generalised cepstral models for the spectrum of vector time series

Maddalena Cavicchioli

Department of Economics "Marco Biagi", University of Modena and Reggio E.
V.le Berengario 51, 41121 Modena, Italy
e-mail: maddalena.cavicchioli@unimore.it

Abstract: The paper treats the modeling of stationary multivariate stochastic processes via a frequency domain model expressed in terms of cepstrum theory. The proposed model nests the vector exponential model of [20] as a special case, and extends the generalised cepstral model of [36] to the multivariate setting, answering a question raised by the last authors in their paper. Contemporarily, we extend the notion of generalised autocovariance function of [35] to vector time series. Then we derive explicit matrix formulas connecting generalised cepstral and autocovariance matrices of the process, and prove the consistency and asymptotic properties of the Whittle likelihood estimators of model parameters. Asymptotic theory for the special case of the vector exponential model is a significant addition to the paper of [20]. We also provide a mathematical machinery, based on matrix differentiation, and computational methods to derive our results, which differ significantly from those employed in the univariate case. The utility of the proposed model is illustrated through Monte Carlo simulation from a bivariate process characterized by a high dynamic range, and an empirical application on time varying minimum variance hedge ratios through the second moments of future and spot prices in the corn commodity market.

MSC 2010 subject classifications: 62M10, 62M15, 62H12.
Keywords and phrases: Stationary vector stochastic processes, spectral density matrix, Wold coefficients, generalised spectrum models, Box-Cox link, spectral estimation, Whittle likelihood.

Received October 2019.

1. Introduction

Vector stationary autoregressive moving-average (VARMA) models and their generalizations have been used extensively to modeling economic, financial and statistical time series. Such models are formulated in the time domain as difference equations with a corresponding covariance structure. However, to fit these models, it is necessary to impose some restrictions on the coefficient matrices to ensure that certain determinants are non trivial. This may increase computational cost, except in special cases. To avoid such difficulties, many authors have considered stationary processes in the frequency domain. Indeed, the spectral density provides a complete characterization of the serial correlation structure of the process, as well as, the necessary information for prediction and interpolations. Tractable methods to derive the spectral representations of a general class of Markov switching (MS) VARMA models have been proposed by [33], [7], and [10]. Spectral representation and autocovariance structure of MS Dynamic Stochastic General Equilibrium (DSGE) models are recently investigated in [8]. A further advantage of spectral analysis is that it can reveal detailed features of real data without using a parametric model. The only assumption needed is the stationarity of the series. A fresh insight can be gained into the time series structure and cyclic behaviour at different time scale, such as seasonal patterns and business cycle. This serves to derive some goodness-of-fit tests on MS models based on spectral density functions.

Frequency domain analysis has become a fundamental component of many scientific inquiries and empirical economic applications. [38] introduce a class of nonlinear multivariate time-frequency functional models to study inference for general scientific processes, especially ecological. An approach for identifying and predicting economic recessions in real-time by using time-frequency functional models has been proposed by [22]. Their methodology extracts information embedded in time-frequency representation of daily returns for stock market indices, and specifically reveals important features of daily NASDAQ index log returns corresponding to economic recessions. See also [15] for a discussion on the last quoted paper. [21] utilize time-frequency functional models to discover several features of animal communication signals, which were overlooked in previous time-domain analyses. [31] propose univariate generalized Gegenbauer processes as a flexible way of modeling time series data which exhibit long memory and seasonal long memory. We refer to [30], [17], and [25] for comprehensive monographs dealing various classes of generalised statistical models that have been found useful in statistical and economic analyses.

One of the most important time-frequency model for scalar times series has been proposed by [4]. It is called the (scalar) exponential model, and arises by truncating the Fourier series expansion of the scalar log-spectrum. The corresponding coefficients are called cepstral coefficients, and their collection constitutes the cepstrum of the model. The cepstral coefficients are then obtained from the discrete Fourier transform of the log-spectrum. These terms were introduced by [3], where cepstral and cepstrum are anagrams of spectral and spectrum, respectively.

[32] study the cepstral random field model, providing recursive formulas which connected the spatial cepstral coefficients to an equivalent moving-average (MA) random field. This facilitates easy computation of the autocovariance matrix, and serves to establish asymptotic results for Bayesian and maximum likelihood (ML) estimators of model parameters.

[19] and [20] treat the modeling of stationary vector time series through a multivariate extension of the scalar exponential model of [4]. Such a process is called vector exponential (VEXP) model. In contrast to VARMA models, the VEXP processes are always stable and invertible. In particular, this implies that the spectral density matrix is nonsingular at all frequencies. Furthermore, the class of VEXP processes is arbitrarily dense in the space of stable invertible vector time series. These authors provide precise mathematical development

and computational algorithms for the proposed model, which differs significantly from the univariate case.

More recently, [36] introduce a class of univariate (generalised) cepstral models by using the Box-Cox transformation of the spectral density. See also [5]. Their class includes various univariate models, such as, the univariate exponential model of [4], and standard ARMA processes. The generalised cepstral coefficients of the proposed model are shown to be related with the generalised autocovariance function of a stationary univariate stochastic process. Such a function has been introduced and studied in a previous paper of the same authors. See [35]. Consistency, asymptotic theory of parameter estimators, Monte Carlo simulations, and empirical applications illustrating the flexibility of the univariate generalised spectral model complete the nice paper of [36].

The main goal of the present paper is to obtain new results on the multivariate specification of the generalised cepstral model introduced by [36]. This answers a suggestion given by those authors in the conclusion section of the quoted paper. We also take advantage from methods and techniques developed by [11] and [20] in the multivariate case. Contemporarily, we show that the proposed multivariate model extends the notion of generalised autocovariance function to vector time series. This, in turn, generalises results obtained by [35] for univariate stochastic processes, and illustrates the utility of the vector model. We also derive explicit matrix relationships between generalised cepstral and autocovariance matrices of the considered vector process. Furthermore, we prove consistency and asymptotic properties of the Whittle likelihood estimators of model parameters. The proposed class of models includes various multivariate stochastic processes, such as, the vector exponential model of [20], and standard VARMA models. Asymptotic theory for the special case of the vector exponential model is a significant addition to the paper of [20]. Finally, we provide a mathematical machinery, based on matrix differentiation, and computational methods, which are very different to those employed in the univariate case.

The paper is organized as follows. In Section 2 we introduce the model, give some preliminaries and definitions of cepstrum theory, and establish notations. Section 3 is devoted to define the generalised cepstral matrices and the generalised matrix cepstrum of a stationary vector stochastic process, and to describe their statistical properties. We show that the generalised matrix spectrum contains all information for prediction and feature extraction of the considered process, and uniquely characterises its spectral properties. Whittle estimation of the proposed vector model, consistency and asymptotic results on the estimators of model parameters are presented in Section 4. Here we formulate the main result (Theorem 1, §4) concerning the closed form expression for the asymptotic covariance matrix of the generalised matrix cepstrum model. The proof of the main theorem is given in Appendix A. Section 5 provides a Monte Carlo simulation from a bivariate VAR(4) process and illustrates the computation of the asymptotic covariance matrix of the estimated parameters (the obtained results are given in Appendix B). In Section 5 we also present an empirical example on the hedge ratios through the second moments of futures and spot prices in the corn commodity market. Section 6 concludes with remarks. For the basic identities and results on matrix calculus the reader is addressed to [14], [26], and [27, 28]. General discussion concerning vector time series and spectral theory is provided in [6], and [18].

2. Generalised linear models for the spectrum

Let $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}}$ be a n-dimensional stationary zero-mean stochastic process. The spectral density matrix (sdm) of $\mathbf{x}$ is a $(n \times n)$ dimensional matrix function given by

$$\mathbf{f}(\omega) \;=\; \sum_{k=-\infty}^{\infty} \mathbf{\Gamma}_k \, z^k \tag{1}$$

where $\mathbf{\Gamma}_k = E(\mathbf{x}_t \, \mathbf{x}'_{t-k})$ is the autocovariance function (acf), $z = \exp(-\mathbf{i}\,\omega)$, $\omega \in [-\pi, \pi]$, and $\mathbf{i}$ denotes the imaginary unit. Assume that $\mathbf{f} = \mathbf{f}(\omega)$ is positive definite (pd) for every $\omega \in [-\pi, \pi]$. For a n-variate time series, $\mathbf{f}(\omega)$ is always Hermitian non negative definite, and is often positive definite. Of course, this may not always be the case. Conditions for positive definiteness are provided after formula (5) below.

Taking the Fourier transform (FT) of $\mathbf{f}$ gives the acf of $\mathbf{x}$:

$$\mathbf{\Gamma}_k \;=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f}(\omega)\, z^{-k} \, d\omega \tag{2}$$

where the integration works component-wise on each entry of the sdm.

Assuming that the process $\mathbf{x}$ is purely nondeterministic, its Wold representation can be written as

$$\mathbf{x}_t \;=\; \mathbf{\Psi}(B)\, \boldsymbol{\epsilon}_t \tag{3}$$

where B is the backshift operator, i.e., $B^k \, \mathbf{x}_t = \mathbf{x}_{t-k}$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_t)$ is a white noise process with zero-mean and positive definite autovariance matrix $\mathbf{\Omega}$, i.e., $\boldsymbol{\epsilon}_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Omega})$, and

$$\mathbf{\Psi}(z) \;=\; \sum_{j\geq 0} \mathbf{\Psi}_j \, z^j \tag{4}$$

is the causal representation of $\mathbf{x}$ (for identifiability, set $\mathbf{\Psi}_0 = \mathbf{I}_n$, the identity $n \times n$ matrix). Assume that $\det \mathbf{\Psi}(z) \neq 0$ for all $|z| \leq 1$, and the matrix coefficients satisfy $\sum_{j\geq 0} j \, ||\mathbf{\Psi}_j||^2 < \infty$.

Then the sdm is also given by

$$\mathbf{f}(\omega) \;=\; \mathbf{\Psi}(z)\, \mathbf{\Omega}\, \mathbf{\Psi}'(\overline{z}) \tag{5}$$

where $\overline{z}$ denotes the complex conjugate of z, as usual. Setting $\mathbf{\Lambda}(z) = \mathbf{\Psi}(z)\, \mathbf{\Omega}^{\frac{1}{2}}$ implies that $\mathbf{f}(\omega) = \mathbf{\Lambda}(z)\, \mathbf{\Lambda}'(\overline{z})$. Here $\mathbf{\Omega}^{\frac{1}{2}}$ denotes the (unique) square root of the pd matrix $\mathbf{\Omega}$. Then a necessary and sufficient condition for $\mathbf{f}(\omega)$ to be pd is that $\mathbf{\Lambda}(z)$ (or a suitable truncation of it) is non singular (at least, up to a set of frequencies that have Lebesgue measure zero). Of course, the positivity of any nested sequence of n principal minors of $\mathbf{f}(\omega)$ is a necessary and sufficient condition for positive definiteness. A computational, simple and fast sufficient criterion to verify positive definiteness of a Hermitian matrix has been described in [37]. Computation geometry of positive definiteness and algorithms to solve the maximal least eigenvalue problem can be found in [23]. A lot of numerical algorithms for correcting non positive definite matrices into pd ones has been collected in [29] (§4).

The acf is related to the matrix coefficients of the Wold filter $\mathbf{\Psi}(B)$ by

$$\mathbf{\Gamma}_k \;=\; \sum_{j\geq 0} \mathbf{\Psi}_{j+k}\, \mathbf{\Omega}\, \mathbf{\Psi}'_j \tag{6}$$

for any integer $k \geq 0$, and $\mathbf{\Gamma}_k \;=\; \mathbf{\Gamma}'_{-k}$ for $k < 0$.

Extending the work of [35, 36] and [19, 20], we consider different representations of the sdm that involve the matrix power and the matrix exponential. Since the sdm is pd by assumption, there are orthogonal matrices (as functions of frequency ω) that diagonalize $\mathbf{f}(\omega)$ for every $\omega \in [-\pi, \pi]$.

By (1) the matrix $\mathbf{f} = \mathbf{f}(\omega)$ is Hermitian for every $\omega \in [-\pi, \pi]$, hence we can write

$$\mathbf{f} \;=\; \mathbf{P}\,\mathbf{A}\,\mathbf{P}^* \tag{7}$$

where $\mathbf{P} \;=\; \mathbf{P}(\omega)$ is a unitary complex-valued matrix, i.e., $\mathbf{P}\,\mathbf{P}^* \;=\; \mathbf{I}_n$, and $\mathbf{A} = \mathbf{A}(\omega)$ is diagonal with real positive entries. Here $\mathbf{P}^*$ denotes the conjugate transpose of $\mathbf{P}$, as usual.

The λ*th matrix power* of $\mathbf{f}$ is defined by

$$\mathbf{f}^{\lambda} \;=\; \mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^* \tag{8}$$

for $\lambda \in \mathbb{R}$, where $\mathbf{A}^{\lambda}$ is the diagonal matrix consisting of the λth powers of the entries of $\mathbf{A}$. Then we can write

$$\mathbf{f} \;=\; \left(\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^*\right)^{1/\lambda} \tag{9}$$

that is, the sdm $\mathbf{f}$ of $\mathbf{x}$ is the matrix power at exponent $1/\lambda$ of the pd matrix $\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^*$.

By Artin (1991, p. 139), the *matrix exponential* of $\mathbf{f}$ is given by

$$\exp(\mathbf{f}) \;=\; \mathbf{P}\,\exp(\mathbf{A})\,\mathbf{P}^* \tag{10}$$

where $\exp(\mathbf{A})$ is the diagonal matrix consisting of the exponential of the entries of $\mathbf{A}$.

In particular, we can write (see Holan *et al.*, 2017)

$$\log_e(\mathbf{f}) \;=\; \mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^* \tag{11}$$

hence

$$\mathbf{f} \;=\; \exp[\mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^*] \tag{12}$$

where the diagonal matrix $\log_e(\mathbf{A})$ consists of the logged entries of $\mathbf{A}$. In other words, by (12) the sdm $\mathbf{f}$ of $\mathbf{x}$ is the matrix exponential of the Hermitian matrix $\mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^*$, which is not pd in general.

The matrix functions in (8) and (11) can be expanded in the Hilbert space with basis $(z^k)_{k\in\mathbb{Z}}$ as

$$\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^{*} \;=\; \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_{\lambda k}\, z^{k} \tag{13}$$

and

$$\mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^{*} \;=\; \sum_{k=-\infty}^{\infty} \mathbf{C}_{k}\, z^{k}. \tag{14}$$

Such expansions can be calculated by determining the matrix coefficients $\boldsymbol{\Gamma}_{\lambda k}$ and $\mathbf{C}_k$ for $\lambda \in \mathbb{R}$ and $k \in \mathbb{Z}$ by inverse FT of $\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^{*}$ and $\mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^{*}$, respectively. Then we have

$$\boldsymbol{\Gamma}_{\lambda k} \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} (\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^{*})(\omega)\, z^{-k}\, d\omega \tag{15}$$

and

$$\mathbf{C}_{k} \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} [\mathbf{P}\,\log_e(\mathbf{A})\,\mathbf{P}^{*}](\omega)\, z^{-k}\, d\omega. \tag{16}$$

For univariate time series, the scalars $(\gamma_{\lambda k})$ from (15) give the *generalised autocovariance function* (gacf) introduced and studied by [35]. However, in the present context they are matrices, and we call them *generalised autocovariance matrices* (gacm). The matrix coefficients $\mathbf{C}_k$ in (16) are the *cepstral matrices* (cm) defined in [20], and denoted there by $\boldsymbol{\Theta}_k$, for $k \in \mathbb{Z}$. For univariate time series, the scalars (c_k) from (16) give the *cepstral coefficients* also considered in [36].

The gacm $\boldsymbol{\Gamma}_{\lambda k}$ depend on two arguments, the integer lag k and the real power λ. For $\lambda = 1$, $\boldsymbol{\Gamma}_{1k} = \boldsymbol{\Gamma}_k$ is the acf of the process $\mathbf{x}$ at lag k. For $\lambda = 0$, $\boldsymbol{\Gamma}_{0k} = \mathbf{0}_{n\times n}$, the null $(n \times n)$ matrix, for $k \neq 0$, and $\boldsymbol{\Gamma}_{00} = \mathbf{I}_n$ (up to a constant), hence we get the acf of a white noise process $\mathrm{WN}(\mathbf{0}, \mathbf{I}_n)$. For $\lambda = -1$, $\boldsymbol{\Gamma}_{-1k} = \boldsymbol{\Gamma}_{\mathrm{inv},k}$ extend to the multivariate case the *inverse autocovariance function* considered in [12]. See also [2].

Finally, from (9) and (12) we obtain the formal expressions

$$\mathbf{f} \;=\; \left(\sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_{\lambda k}\, z^{k}\right)^{1/\lambda} \tag{17}$$

and

$$\mathbf{f} \;=\; \exp\left(\sum_{k=-\infty}^{\infty} \mathbf{C}_{k}\, z^{k}\right). \tag{18}$$

Thus a generic sdm $\mathbf{f}(\omega)$ of a covariance stationary vector time series $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}}$ can be written in the forms (17) and (18), using the matrix power and the matrix exponential, respectively. Equation (18) arises from [20], by setting $\mathbf{C}_k = \boldsymbol{\Theta}_k$, for $k \in \mathbb{Z}$. Equation (17) extends the gacf from [35] to the multivariate setting.

3. Generalised cepstral matrices and matrix cepstrum

Extending the univariate case of [36] to multivariate time series, consider the Box-Cox transform (see [5]) of the sdm $\mathbf{f}(\omega)$ with transformation parameter $\lambda \in \mathbb{R}$

$$\mathbf{g}_\lambda(\omega) \;=\; \begin{cases} \frac{1}{\lambda}\left[\mathbf{f}^\lambda(\omega) \,-\, \mathbf{I}_n\right] & \lambda \neq 0 \\ \log_e \, \mathbf{f}(\omega) & \lambda = 0. \end{cases} \tag{19}$$

For every $\omega \in [-\pi, \pi]$, $\mathbf{g}_\lambda(\omega)$ is a well-defined $(n \times n)$ matrix, which we call the *generalised cepstral density matrix* (gcdm) of the process $\mathbf{x}$. In Appendix A, we show that

$$\lim_{\lambda \to 0} \frac{1}{\lambda}\left[\mathbf{f}^\lambda(\omega) \,-\, \mathbf{I}_n\right] \;=\; \log_e \, \mathbf{f}(\omega) \tag{20}$$

hence $\mathbf{g}_\lambda(\omega)$ is a continuous matrix function with respect to frequency ω.

By (13) and (14) the gcdm $\mathbf{g}_\lambda(\omega)$ can be represented as

$$\mathbf{g}_\lambda(\omega) \;=\; \sum_{k=-\infty}^{\infty} \mathbf{C}_{\lambda k} \, z^k \tag{21}$$

which is linear in the matrix coefficients $\mathbf{C}_{\lambda k}$

$$\mathbf{C}_{\lambda k} \;=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{g}_\lambda(\omega) \, z^{-k} \, d\omega \tag{22}$$

given by the inverse FT of $\mathbf{g}_\lambda(\omega)$.

When λ equals zero, we obtain the exponential model of [20] as a special case, that is, $\mathbf{C}_{0k} = \mathbf{C}_k$ from (14), for $k \in \mathbb{Z}$. See also [4] for univariate time series. In the univariate case, the scalars $(c_{\lambda k})$ from (22) are the *generalised cepstral coefficients* at lag k introduced and studied by [36]. However, in the present context they are $(n \times n)$ matrices, which we call *generalised cepstral matrices* (gcm).

Assume that $\mathbf{g}_\lambda(\omega)$ can be represented by a finite polynomial

$$\mathbf{g}_\lambda(\omega) \;=\; \sum_{k=-K}^{K} \mathbf{C}_{\lambda k} \, z^k \tag{23}$$

for some positive integer K. Then the set of $(n \times n)$ matrices

$$\mathcal{C}_{\mathbf{x},K} \;= \{\mathbf{C}_{\lambda k} \,:\, \lambda \in \mathbb{R},\, k = 0, \ldots, K\}$$

is called the *generalised matrix cepstrum* (of order K) for the process $\mathbf{x}$. Later, we show that $\mathbf{C}_{\lambda,-k} = \mathbf{C}'_{\lambda k}$ for every integer $k < 0$.

The spectral model with Box-Cox link and mean function

$$\mathbf{f}(\omega) \;=\; \begin{cases} \left[\mathbf{I}_n \,+\, \lambda \, \mathbf{g}_\lambda(\omega)\right]^{\frac{1}{\lambda}} & \lambda \neq 0 \\ \exp \, \mathbf{g}_\lambda(\omega) & \lambda = 0 \end{cases} \tag{24}$$

will be referred to as *generalised matrix cepstral model* for the spectrum with parameter $\lambda \in \mathbb{R}$ and order K. In short, we denote it by GMCM(λ, K). Equation (24) is well-defined. See Appendix A for its computation.

To our knowledge, the idea of the generalised matrix cepstrum is completely new in the literature, and is a nontrivial multivariate extension (based on differential matrix calculus) of the univariate model from [36]. Moreover, the asymptotics for the parameter vector by using Whittle likelihood method provides matrix expressions in closed form, which can not be easily derived from the univariate case (see the statement of Theorem 1, §4). This asymptotic theory also gives a significant contribution to the paper of [20], where such a theory has not been discussed in the case of the vector exponential model (see Theorem 2, §4).

Remark 1. The GMCM(λ, K) assumes that the spectrum can be represented by a finite set of $(n \times n)$ matrix coefficients. These matrices represent the Fourier coefficients of the Box-Cox transformation of the spectrum.

For example, consider the GMCM(λ, K) for $\lambda = K = 1$, that is,

$$\mathbf{f}(\omega) = \mathbf{I}_n + \mathbf{g}_1(\omega) = \mathbf{I}_n + \mathbf{C}_{1,0} + \mathbf{C}_{1,-1}\,\overline{z} + \mathbf{C}_{1,1}\,z$$

where $z = \exp(-\mathbf{i}\,\omega)$ and $\overline{z} = z^{-1}$. The right-hand side is the sdm of the 1st order vector moving-average (VMA) process $\mathbf{x}_t = \boldsymbol{\epsilon}_t + \boldsymbol{\Psi}\,\boldsymbol{\epsilon}_{t-1}$ such that

$$\mathbf{I}_n + \mathbf{C}_{1,0} = \boldsymbol{\Omega} + \boldsymbol{\Psi}\,\boldsymbol{\Omega}\,\boldsymbol{\Psi}' \qquad \mathbf{C}_{1,1} = \boldsymbol{\Psi}\,\boldsymbol{\Omega} = \mathbf{C}'_{1,-1}.$$

More generally, GMCM$(1, K)$ is a VMA(K) model. Conversely, GMCM$(-1, K)$ is a vector autoregressive (VAR) model of order K.

For example, the sdm of GMCM$(-1, 1)$ is given by

$$\mathbf{f}(\omega) = [\mathbf{I}_n - \mathbf{g}_{-1}(\omega)]^{-1} = [\mathbf{I}_n - \mathbf{C}_{-1,0} - \mathbf{C}_{-1,-1}\,\overline{z} - \mathbf{C}_{-1,1}\,z]^{-1}.$$

This is also the sdm of the VAR(1) process $\mathbf{x}_t = \boldsymbol{\Phi}\,\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t$ with

$$\mathbf{I}_n - \mathbf{C}_{-1,0} = \boldsymbol{\Omega}^{-1} + \boldsymbol{\Phi}'\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Phi} \qquad \mathbf{C}_{-1,1} = \boldsymbol{\Omega}^{-1}\,\boldsymbol{\Phi} = \mathbf{C}'_{-1,-1}.$$

Remark 2. The methodology in this paper is based on the assumption that the process $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}}$ is stationary. This is also a standard assumption in [35, 36] and [20]. As pointed out by one of the referees, the resulting generalised matrix cepstral model (GMCM) is more general than the often used vector ARMA models, but it is hard to impose the conditions on $\mathbf{C}_{\lambda\,k}$ such that the GMCM is stationary. This is not a problem for VARMA models, or other non-linear vector time series models. From Remark 1, it is clear that some conditions on $\mathbf{C}_{\lambda\,k}$ are needed for the stationarity of $\mathbf{x}$, but these conditions can not be easily derived for the general GMCM. Thus we can circumvent this problem to make the data stationary by differencing or capturing a common trend. Furthermore, if a Bayesian treatment is desired, prior parameter restrictions may cause some difficulties since they imply that the model parameters must be supported on a complicated manifold. However, the cepstral approach allows for the entries of each parameter matrix to be any real number, so that taking independent vague Gaussian priors is a coherent choice that guarantees a stable outcome. See [20] (§3.2).

To complete the section we derive some properties of the generalised cepstral matrices. It is shown that $\mathcal{C}_{\mathbf{x},K}$ contains all the necessary information for prediction and feature extraction of the process $\mathbf{x}$, and uniquely characterises its spectral properties.

Any sdm is Hermitian and has the property $\mathbf{f}(-\omega) = \mathbf{f}'(\omega)$. This also holds for $\mathbf{f}^\lambda$ and $\log_e(\mathbf{f})$, hence for $\mathbf{g}_\lambda(\omega)$ by (19). It follows from (21) that the gcm $\mathbf{C}_{\lambda k}$ are real-valued and satisfy

$$\mathbf{C}_{\lambda,-k} \;=\; \mathbf{C}'_{\lambda k}. \tag{25}$$

In particular, for the gacm $\mathbf{\Gamma}_{\lambda k}$ we have the relation

$$\mathbf{\Gamma}_{\lambda,-k} \;=\; \mathbf{\Gamma}'_{\lambda k}. \tag{26}$$

See [35] for the univariate case, i.e., $\gamma_{\lambda,-k} \;=\; \gamma_{\lambda k}$.

For $\lambda = k = 0$ the Szegö-Kolmogorov formula for the prediction error variance gives

$$\mathbf{\Omega} \;=\; \exp\left(\frac{1}{2\,\pi}\int_{-\pi}^{\pi} \log_e\, \mathbf{f}(\omega)\, d\omega\right) \;=\; \exp(\mathbf{C}_{00})$$

hence

$$\mathbf{C}_{00} \;=\; \log_e\, \mathbf{\Omega}. \tag{27}$$

For $\lambda \neq 0$ the gcm $\mathbf{C}_{\lambda k}$ are related to gacm $\mathbf{\Gamma}_{\lambda k}$ by the following relationships:

$$\mathbf{C}_{\lambda 0} \;=\; \frac{1}{\lambda}\,(\mathbf{\Gamma}_{\lambda 0} \,-\, \mathbf{I}_n) \qquad\qquad \mathbf{C}_{\lambda k} \;=\; \frac{1}{\lambda}\,\mathbf{\Gamma}_{\lambda k} \qquad (k \neq 0). \tag{28}$$

Remark 3. For $\lambda = 1$, $\mathbf{C}_{1k} = \mathbf{\Gamma}_{1k} = \mathbf{\Gamma}_k$ $(k > 0)$ from (2) and (28), i.e., the acf of $\mathbf{x}$. For $\lambda = -1$ and $k \neq 0$, $\mathbf{C}_{-1,k} = -\mathbf{\Gamma}_{-1,k}$, which is the inverse autocovariance of $\mathbf{x}$ (see Cleveland, 1972, for univariate time series). The matrix $\mathbf{C}_{\lambda 0}$ for $\lambda = -1, 0, 1$ is related to some statistical characteristics of $\mathbf{x}$. Namely, $(\mathbf{I}_n - \mathbf{C}_{-1,0})^{-1}$ is the interpolation error variance, $\exp(\mathbf{C}_{00}) = \mathbf{\Omega}$ is the prediction error variance, and $\mathbf{I}_n + \mathbf{C}_{1,0} = \mathbf{\Gamma}_{1,0} = \mathbf{\Gamma}_0$ is the unconditional autocovariance matrix of $\mathbf{x}$.

The coefficients of the Wold and autoregressive representations of $\mathbf{x}$ can be obtained using a spectral factorization that arises naturally after a reparameterization of the gcm $\mathbf{C}_{\lambda k}$. By the representation theorem of Fejér and Riesz, we can write

$$\mathbf{f}^\lambda(\omega) \;=\; \boldsymbol{\varphi}_\lambda(z)\,\mathbf{\Omega}_\lambda\,\boldsymbol{\varphi}'_\lambda(\overline{z}) \tag{29}$$

where

$$\boldsymbol{\varphi}_\lambda(z) \;=\; \sum_{k=0}^{K} \boldsymbol{\varphi}_{\lambda\,k}\, z^k \qquad\qquad\qquad \boldsymbol{\varphi}_{\lambda\,0} = \mathbf{I}_n.$$

From (29) when $\lambda \neq 0$ the gcm $\mathbf{C}_{\lambda k}$ are obtained as

$$\mathbf{C}_{\lambda 0} \;=\; \frac{1}{\lambda}\left[\sum_{k=0}^{K} \boldsymbol{\varphi}_{\lambda\, k}\, \boldsymbol{\Omega}_{\lambda}\, \boldsymbol{\varphi}'_{\lambda,k} \;-\; \mathbf{I}_n\right] \tag{30}$$

and

$$\mathbf{C}_{\lambda k} \;=\; \frac{1}{\lambda} \sum_{j=k}^{K} \boldsymbol{\varphi}_{\lambda\, j}\, \boldsymbol{\Omega}_{\lambda}\, \boldsymbol{\varphi}'_{\lambda,j-k} \qquad\qquad (k \neq 0). \tag{31}$$

Using arguments from [36], we can always assume that $\det \boldsymbol{\varphi}_{\lambda}(z) \neq 0$ for $|z| \leq 1$.

Extending the method proposed by [34] to power series with matrix coefficients, for $\lambda = 0$ the coefficients of the Wold representation can be obtained recursively as

$$\boldsymbol{\Psi}_j \;=\; \frac{1}{j} \sum_{r=1}^{j} r\, \mathbf{C}_{0r}\, \boldsymbol{\Psi}_{j-r} \qquad (j > 0) \qquad\qquad \boldsymbol{\Psi}_0 \;=\; \mathbf{I}_n.$$

For $\lambda \neq 0$, we have the recursive relations

$$\boldsymbol{\Psi}_j \;=\; \frac{1}{j} \sum_{r=1}^{j} \left(r\, \frac{\lambda+1}{\lambda} \;-\; j \right) \boldsymbol{\varphi}_{\lambda\, r}\, \boldsymbol{\Psi}_{j-r} \qquad (j > 0) \qquad\qquad \boldsymbol{\Psi}_0 \;=\; \mathbf{I}_n$$

by using [16]. The coefficients of the VAR(∞) representation of $\mathbf{x}$ are derived as follows. Set $\mathbf{x}_t = \boldsymbol{\Psi}(B)\, \boldsymbol{\epsilon}_t$ as in (3) and $\boldsymbol{\epsilon}_t = \boldsymbol{\Phi}(B)\, \mathbf{x}_t$ with $\boldsymbol{\Psi}(B) = \sum_{k \geq 0} \boldsymbol{\Psi}_k\, B^k$ ($\boldsymbol{\Psi}_0 = \mathbf{I}_n$) and $\boldsymbol{\Phi}(B) = \mathbf{I}_n \;-\; \sum_{k \geq 1} \boldsymbol{\Phi}_k\, B^k$. Following [17] (§10), the matrices $\boldsymbol{\Phi}_k$ can be evaluated by requiring

$$\left(\mathbf{I}_n \;-\; \boldsymbol{\Phi}_1\, B^1 \;-\; \boldsymbol{\Phi}_2\, B^2 \;-\; \cdots\right) \left(\mathbf{I}_n \;+\; \boldsymbol{\Psi}_1\, B^1 \;+\; \boldsymbol{\Psi}_2\, B^2 \;-\; \cdots\right) \;=\; \mathbf{I}_n.$$

Setting the coefficient of B^1 equal to zero produces $\boldsymbol{\Phi}_1 \;=\; \boldsymbol{\Psi}_1$. Similarly, setting the coefficient of B^2 equal to zero gives $\boldsymbol{\Phi}_2 \;=\; -\,\boldsymbol{\Phi}_1\, \boldsymbol{\Psi}_1 \;+\; \boldsymbol{\Psi}_2$, hence $\boldsymbol{\Phi}_2 \;=\; -\,\boldsymbol{\Psi}_1^2 \;+\; \boldsymbol{\Psi}_2$. In general, setting the coefficient of B^k equal to zero produces the recursive relation

$$\boldsymbol{\Phi}_k \;=\; -\sum_{j=1}^{k-1} \boldsymbol{\Phi}_j\, \boldsymbol{\Psi}_{k-j} \;+\; \boldsymbol{\Psi}_k$$

for all $k = 2, 3, \ldots$ with $\boldsymbol{\Phi}_k = \mathbf{0}$ for $k < 0$, $\boldsymbol{\Phi}_0 = \mathbf{I}_n$, and $\boldsymbol{\Phi}_1 = \boldsymbol{\Psi}_1$.

Thus all the relevant informations for prediction are available from the $K+1$ matrices of the generalised matrix cepstrum $\mathcal{C}_{\mathbf{x},K}$.

4. Whittle likelihood estimation and asymptotic properties

Let $\mathbf{x}$ be generated by (3) with Gaussian disturbances (having finite eight order moments), and $\sum_{j \geq 0} j\, ||\boldsymbol{\Psi}_j||^2 \;<\; \infty$. Assume $\mathbf{x}_1, \ldots, \mathbf{x}_T$ from the process $\mathbf{x}$

are available and consider the sample spectral density matrix (periodogram) defined by

$$\mathbf{I}_T(\omega) \;=\; \frac{1}{2\,\pi\,T}\left(\sum_{t=1}^{T}\mathbf{x}_t\,\exp(-\,\mathbf{i}\omega\,t)\right)\left(\sum_{t=1}^{T}\mathbf{x}_t'\,\exp(\,\mathbf{i}\omega\,t)\right) \tag{32}$$

for $\omega\in[-\pi,\pi]$. Set $\widetilde{\mathbf{x}}=(\mathbf{x}_1'\cdots\mathbf{x}_T')'\in\mathbb{R}^{nT}$. Then $\widetilde{\mathbf{x}}\sim N(\mathbf{0},\widetilde{\boldsymbol{\Gamma}})$, where $\widetilde{\boldsymbol{\Gamma}}$ denotes the $(nT)\times(nT)$ covariance matrix of the sample.

The log Gaussian likelihood for a mean-zero sample, apart from a constant, is

$$\mathcal{L}(\lambda,\boldsymbol{\theta}) \;=\; -\frac{1}{2}\,\log_e\det\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}) \;-\; \frac{1}{2}\,\widetilde{\mathbf{x}}'\,\widetilde{\boldsymbol{\Gamma}}^{-1}(\boldsymbol{\theta})\,\widetilde{\mathbf{x}} \tag{33}$$

where

$$\boldsymbol{\theta}=\boldsymbol{\theta}_\lambda=([\mathrm{vec}\,\mathbf{C}_{\lambda 0}]'\quad[\mathrm{vec}\,\mathbf{C}_{\lambda 1}]'\cdots[\mathrm{vec}\,\mathbf{C}_{\lambda K}]')'\in\mathbb{R}^R$$

where $R=n^2(K+1)$ and K is as in (23). Assuming conditions sufficient to guarantee efficiency of the maximum likelihood (ML) estimators (see later), the inverse of the Hessian can be used to approximate their asymptotic covariance matrix.

Because there is some computational cost associated with the inversion of $\widetilde{\boldsymbol{\Gamma}}$, an approximated version of (33), known as the Whittle likelihood, may be preferable for very large sample sizes. Therefore, for the mean-zero case we can consider the following Whittle likelihood

$$\mathcal{W}_T(\lambda,\boldsymbol{\theta}) \;=\; -\,\frac{1}{2\,\pi}\int_{-\pi}^{\pi}\log_e\,\det\,\mathbf{f}(\omega)\,d\omega \;-\; \frac{1}{2\,\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left[\mathbf{I}_T(\omega)\,\mathbf{f}(\omega)^{-1}\right]\,d\omega. \tag{34}$$

Observe that the parameter λ appears on the right-hand side of (34) writing $\mathbf{f}(\omega)$ as in (24). Furthermore, the elements of $\boldsymbol{\theta}=\boldsymbol{\theta}_\lambda$ arise from the expression of $\mathbf{g}_\lambda(\omega)$ given in (23), by using the vec operator. These observations will always be assumed in the sequel to simplify notations.

Notice that the integral on the left-hand side in (34) equals the trace of $\mathbf{C}_{00}=\mathbf{C}_0$ for $\lambda=0$ (see Appendix A), that is,

$$\frac{1}{2\,\pi}\int_{-\pi}^{\pi}\log_e\,\det\,\mathbf{f}(\omega)\,d\omega \;=\; \mathrm{tr}\,\mathbf{C}_0. \tag{35}$$

Approximating (34) by the Riemann sum over Fourier frequencies $\omega_j=(\pi\,j)/T$ for $j=-T,\ldots,T$, the Whittle likelihood can be written

$$\mathcal{W}_T(\lambda,\boldsymbol{\theta}) \;=\; -\,\frac{1}{2\,T}\sum_{j=-T}^{T}\left\{\log_e\det\mathbf{f}(\omega_j) \;+\; \mathrm{tr}\left[\mathbf{I}_T(\omega_j)\,\mathbf{f}(\omega_j)^{-1}\right]\right\} \tag{36}$$

which is to be maximized with respect to the transformation parameter λ and the vector $\boldsymbol{\theta}=\boldsymbol{\theta}_\lambda$ containing the elements of the gcm $\mathbf{C}_{\lambda k}$ for $k=0,\ldots,K$.

Set

$$\mathbf{z}(\omega) \;=\; (1/2 \qquad z \qquad z^2 \quad \cdots \quad z^K)'\in\mathbb{R}^{K+1}$$

and

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_\lambda \;=\; [\mathbf{C}'_{\lambda 0} \qquad \mathbf{C}'_{\lambda 1} \quad \cdots \quad \mathbf{C}'_{\lambda K}]' \in \mathbb{R}^{[n(K+1)]\times n}$$

hence $\boldsymbol{\theta} = \operatorname{vec}\boldsymbol{\Theta}$, and let $\overline{\mathbf{z}}(\omega)$ be the complex conjugate of $\mathbf{z}(\omega)$. Then (23) becomes

$$\mathbf{g}_\lambda(\omega) \;=\; (\mathbf{z}(\omega)' \otimes \mathbf{I}_n)\,\boldsymbol{\Theta} \,+\, \boldsymbol{\Theta}'\,(\overline{\mathbf{z}}(\omega) \otimes \mathbf{I}_n). \tag{37}$$

Thus (36) can be expressed as

$$\mathcal{W}_T(\lambda, \boldsymbol{\theta}) \;=\; -\,\frac{1}{2\,T} \sum_{j=-T}^{T} \mathcal{W}_{T,j}(\lambda, \boldsymbol{\theta})$$

where

$$\mathcal{W}_{T,j}(\lambda, \boldsymbol{\theta}) \;=\; \log_e \det[\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_\lambda(\omega_j)]^{\frac{1}{\lambda}} \,+\, \operatorname{tr}\Big(\mathbf{I}_T(\omega_j)\,[\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_\lambda(\omega_j)]^{-\frac{1}{\lambda}}\Big)$$

for $\lambda \neq 0$, and

$$\mathcal{W}_{T,j}(\lambda, \boldsymbol{\theta}) \;=\; \operatorname{tr}\mathbf{g}_\lambda(\omega_j) \,+\, \operatorname{tr}\big(\mathbf{I}_T(\omega_j)\,[\exp\mathbf{g}_\lambda(\omega_j)]^{-1}\big)$$

for $\lambda = 0$ (use Formula (4) from [11]. Here $\mathbf{g}_\lambda(\omega)$ is given by (37), and $\mathbf{I}_n + \lambda\,\mathbf{g}_\lambda(\omega_j)$ is assumed to be positive definite.

As suggested in [36], the profile likelihood of the model as λ varies can be used to select the spectral model for $\mathbf{x}$. Let $\mathcal{W}_{T,\max}(\lambda)$ denote the partially maximised (or, profile) Whittle likelihood, i.e., $\mathcal{W}_{T,\max}(\lambda) = \mathcal{W}_T(\lambda, \widetilde{\boldsymbol{\theta}}_\lambda)$, where $\widetilde{\boldsymbol{\theta}}_\lambda = \operatorname{argmax}_{\boldsymbol{\theta}} \mathcal{W}_T(\lambda, \boldsymbol{\theta})$. The ML estimate of λ is obtained as the value which maximises the profile Whittle likelihood.

The truncation parameter K can be chosen as the value minimizing an information criterion, such as AIC or BIC, given by

$$\mathrm{AIC}(K, \lambda) \;=\; -\,2\,\mathcal{W}_T(\lambda, \widetilde{\boldsymbol{\theta}}_\lambda) \,+\, 2\,K$$

and

$$\mathrm{BIC}(K, \lambda) \;=\; -\,2\,\mathcal{W}_T(\lambda, \widetilde{\boldsymbol{\theta}}_\lambda) \,+\, \log_e(T)\,K,$$

respectively.

We now prove the consistency and asymptotic normality of the Whittle ML estimator, denoted by $(\widetilde{\lambda}_T, \widetilde{\boldsymbol{\theta}}'_T)'$, of the parameter vector $(\lambda, \boldsymbol{\theta}')'$. In practice, the transformation parameter is estimated by maximizing the profile likelihood, as discussed above. As usual, denote $(\lambda_0, \boldsymbol{\theta}'_0)'$ the true parameter vector.

The asymptotic theory for the Whittle MLE is based on the following assumptions.

Assumption A1. The process $\mathbf{x} = (\mathbf{x}_t)$ is a stationary Gaussian process, generated according to (3), where $(\boldsymbol{\epsilon}_t)$ is a sequence of independent, identically distributed normal random variables with mean zero, positive definite covariance matrix $\boldsymbol{\Omega}$, and $\boldsymbol{\epsilon}_t$ has finite absolute eight order moments. Furthermore

$$\sum_{j\geq 0} j\,||\boldsymbol{\Psi}_j||^2 \;<\; \infty.$$

Assumption A2. The true parameter vector of the model is in the interior of the parameter space, and the $(n \times n)$ matrix $\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)$ is positive definite for every $\omega \in [-\pi, \pi]$, with determinant $\geq m$ for some $m > 0$.

Assumption A3. The process $\mathbf{x}$ is absolutely regular, that is, its cepstral matrix coefficients satisfy $\sum_{j=1}^{\infty} j\, ||\mathbf{C}_j||^2 < \infty$.

Theorem 1. *Assume that Assumptions A1–A3 are fulfilled. Then we have*

$$\text{plim}_{T\to\infty}\, \widetilde{\lambda}_T \,=\, \lambda_0 \qquad\qquad \text{plim}_{T\to\infty}\, \widetilde{\boldsymbol{\theta}}_T \,=\, \boldsymbol{\theta}_0$$

and

$$\sqrt{T}\begin{pmatrix}\widetilde{\lambda}_T \,-\, \lambda_0 \\ \\ \widetilde{\boldsymbol{\theta}}_T \,-\, \boldsymbol{\theta}_0\end{pmatrix} \underset{d}{\longrightarrow} \mathrm{N}(\mathbf{0}, \mathbf{V}(\lambda_0, \boldsymbol{\theta}_0))$$

with

$$[\mathbf{V}(\lambda, \boldsymbol{\theta})]^{-1} \,=\, \begin{pmatrix} V^{(\lambda,\lambda)} & V^{(\lambda,\boldsymbol{\theta})} \\ \\ V^{(\boldsymbol{\theta},\lambda)} & V^{(\boldsymbol{\theta},\boldsymbol{\theta})}\end{pmatrix} \in \mathbb{R}^{(1+R)\times(1+R)}$$

where

$$V^{(\lambda,\lambda)} \,=\, \frac{n}{2\,\pi}\int_{-\pi}^{\pi}\left\{\mathrm{tr}\left[\mathbf{f}(\omega)^{-\lambda}\,\mathbf{g}_\lambda(\omega)^2\,\mathbf{M}(\omega)\right]\right\}^2\, d\omega$$

$$V^{(\lambda,\boldsymbol{\theta})} \,=\, \frac{n}{2\,\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left[\mathbf{f}(\omega)^{-\lambda}\,\mathbf{g}_\lambda(\omega)^2\,\mathbf{M}(\omega)\right]\,[\mathrm{vec}\,\mathbf{f}(\omega)^{-\lambda}] \times\, [\mathbf{z}'(\omega)\otimes\mathbf{I}_{n^2} \,+\, \overline{\mathbf{z}}'(\omega)\otimes\mathbf{I}_{n^2}]\, d\omega$$

and

$$V^{(\boldsymbol{\theta},\boldsymbol{\theta})} \,=\, \frac{n}{2\,\pi}\int_{-\pi}^{\pi}[\mathbf{z}(\omega)\otimes\mathbf{I}_{n^2} \,+\, \overline{\mathbf{z}}(\omega)\otimes\mathbf{I}_{n^2}]\,[\mathrm{vec}\,\mathbf{f}(\omega)^{-\lambda}] \times\, [\mathrm{vec}\,\mathbf{f}(\omega)^{-\lambda}]'\,[\mathbf{z}'(\omega)\otimes\mathbf{I}_{n^2} \,+\, \overline{\mathbf{z}}'(\omega)\otimes\mathbf{I}_{n^2}]\, d\omega$$

with $V^{(\boldsymbol{\theta},\lambda)} = V^{(\lambda,\boldsymbol{\theta})'}$, $R = n^2(K+1)$ *and*

$$\mathbf{M}(\omega) = [\mathbf{g}_\lambda(\omega)]^{-2}\left\{\frac{1}{\lambda^2}\left(\lambda\,\mathbf{g}_\lambda(\omega) \,-\, [\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_\lambda(\omega)]\,\log_e\,[\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_\lambda(\omega)]\right)\right\}.$$

This theorem generalizes Theorem 2 by [36] to multivariate time series. The above matrix expressions give those in the quoted paper when $n = 1$. Furthermore, the above theorem provides the asymptotic theory for the special case of the vector exponential model of [20]. More precisely, for $\lambda_0 = 0$ (that is, the exponential case) A1 and A2 are not needed, and the finiteness of $\sum_{j=1}^{\infty} j\, ||\mathbf{C}_j||^2$ implies $\sum_{j=1}^{\infty} j\, ||\boldsymbol{\Gamma}_j||^2 < \infty$. Then we have

Theorem 2. *Under Assumption A3, the asymptotic matrix* $[\mathbf{V}(0, \boldsymbol{\theta})]^{-1}$ *for the vector exponential model (case* $\lambda_0 = 0$*) has block elements*

$$V^{(0,0)} \,=\, \frac{n}{8\,\pi}\int_{-\pi}^{\pi}\left\{\mathrm{tr}[\mathbf{g}_\lambda(\omega)^2]\right\}^2\, d\omega$$

$$V^{(0,\boldsymbol{\theta})} \;=\; -\,\frac{n}{4\,\pi}\int_{-\pi}^{\pi} \mathrm{tr}[\mathbf{g}_\lambda(\omega)^2]\,[\mathrm{vec}(\mathbf{I}_n)]'\,[\mathbf{z}'(\omega)\otimes\mathbf{I}_{n^2} \;+\; \overline{\mathbf{z}}'(\omega)\otimes\mathbf{I}_{n^2}]\,d\omega$$

and

$$V^{(\boldsymbol{\theta},\boldsymbol{\theta})} \;=\; \frac{n}{2\,\pi}\int_{-\pi}^{\pi} [\mathbf{z}(\omega)\otimes\mathbf{I}_{n^2} \;+\; \overline{\mathbf{z}}(\omega)\otimes\mathbf{I}_{n^2}]\,[\mathrm{vec}\,\mathbf{I}_n] \\ \times\;\; [\mathrm{vec}\,\mathbf{I}_n]'\,[\mathbf{z}'(\omega)\otimes\mathbf{I}_{n^2} \;+\; \overline{\mathbf{z}}'(\omega)\otimes\mathbf{I}_{n^2}]\,d\omega.$$

The matrix formulas in the statement of Theorem 2 are obtained by taking the limit for $\lambda \to 0$ of $[\mathbf{V}(\lambda,\boldsymbol{\theta})]^{-1}$, as given in Theorem 1.

5. Simulated and empirical examples

5.1. *Simulated* VAR(4) *process*

This example estimates the spectrum of a bivariate VAR(4) process defined by

$$\boldsymbol{\Phi}(L)\,\mathbf{x}_t \;=\; \boldsymbol{\epsilon}_t, \qquad\qquad \boldsymbol{\epsilon}_t \sim \mathrm{NID}(\mathbf{0},\boldsymbol{\Omega})$$

where

$$\boldsymbol{\Phi}(L) \;=\; \mathbf{I}_2 \;-\; \boldsymbol{\Phi}_1\,L \;-\; \boldsymbol{\Phi}_2\,L^2 \;-\; \boldsymbol{\Phi}_3\,L^3 \;-\; \boldsymbol{\Phi}_4\,L^4$$

with

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} -0.1731 & -0.0212 \\ 0.1549 & -0.1405 \end{pmatrix} \qquad\qquad \boldsymbol{\Phi}_2 \;=\; \begin{pmatrix} 0.0412 & -0.0236 \\ 0.0989 & 0.0519 \end{pmatrix}$$

$$\boldsymbol{\Phi}_3 = \begin{pmatrix} 0.0051 & 0.0282 \\ 0.0351 & 0.0248 \end{pmatrix} \qquad\qquad \boldsymbol{\Phi}_4 \;=\; \begin{pmatrix} 0.0205 & 0.0111 \\ 0.0225 & 0.0003 \end{pmatrix}$$

and $\boldsymbol{\Omega} \;=\; \mathbf{I}_2$. Let $\boldsymbol{\Phi}(z)$ be the AR matrix polynomial of the model. Then $\det\boldsymbol{\Phi}(z)$ has two real roots -2.3015 and 2.1488, and three pairs of complex conjugate roots with modulus 2.7016, 3.2562 and 3.6144 and phases 1.76, 2.68 and 0.89, respectively. The process is second-order stationary as $\det\boldsymbol{\Phi}(z) \neq 0$ for every $|z| \leq 1$. Let $\mathbf{f}(\omega) \;=\; (f_{ij}(\omega)) \in \mathbb{R}^2$ be the spectral matrix of the process. Rescaling the squared argument of the cross-spectrum $f_{12}(\omega)$ yields the squared coherence $\rho^2(\omega)$ defined as

$$\rho^2(\omega) = \frac{|f_{12}(\omega)|^2}{f_{11}(\omega)\,f_{22}(\omega)}.$$

We conduct a Monte Carlo experiment in which 1,000 replications of length $T = 100$ are generated according to the bivariate VAR(4) Gaussian process above. A typical realization is plotted in Figure 1, with $T \;=\; 100$ observations. For each replication, we estimate a sequence of GMCM(λ, K) models, with λ taking values on a grid from -2.1 and -0.6 with step 0.05, and for K ranging from 0 to 7. This range of values covers the subset of interest of the parameter space

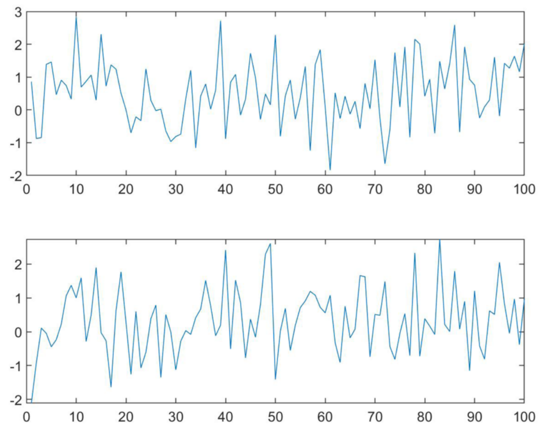

FIG 1. *A typical realization of the simulated bivariate VAR(4) described in Section 5.1.*

and that in which the maximization of the Whittle likelihood by a numerical algorithm works well. Model selection is carried out using BIC criterion. The main results are in Table B (see Appendix B), which reports the means of the estimated cepstral matrix parameters, the corresponding standard deviations, and the true parameter values. The standard errors are computed by using the asymptotic matrix expressions given in the statement of Theorem 1. Estimates are rather close to the true parameter values and they tend to be closer as the number of replications increases. It can be observed that the Whittle estimation method has small bias and reasonable standard errors. A typical profile of the Whittle likelihood of the GMCM(λ, K) models as a function of λ (on the x-axis) is depicted in Figure 2. This shows that the optimal value of the power transformation parameter λ is equal to -1. Given $\lambda = -1$, model selection by BIC criterion for the order of the GMCM(λ, K) selects the true model parameter $K = 4$ the 81% of the replications. Finally, the estimated squared coherence is reported in Figure 3 together with the centered periodogram in points. The comparison shows that the selected model exhibits a squared coherence that approximate the periodogram quite well.

5.2. Empirical example on hedge ratios

[24] use a bivariate 2-state Markov switching generalised autoregressive conditional heteroscedastic model (in short, MS GARCH) to estimate the hedge ratios through the second moments of futures and spot prices in the corn commodity market. It is shown in the quoted paper that the MS-GARCH specification is more capable than standard VARMA models to capture the empirical features of the considered real data. Let $r_{c,t}$ and $r_{f,t}$ denote the returns on the spot and futures at time t, respectively. Then the bivariate process is $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}}$, where $\mathbf{x}_t = (r_{c,t} \quad r_{f,t})'$. The data series (traded on the Chicago Board of Trade)

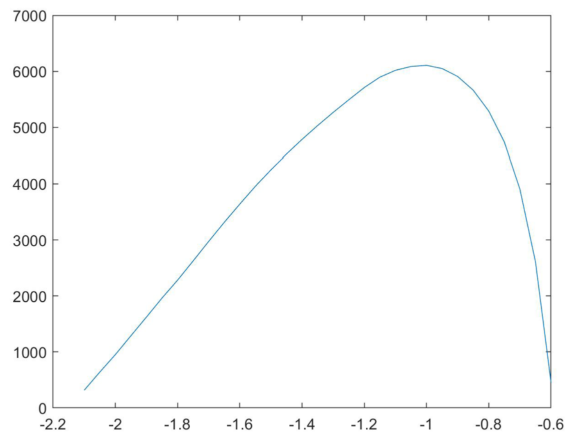

FIG 2. *A typical profile of the Whittle likelihood estimation of GMCM(λ, K) model in the simulated experiment described in Section 5.1.*

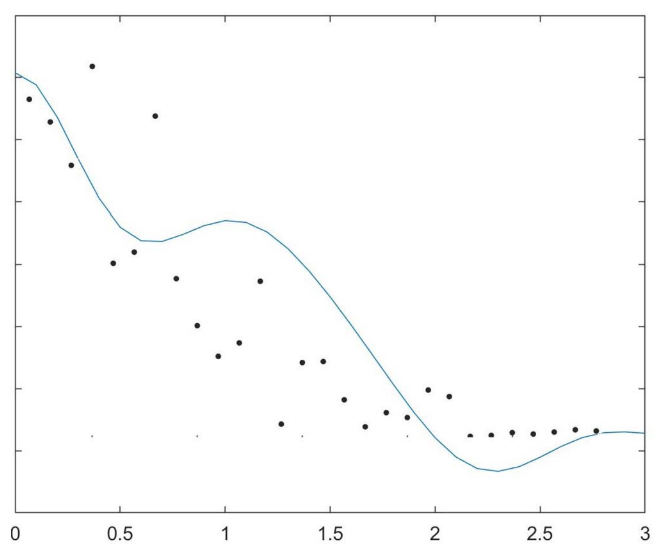

FIG 3. *Estimated squared coherence of GMCM(−1, 4) model (line) and periodogram (points). See Section 5.1.*

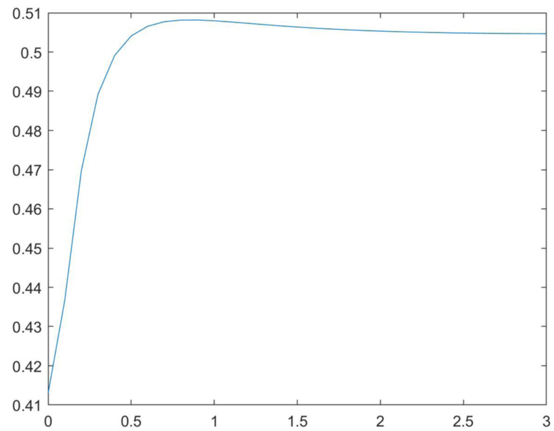

FIG 4. *Hedge ratio. Squared coherence $\rho^2(\omega)$ of the process described in Section 5.2.*

is taken from the period 2 January 1991 to 31 December 2003. The spectral density functions of the squares $r^2_{c,t}$ and $r^2_{f,t}$, and the cross product $r_{c,t}\, r_{f,t}$ have been derived in [9] by starting from the MS GARCH specification. By using these functions, the squared coherence $\rho^2(\omega)$ of the process is plotted in Figure 4. We now investigate what GMCM(λ, K) representation provides the best fit to the curve in Figure 4. Estimating the GMCM(λ, K) on a grid of values for λ in the range $[-12.00; 2.00]$ with step 0.45, and for K ranging from 0 to 7, and computing the BIC criterion, leads to selecting $\widetilde{\lambda} = -5.56$ and $\widetilde{K} = 5$. Figure 5 displays the profile Whittle likelihood of GMCM$(\lambda, \widetilde{K})$ models, as a function of λ in the grid. It shows that the optimal value of the power transformation parameter is $\widetilde{\lambda} = -5.56$. Figure 6 plots the squared coherences fitted by the GMCM$(\lambda, \widetilde{K})$ models with $\widetilde{K} = 5$ and λ set equal to 1, 0, -1, and -5.56, respectively. The case $\lambda = 1$ corresponds to fitting a bivariate MA(5) model to the series; the case $\lambda = 0$ corresponds to fitting the bivariate exponential model of order $\widetilde{K} = 5$ from [20]; and $\lambda = -1$ corresponds to fitting a bivariate AR(5). For none of these cases the estimated squared coherence has the form of that plotted in Figure 4. On the contrary, the squared coherence fitted by the maximum likelihood, i.e., $\widetilde{\lambda} = -5.56$ and $\widetilde{K} = 5$, has the same profile shown in Figure 4.

6. Conclusion

We have proposed a general class of frequency domain models for multivariate stationary stochastic processes, called generalised matrix cepstral models (GMCM). Such models, expressed in terms of cepstrum theory, extend the generalised cepstral model of [36] to the multivariate setting, and nest the vector exponential model of [19, 20] as a special case. Any multivariate short memory time series application can be faced by using the GMCM framework, providing a

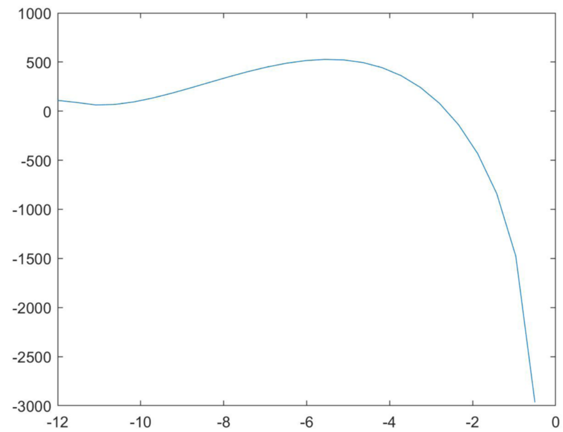

FIG 5. *Hedge ratio. Profile of the Whittle likelihood as a function of* λ *for* $GMCM(\lambda, \widetilde{K})$ *models with* $\widetilde{K} = 5$.

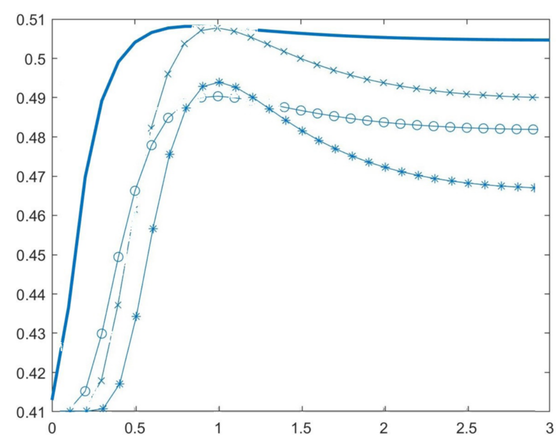

FIG 6. *Hedge ratio. Comparison of the squared coherences arising from different values of* λs *in* $GMCM(\lambda, \widetilde{K})$ *models with* $\widetilde{K} = 5$. $\lambda = 1$ *(starred line),* $\lambda = 0$ *(dotted line),* $\lambda = -1$ *(crossed line),* $\lambda = -5.56$ *(solid line).*

large class of models for vector time series. Then we derive the statistical properties of the model, discuss the Whittle maximum likelihood estimation, and prove the consistency and asymptotic properties of the parameter estimators. Other approaches to estimation can be immediately discussed, such as Bayesian estimation along the treatment shown in [20]. A further development of the research may be to allow the elements of cepstral matrices to vary over time. This permits to model time-varying spectra of locally stationary vector time series. Finally, alternative cepstral models can be investigated along the lines of the proposed model.

Appendix A

Derivation of (20). We have

$$\begin{aligned}
\frac{1}{\lambda}\,[\mathbf{f}^{\lambda} \,-\, \mathbf{I}_n] \,&=\, \frac{1}{\lambda}\,[\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^{*} \,-\, \mathbf{I}_n] \\
&= \frac{1}{\lambda}\,[\mathbf{P}\,\mathbf{A}^{\lambda}\,\mathbf{P}^{*} \,-\, \mathbf{P}\,\mathbf{P}^{*}] \\
&= \mathbf{P}\,\left[\frac{1}{\lambda}\,(\mathbf{A}^{\lambda} \,-\, \mathbf{I}_n)\right]\,\mathbf{P}^{*} \\
&= \mathbf{P}\ \mathrm{diag}\left((a_{11}^{\lambda}-1)/\lambda, \ldots, (a_{nn}^{\lambda}-1)/\lambda\right)\,\mathbf{P}^{*}
\end{aligned}$$

where $\mathbf{A} = \mathrm{diag}(a_{11}, \ldots, a_{nn})$. Then

$$\begin{aligned}
\lim_{\lambda\to 0}\frac{1}{\lambda}\,[\mathbf{f}^{\lambda} \,-\, \mathbf{I}_n] \,&=\, \mathbf{P}\ \mathrm{diag}\left(\lim_{\lambda\to 0}\frac{a_{11}^{\lambda}-1}{\lambda}, \ldots, \lim_{\lambda\to 0}\frac{a_{nn}^{\lambda}-1}{\lambda}\right)\,\mathbf{P}^{*} \\
&= \mathbf{P}\ \mathrm{diag}\left(\log_e a_{11}, \ldots, \log_e a_{nn}\right)\,\mathbf{P}^{*} \\
&= \mathbf{P}\ \log_e(\mathbf{A})\,\mathbf{P}^{*} \,=\, \log_e \mathbf{f}.
\end{aligned}$$

Derivation of (24). Given a factorization $\mathbf{g}_{\lambda} = \mathbf{Q}_{\lambda}\,\mathbf{D}_{\lambda}\,\mathbf{Q}_{\lambda}^{*}$ with $\mathbf{Q}_{\lambda}$ unitary matrix and $\mathbf{D}_{\lambda}$ diagonal matrix, we get

$$\begin{aligned}
\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_{\lambda} &= \mathbf{I}_n \,+\, \mathbf{Q}_{\lambda}\,\lambda\,\mathbf{D}_{\lambda}\,\mathbf{Q}_{\lambda}^{*} \\
&= \mathbf{Q}_{\lambda}\,\mathbf{Q}_{\lambda}^{*} \,+\, \mathbf{Q}_{\lambda}\,\lambda\,\mathbf{D}_{\lambda}\,\mathbf{Q}_{\lambda}^{*} \\
&= \mathbf{Q}_{\lambda}\,(\mathbf{I}_n \,+\, \lambda\,\mathbf{D}_{\lambda})\,\mathbf{Q}_{\lambda}^{*} \\
&= \mathbf{Q}_{\lambda}\ \mathrm{diag}(1 \,+\, \lambda\,d_{11,\lambda}, \ldots, 1 \,+\, \lambda\,d_{nn,\lambda})\,\mathbf{Q}_{\lambda}^{*}
\end{aligned}$$

where $\mathbf{D}_{\lambda} = \mathrm{diag}(d_{11,\lambda}, \ldots, d_{nn,\lambda})$. Then

$$[\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_{\lambda}]^{\frac{1}{\lambda}} \,=\, \mathbf{Q}_{\lambda}\ \mathrm{diag}\left((1 \,+\, \lambda\,d_{11,\lambda})^{\frac{1}{\lambda}}, \ldots, (1 \,+\, \lambda\,d_{nn,\lambda})^{\frac{1}{\lambda}}\right)\,\mathbf{Q}_{\lambda}^{*}$$

hence

$$\begin{aligned}
\lim_{\lambda\to 0}[\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_{\lambda}]^{\frac{1}{\lambda}} \,&=\, \mathbf{Q}_{\lambda}\ \mathrm{diag}\left(\lim_{\lambda\to 0}(1 \,+\, \lambda\,d_{11,\lambda})^{\frac{1}{\lambda}}, \ldots, \lim_{\lambda\to 0}(1 \,+\, \lambda\,d_{nn,\lambda})^{\frac{1}{\lambda}}\right)\,\mathbf{Q}_{\lambda}^{*} \\
&= \mathbf{Q}_{\lambda}\ \mathrm{diag}\left(\exp(d_{11,\lambda}), \ldots, \exp(d_{nn,\lambda})\right)\,\mathbf{Q}_{\lambda}^{*} \\
&= \mathbf{Q}_{\lambda}\ \exp(\mathbf{D}_{\lambda})\,\mathbf{Q}_{\lambda}^{*} \\
&= \exp(\mathbf{Q}_{\lambda}\,\mathbf{D}_{\lambda}\,\mathbf{Q}_{\lambda}^{*}) = \exp \mathbf{g}_{\lambda}.
\end{aligned}$$

Derivation of (28). For $\lambda \neq 0$ using (15) and (19) we get

$$\begin{aligned}\boldsymbol{\Gamma}_{\lambda k} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{f}^{\lambda}(\omega)\, z^{-k}\, d\omega \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} (\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_{\lambda}(\omega))\; z^{-k}\, d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{I}_n\, z^{-k}\, d\omega \,+\, \lambda\,\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{g}_{\lambda}(\omega)\, z^{-k}\, d\omega \\ &= \delta_0^k\,\mathbf{I}_n \,+\, \lambda\,\mathbf{C}_{\lambda k}\end{aligned}$$

by (22). Here δ_j^i denotes the Kronecker symbol, i.e., $\delta_i^i = 1$ and $\delta_j^i = 0$ for $i \neq j$. In fact, the integral on the left-hand side vanishes for $k \neq 0$ while equals $\mathbf{I}_n$ for $k = 0$.

Derivation of (35). Using (18) and Formula (4) from [11], we get ($\lambda = 0$)

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log_e\, \det \mathbf{f}(\omega)\, d\omega &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log_e\, \det \exp\left(\sum_{k=-\infty}^{\infty} \mathbf{C}_k\, z^k\right) d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left(\sum_{k=-\infty}^{\infty} \mathbf{C}_k\, z^k\right) d\omega \\ &= \frac{1}{2\pi}\,\mathrm{tr}\left(\sum_{k=-\infty}^{\infty} \mathbf{C}_k \int_{-\pi}^{\pi} z^k\, d\omega\right) \;=\; \mathrm{tr}\,\mathbf{C}_0\end{aligned}$$

as the last integral vanishes for $k \neq 0$ and equals 2π for $k = 0$.

Derivation of (37). From (23) and (25) we get

$$\begin{aligned}\mathbf{g}_{\lambda}(\omega) = \sum_{k=-K}^{K} \mathbf{C}_{\lambda k}\, z^k \;&=\; \mathbf{C}_{\lambda 0} \,+\, \sum_{k=-K}^{-1} \mathbf{C}_{\lambda k}\, z^k \,+\, \sum_{k=1}^{K} \mathbf{C}_{\lambda k}\, z^k \\ &= \mathbf{C}_{\lambda 0} \,+\, \sum_{k=1}^{K} \mathbf{C}_{\lambda,-k}\, \overline{z}^k \,+\, \sum_{k=1}^{K} \mathbf{C}_{\lambda k}\, z^k \\ &= \mathbf{C}_{\lambda 0} \,+\, \sum_{k=1}^{K} \mathbf{C}'_{\lambda k}\, \overline{z}^k \,+\, \sum_{k=1}^{K} \mathbf{C}_{\lambda k}\, z^k \\ &= (\mathbf{z}(\omega)' \otimes \mathbf{I}_n)\,\boldsymbol{\Theta} \,+\, \boldsymbol{\Theta}'\,(\overline{\mathbf{z}}(\omega) \otimes \mathbf{I}_n).\end{aligned}$$

Proof of the Theorem 1. Under assumptions A1–A3, Theorem II.2.1 (consistency) and Theorem II.2.2 from [13] hold, as proved below. Condition A2 states that $\mathbf{I}_n \,+\, \lambda\,\mathbf{g}_{\lambda}(\omega)$ is pd, hence it is a proper spectral density. To guarantee the positivity of the spectral density, one can use a reparameterization of the model based on a set of generalised inverse partial autocorrelations as described in [36]. Furthermore, under A1 the gacm $\boldsymbol{\Gamma}_{\lambda k}$ form a pd sequence. In addition, by (37) $\mathbf{g}_{\lambda}(\omega)$ is a smooth and Hermitian matrix function of $\omega \in [-\pi, \pi]$. Note that the GMCM is also identified, i.e., $(\lambda_1, \boldsymbol{\theta}_1')' \neq (\lambda_2, \boldsymbol{\theta}_2')'$ implies that $\mathbf{f}_1(\omega) \neq \mathbf{f}_2(\omega)$ for almost all frequencies ω and parameter pairs $(\lambda_i, \boldsymbol{\theta}_i')'$ in the interior of parameter space, $i = 1, 2$. Assumption A3 implies that $\sum_{j=-\infty}^{\infty} j\,||\boldsymbol{\Gamma}_k||^2 < \infty$. This fact, together with $\mathbf{f}(\omega)$ pd and continuous, guarantees that the principal part of the Gaussian log-likelihood can be approximated by the Whittle likelihood. See Corollary I.3.1 from [13].

We now compute the derivatives of the inverse matrix of the sdm $\mathbf{f}(\omega)$. The derivative of $\mathbf{f}(\omega)^{-1}$ with respect to λ is an $(n \times n)$ matrix. The derivative of $\mathbf{f}(\omega)^{-1}$ with respect to $\boldsymbol{\Theta}$ is an $n \times [n(K+1)]$ matrix. For computational convenience we use $\boldsymbol{\Theta}$ instead of $\boldsymbol{\theta} = \operatorname{vec} \boldsymbol{\Theta}$ taking in mind the expression of $\mathbf{g}_\lambda(\omega)$ in (37). For $\lambda \neq 0$, we have

$$\begin{aligned}
\frac{\partial\, \mathbf{f}(\omega)^{-1}}{\partial\, \lambda} &= \frac{\partial}{\partial\, \lambda}\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-\frac{1}{\lambda}} \\
&= \frac{\partial}{\partial\, \lambda} \exp \log_e\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-\frac{1}{\lambda}} \\
&= \exp \log_e\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-\frac{1}{\lambda}}\, \frac{\partial}{\partial\, \lambda} \log_e\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-\frac{1}{\lambda}} \\
&= \mathbf{f}(\omega)^{-1}\, \frac{\partial}{\partial\, \lambda} \left\{ -\frac{1}{\lambda} \log_e\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)] \right\} \\
&= \mathbf{f}(\omega)^{-1} \left\{ \frac{1}{\lambda^2} \log_e\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)] - \frac{1}{\lambda}\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-1}\, \mathbf{g}_\lambda(\omega) \right\} \\
&= -\, \mathbf{f}(\omega)^{-1}\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-1}\, [\mathbf{g}_\lambda(\omega)]^2\, \mathbf{M}(\omega) \in \mathbb{R}^{n \times n}
\end{aligned}$$

where $\mathbf{M}(\omega)$ is as in the statement of the theorem.

Note that

$$\lim_{\lambda \to 0} \mathbf{M}(\omega) = -\frac{1}{2}\, \mathbf{I}_n$$

hence

$$\lim_{\lambda \to 0} \frac{\partial\, \mathbf{f}(\omega)^{-1}}{\partial \lambda} = \frac{1}{2}\, \mathbf{f}(\omega)^{-1}\, [\mathbf{g}_\lambda(\omega)]^2$$

where

$$\mathbf{f}(\omega) = \exp\, \mathbf{g}_\lambda(\omega).$$

The derivative of $\mathbf{f}(\omega)^{-1}$ with respect to $\boldsymbol{\Theta}$ is the $n \times [n(K+1)]$ matrix function given by

$$\begin{aligned}
\frac{\partial\, \mathbf{f}(\omega)^{-1}}{\partial\, \boldsymbol{\Theta}} &= -\frac{1}{\lambda}\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-\frac{1}{\lambda} - 1}\, \lambda\, \frac{\partial\, \mathbf{g}_\lambda(\omega)}{\partial\, \boldsymbol{\Theta}} \\
&= -\, \mathbf{f}(\omega)^{-1}\, [\mathbf{I}_n + \lambda\, \mathbf{g}_\lambda(\omega)]^{-1}\, [\mathbf{z}'(\omega) \otimes \mathbf{I}_n + \overline{\mathbf{z}}'(\omega) \otimes \mathbf{I}_n].
\end{aligned}$$

In particular, we get

$$\lim_{\lambda \to 0} \frac{\partial\, \mathbf{f}(\omega)^{-1}}{\partial\, \boldsymbol{\Theta}} = -\, \mathbf{f}(\omega)^{-1}\, [\mathbf{z}'(\omega) \otimes \mathbf{I}_n + \overline{\mathbf{z}}'(\omega) \otimes \mathbf{I}_n]$$

where

$$\mathbf{f}(\omega) = \exp\, \mathbf{g}_\lambda(\omega).$$

Hence, the derivatives of $\mathbf{f}(\omega)^{-1}$ exist and are continuous in $(\lambda, \boldsymbol{\theta})$. By Theorem II.2.1 in [13], it follows that $\text{plim}_{T\to\infty}\, \widetilde{\lambda}_T = \lambda_0$ and $\text{plim}_{T\to\infty}\, \widetilde{\boldsymbol{\theta}}_T = \boldsymbol{\theta}_0$.

Furthermore, the spectral density matrix $\mathbf{f}(\omega)$ is a twice-differentiable matrix function of λ and $\boldsymbol{\Theta}$, and the second derivatives are

$$\frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \lambda^2} \;=\; \mathbf{f}(\omega)\, \Big\{[\mathbf{M}_1(\omega)]^2 \,+\, \mathbf{M}_2(\omega)\Big\} \in \mathbb{R}^{n\times n}$$

$$\frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \lambda\, \partial\boldsymbol{\Theta}} \;=\; \mathbf{f}(\omega)\, \mathbf{N}(\omega)\, [\mathbf{z}'(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}'(\omega)\otimes \mathbf{I}_n] \in \mathbb{R}^{n\times[n(K+1)]}$$

and

$$\frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \boldsymbol{\Theta}'\, \partial\boldsymbol{\Theta}} \;=\; (1-\lambda)\, \mathbf{f}(\omega)\, \mathbf{L}(\omega) \;\in \mathbb{R}^{n\times n}$$

where

$$\mathbf{M}_1(\omega) \;=\; -\,\frac{1}{\lambda^2}\, \log_e\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)] \,+\, \frac{1}{\lambda}\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)]^{-1}\, \mathbf{g}_\lambda(\omega)$$

$$\begin{aligned}\mathbf{M}_2(\omega) \;=\; &\frac{2}{\lambda^3}\, \log_e\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)] \,-\, \frac{2}{\lambda^2}\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)]^{-1}\, \mathbf{g}_\lambda(\omega)\\ &-\, \frac{1}{\lambda}\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)]^{-2}\, [\mathbf{g}_\lambda(\omega)]^2\end{aligned}$$

$$\mathbf{N}(\omega) \;=\; \left(\frac{1\,-\,\lambda}{\lambda}\right)\, [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)]^{-2}\, \mathbf{g}_\lambda(\omega)$$

and

$$\mathbf{L}(\omega) \;=\; [\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)]^{-2}\, [\mathbf{z}'(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}'(\omega)\otimes \mathbf{I}_n]\, [\mathbf{z}(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}(\omega)\otimes \mathbf{I}_n].$$

Under the stated assumption, i.e., $\mathbf{I}_n \,+\, \lambda\, \mathbf{g}_\lambda(\omega)$ is positive definite, the second partial derivatives are continuous in ω. Furthermore, we have

$$\lim_{\lambda\to 0}\, \frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \lambda^2} \;=\; \mathbf{f}(\omega)\, [\mathbf{g}_\lambda(\omega)]^3\, \left[\frac{1}{4}\, \mathbf{g}_\lambda(\omega) \,+\, \frac{2}{3}\, \mathbf{I}_n\right]$$

$$\lim_{\lambda\to 0}\, \frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \lambda\, \partial\boldsymbol{\Theta}} \;=\; -\, \mathbf{f}(\omega)\, \mathbf{g}_\lambda(\omega)\, \left[\mathbf{I}_n \,+\, 2\, \mathbf{g}_\lambda(\omega)\right]\, [\mathbf{z}'(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}'(\omega)\otimes \mathbf{I}_n]$$

and

$$\lim_{\lambda\to 0}\, \frac{\partial^2\, \mathbf{f}(\omega)}{\partial\, \boldsymbol{\Theta}'\, \partial\boldsymbol{\Theta}} \;=\; \mathbf{f}(\omega)\, [\mathbf{z}'(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}'(\omega)\otimes \mathbf{I}_n]\, [\mathbf{z}(\omega)\otimes \mathbf{I}_n \,+\, \overline{\mathbf{z}}(\omega)\otimes \mathbf{I}_n]$$

where

$$\mathbf{f}(\omega) \;=\; \exp\, \mathbf{g}_\lambda(\omega).$$

The above relations follow from

$$\lim_{\lambda\to 0}\, \mathbf{M}_1(\omega) \;=\; -\frac{1}{2}\, [\mathbf{g}_\lambda(\omega)]^2 \qquad\qquad \lim_{\lambda\to 0}\, \mathbf{M}_2(\omega) \;=\; \frac{2}{3}\, [\mathbf{g}_\lambda(\omega)]^3$$

$$\lim_{\lambda \to 0} \mathbf{N}(\omega) \;=\; -\,\mathbf{g}_\lambda(\omega) \;-\; 2\,[\mathbf{g}_\lambda(\omega)]^2$$

and

$$\lim_{\lambda \to 0} \mathbf{L}(\omega) \;=\; [\mathbf{z}'(\omega) \otimes \mathbf{I}_n \;+\; \overline{\mathbf{z}}'(\omega) \otimes \mathbf{I}_n]\,[\mathbf{z}(\omega) \otimes \mathbf{I}_n \;+\; \overline{\mathbf{z}}(\omega) \otimes \mathbf{I}_n].$$

Set $\boldsymbol{\delta} = (\lambda \quad \boldsymbol{\theta}')' \in \mathbb{R}^{R+1}$, where $R = n^2(K+1)$. The score associated with the Whittle likelihood in (36) is given by

$$\mathcal{S}(\boldsymbol{\delta}) \;=\; \frac{1}{2T} \sum_{j=-T}^{T} \operatorname{tr}\left[\mathbf{I}_T(\omega_j)\,\mathbf{f}(\omega_j)^{-1} \;-\; \mathbf{I}_n\right] \,\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'} \in \mathbb{R}^{1\times(R+1)}.$$

In fact, we have

$$\begin{aligned}
\frac{\partial \log_e \det \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'} &= -\,\frac{\partial \log_e \det \mathbf{f}(\omega)^{-1}}{\partial \boldsymbol{\delta}'} \\
&= -\,\frac{\partial \log_e \det \mathbf{f}(\omega)^{-1}}{\partial\,[\operatorname{vec}\mathbf{f}(\omega)^{-1}]'}\;\frac{\partial \operatorname{vec}\mathbf{f}(\omega)^{-1}}{\partial \boldsymbol{\delta}'} \\
&= [\operatorname{vec}\mathbf{f}(\omega)]'\,[\operatorname{vec}\mathbf{f}(\omega)^{-1}]\,(\operatorname{vec}\mathbf{I}_n)'\;\frac{\partial \operatorname{vec}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'} \\
&= \operatorname{tr}\left[\mathbf{f}(\omega)\,\mathbf{f}(\omega)^{-1}\right]\;\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \operatorname{tr}\left[\mathbf{I}_T(\omega)\,\mathbf{f}(\omega)^{-1}\right]}{\partial \boldsymbol{\delta}'} &= \frac{\partial \operatorname{tr}\left[\mathbf{I}_T(\omega)\,\mathbf{f}(\omega)^{-1}\right]}{\partial\,[\operatorname{vec}\mathbf{f}(\omega)^{-1}]'}\;\frac{\partial \operatorname{vec}\mathbf{f}(\omega)^{-1}}{\partial \boldsymbol{\delta}'} \\
&= [\operatorname{vec}\mathbf{I}_T(\omega)]'\;\frac{\partial \operatorname{vec}\mathbf{f}(\omega)^{-1}}{\partial \boldsymbol{\delta}'} \\
&= -\,[\operatorname{vec}\mathbf{I}_T(\omega)]'\,[\operatorname{vec}\mathbf{f}(\omega)^{-1}]\,(\operatorname{vec}\mathbf{I}_n)'\;\frac{\partial \operatorname{vec}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'} \\
&= -\,\operatorname{tr}\left[\mathbf{I}_T(\omega)\,\mathbf{f}(\omega)^{-1}\right]\;\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'}.
\end{aligned}$$

The Hessian matrix associated with the Whittle likelihood in (36) is the $(R+1)\times(R+1)$ matrix function given by

$$\begin{aligned}
\mathcal{H}(\boldsymbol{\delta}) \;=\; &-\frac{1}{2T} \sum_{j=-T}^{T} \operatorname{tr}\left[\mathbf{I}_T(\omega_j)\,\mathbf{f}(\omega_j)^{-1} \;-\; \mathbf{I}_n\right]\,\frac{\partial^2 \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}\,\partial \boldsymbol{\delta}'} \\
&-\frac{1}{2T} \sum_{j=-T}^{T} \operatorname{tr}\left[\mathbf{I}_T(\omega_j)\,\mathbf{f}(\omega_j)^{-1}\right]\,\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}}\;\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'}.
\end{aligned}$$

Hence, the information matrix $\mathcal{I}(\boldsymbol{\delta}) \;=\; E[\,-\,\mathcal{H}(\boldsymbol{\delta})]$ is such that

$$\mathcal{I}(\boldsymbol{\delta}) \;=\; \frac{n}{2\,\pi} \int_{-\pi}^{\pi} \frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}}\;\frac{\partial \operatorname{tr}\log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'}\,d\omega$$

by using [6] (Proposition 11.7.2). Furthermore, we have

$$\frac{\partial \operatorname{tr} \log_e \mathbf{f}(\omega)}{\partial \lambda} = \operatorname{tr}\left\{[\mathbf{I}_n + \lambda \mathbf{g}_\lambda(\omega)]^{-1} [\mathbf{g}_\lambda(\omega)]^2 \mathbf{M}(\omega)\right\} \in \mathbb{R}$$

and

$$\frac{\partial \operatorname{tr} \log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\theta}'} = \left\{\operatorname{vec}[\mathbf{I}_n + \lambda \mathbf{g}_\lambda(\omega)]^{-1}\right\}' [\mathbf{z}'(\omega) \otimes \mathbf{I}_{n^2} + \overline{\mathbf{z}}'(\omega) \otimes \mathbf{I}_{n^2}] \in \mathbb{R}^{1\times R}.$$

Thus $\widetilde{\lambda}$ and $\widetilde{\boldsymbol{\theta}}$ are asymptotically normal as stated in the theorem. Moreover, the asymptotic covariance matrix $\mathbf{V}(\lambda_0, \boldsymbol{\theta}_0)$ is given by

$$\mathbf{V}(\lambda, \boldsymbol{\theta}) = \left[\frac{n}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \operatorname{tr} \log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}} \frac{\partial \operatorname{tr} \log_e \mathbf{f}(\omega)}{\partial \boldsymbol{\delta}'} d\omega\right]^{-1}.$$

This completes the proof of the theorem.

Appendix B

TABLE B

Summary statistics of the simulated experiment of the Whittle likelihood estimation of the parameters of the GMCM(λ, K) model, based on 1,000 replications from the bivariate VAR(4) model described in Section 5.

Parameters	Mean	Sd	True
$C_{-1,0}(1,1)$	- 0.0802	0.1148	-0.0677
$C_{-1,0}(1,2)$	0.0341	0.2424	0.0127
$C_{-1,0}(2,2)$	- 0.0286	0.0609	-0.0250
$C_{-1,-1}(1,1)$	- 0.2354	0.2008	-0.1859
$C_{-1,-1}(1,2)$	0.1276	0.1874	0.1669
$C_{-1,-1}(2,1)$	- 0.0381	0.0843	-0.0371
$C_{-1,-1}(2,2)$	- 0.0862	0.1018	-0.1347
$C_{-1,-2}(1,1)$	0.03521	0.1930	0.0336
$C_{-1,-2}(1,2)$	0.1204	0.2286	0.1033
$C_{-1,-2}(2,1)$	- 0.0183	0.0271	-0.0231
$C_{-1,-2}(2,2)$	0.0372	0.3397	0.0562
$C_{-1,-3}(1,1)$	0.0241	0.0690	0.0052
$C_{-1,-3}(1,2)$	0.0275	0.0474	0.0387
$C_{-1,-3}(2,1)$	0.0293	0.0623	0.0301
$C_{-1,-3}(2,2)$	0.0156	0.0920	0.0251
$C_{-1,-4}(1,1)$	0.0169	0.0358	0.0205
$C_{-1,-4}(1,2)$	0.0186	0.0080	0.0225
$C_{-1,-4}(2,1)$	0.0059	0.0085	0.0111
$C_{-1,-4}(2,2)$	0.0011	0.0222	0.0003
λ	- 1.0011	0.0204	-1
K	4.0023	0.4025	4
$\Omega(1,1)$	1.0641	0.5403	1
$\Omega(1,2)$	0.0025	0.0036	0
$\Omega(2,2)$	1.1430	0.8204	1

Acknowledgements

This work is financially supported by FAR(2019) research grant of the University of Modena and Reggio E., Italy. We should like to thank the Editor in Chief of the journal, Professor Domenico Marinucci, the Associate Editor and the two anonymous referees for their valuable and constructive comments on the original version of the paper.

References

[1] Artin, M. (1991). *Algebra*. Prentice Hall, Englewood Cliffs, New Jersey, USA.

[2] Battaglia, F. (1983). Inverse autocovariances and a measure of linear determinism for a stationary process. *Journal of Time Series Analysis* **4** 79–87. MR0724539

[3] Bogert, B. P., Healy, M., and Tukey, J. (1963). *The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariances, cross-cepstrum, and saphe cracking*. Proceedings of the Symposium on Time Series Analysis (pp. 209–243), John Wiley & Sons, New York, U.S.

[4] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**(2), 217–226. MR0323048

[5] Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society Ser. B* **26**, 211–246. MR0192611

[6] Brockwell, P. J., and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlag, Berlin-Heidelberg-New York. MR0868859

[7] Cavicchioli, M. (2013). Spectral density of Markov-switching VARMA models. *Economics Letters* **121**, 218–220. MR3115354

[8] Cavicchioli, M. (2019a). Spectral representation and autocovariance structure of Markov switching DSGE models. *Communications in Statistics–Theory and Methods*. `https://doi.org/10.1080/03610926.2018.1563184`.

[9] Cavicchioli, M. (2019b). Fourth moment structure of Markov switching multivariate GARCH models. *Journal of Financial Econometrics*. `https://doi.org/10.1093/jjfinec/nbz020`.

[10] Cheng, J. (2016). Spectral density of Markov switching models: derivation, simulation studies and application. *Model Assisted Statistics and Applications* **11**(4), 277–291.

[11] Chiu, T.Y.M., Leonard, T., and Tsui, K-W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association* **91**, 198–210. MR1394074

[12] Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics* **14**(2), 277–293.

[13] Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer Science & Business Media. Springer-Verlag, Berlin-Heidelberg-New York. MR0812272

[14] Fackler, P. L. (2005). *Notes on Matrix Calculus*. North Carolina State University, U.S.

[15] Ghysels, E., and Owyang, M. T. (2012). Discussion of "An approach for identifying and predicting economic recessions in real-time using time-frequency functional models" by Holan, Yang, Matteson, and Wikle. *Applied Stochastic Models in Business and Industry* **28**, 500–501. MR3008940

[16] Gould, H. W. (1974). Coefficient identities for powers of Taylor and Dirichlet series. *The American Mathematical Monthly* **81**(1), 3–14. MR0340038

[17] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, N.J. MR1278033

[18] Hannan, E. J., and Deistler, M. (2012). *The Statistical Theory of Linear Systems*, vol. 70. SIAM. MR3397291

[19] Holan, S. H. (2004). *Time series exponential models: Theory and methods*. Ph.D. Thesis, Texas A&M University. MR2705974

[20] Holan, S. H., McElroy, T. S., and Wu, G. (2017). The cepstral model for multivariate time series: the vector exponential model. *Statistica Sinica* **27**, 23–42. MR3618159

[21] Holan, S. H., Wikle, C. K., Sullivan-Beckers, L. E., and Cocroft, R. B. (2010). Modeling complex phenotypes: generalized linear models using spectrogram predictors of animal communication signals. *Biometrics* **66**(3), 914–924. MR2758228

[22] Holan, S. H., Yang, W-H., Matteson, D. S., and Wikle, C. K. (2012). An approach for identifying and predicting economic recessions in real-time using time-frequency functional models. *Applied Stochastic Models in Business and Industry* **28**(6), 485–499. MR3008939

[23] Huhtanen, M., and Seiskari, O. (2012). Computational geometry of positive definiteness. *Linear Algebra and its Applications* **437**(7), 1562–1578. MR2946342

[24] Lee, H. T., and Yoder, J. K. (2007). A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratios. *Applied Economics* **39**(10), 1253–1265.

[25] Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. 2nd ed. Springer-Verlag, Berlin-Heidelberg-New York. MR2172368

[26] Magnus, J. R. (1988). *Linear Structures*. Griffin ed., London. MR0947343

[27] Magnus, J. R., and Neudecker, H. (1986). Symmetry, $0-1$ matrices and Jacobians. A Review. *Econometric Theory* **2**, 157–190.

[28] Magnus, J. R., and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Second ed. John Wiley and Sons, Chichester, New York. MR1698873

[29] Marée, S. (2012). *Correcting non positive definite correlation matrices*. BSc thesis Applied Mathematics, Delft University of Technology, Delft, Netherlands.

[30] McCullagh, T. S., and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, Cambridge, UK. MR3223057

[31] McElroy, T. S., and Holan, S. H. (2012). On the computation of

autocovariances for generalized Gegenbauer processes. *Statistica Sinica* **22**, 1661–1687. MR3027103

[32] McElroy, T. S., and Holan, S. H. (2014). Asymptotic theory of cepstral random fields. *The Annals of Statistics* **42**(1), 64–86. MR3161461

[33] Pataracchia, B. (2011). The spectral representation of Markov switching ARMA models. *Economics Letters* **112**, 11–15. MR2849173

[34] Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics–Theory and Methods* **12**(18), 2085–2094. MR0710561

[35] Proietti, T., and Luati, A. (2015). The generalised autocovariance function. *Journal of Econometrics* **186**, 245–257. MR3321536

[36] Proietti, T., and Luati, A. (2019). Generalised linear cepstral models for the spectrum of a time series. *Statistica Sinica* **29**, 1561–1583. MR3932530

[37] Rump, S. M. (2006). Verification of positive definiteness. *BIT Numerical Mathematics* **46**, 433–452. MR2238681

[38] Yang, W-H., Wikle, C. K., Holan, S. H., and Wildhaber, M. L. (2013). Ecological prediction with nonlinear multivariate time-frequency functional data models. *Journal of Agricultural, Biological, and Environmental Statistics* **18**(3), 450–474. MR3110902