JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 759–784 (2020)
Published online 12 July 2020 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12534

ORIGINAL ARTICLE

EFFICIENT BAYESIAN PARCOR APPROACHES FOR DYNAMIC MODELING OF MULTIVARIATE TIME SERIES

WENJIE ZHAO[a] AND RAQUEL PRADO[a*]

[a] *Department of Statistics, University of California at Santa Cruz, Santa Cruz, CA, USA*

A Bayesian lattice filtering and smoothing approach is proposed for fast and accurate modeling and inference in multivariate non-stationary time series. This approach offers computational feasibility and interpretable time-frequency analysis in the multivariate context. The proposed framework allows us to obtain posterior estimates of the time-varying spectral densities of individual time series components, as well as posterior measurements of the time-frequency relationships across multiple components, such as time-varying coherence and partial coherence. The proposed formulation considers multivariate dynamic linear models (MDLMs) on the forward and backward time-varying partial autocorrelation coefficients (TV-VPARCOR). Computationally expensive schemes for posterior inference on the multivariate dynamic PARCOR model are avoided using approximations in the MDLM context. Approximate inference on the corresponding time-varying vector autoregressive (TV-VAR) coefficients is obtained via Whittle's algorithm. A key aspect of the proposed TV-VPARCOR representations is that they are of lower dimension, and therefore more efficient, than TV-VAR representations. The performance of the TV-VPARCOR models is illustrated in simulation studies and in the analysis of multivariate non-stationary temporal data arising in neuroscience and environmental applications. Model performance is evaluated using goodness-of-fit measurements in the time-frequency domain and also by assessing the quality of short-term forecasting.

Received 18 July 2019; Accepted 04 May 2020

Keywords: Multivariate time series; Bayesian dynamic linear models; time-varying partial autocorrelations; time-varying vector autoregressions.

MOS subject classification: 62M10, 62M15, 62H12.

1. INTRODUCTION

Recent technological advances in scientific areas have led to multi-dimensional datasets with a complex temporal structure, often consisting of several time series components that are related over time. Inferring changes in the spectral content of each component, as well as time-varying relationships across components, is often relevant in applied areas. For example, understanding the interplay across temporal components derived from multi-channel/multi-location brain signals and brain imaging data is a key feature in brain connectivity studies (e.g., Astolfi *et al.*, 2008; Milde *et al.*, 2009; Cheung *et al.*, 2010; Omidvarnia *et al.*, 2014; Schmidt *et al.*, 2016; Yu *et al.*, 2016; Chiang *et al.*, 2017; Ting *et al.*, 2017, among others). Multivariate time series analysis is also important for filtering, smoothing, and prediction in environmental studies and finance where many variables are simultaneously measured over time (e.g., Tsay, 2013; Zhang, 2017).

Several time-domain, frequency-domain, and time-frequency approaches are available for modeling and inferring spectral characteristics of univariate non-stationary time series. However, a much more limited number of approaches are available for computationally efficient and scientifically interpretable analysis of multivariate non-stationary time series. Furthermore, currently available statistical tools have important practical limitations. For instance, vector autoregressions (VARs) are often used in the analysis of multi-channel electroencephalogram

* Correspondence to: Raquel Prado, Department of Statistics, University of California at Santa Cruz, Santa Cruz, CA, USA.
E-mail: raquel@soe.ucsc.edu

(EEG) data and estimation of cortical connectivity (see e.g., Cheung *et al.*, 2010; Chiang *et al.*, 2017); however, these models cannot capture the time-varying characteristics of these data.

Other approaches based on time-varying VARs are able to adapt to the non-stationary features of multi-channel EEG data, but to allow scalability only lead to point estimates of the spectral characteristics of the data and are highly dependent on a set of tuning parameters that are hard to elicit in practice. Alternative modeling frameworks that allow for full posterior inference while incorporating flexible and realistic dynamic structures (e.g., West *et al.*, 1999; Prado *et al.*, 2001; Nakajima and West, 2017), are either not available for multivariate time series, or they are highly computationally intensive, requiring Markov chain Monte Carlo (MCMC) sampling for posterior inference. Frequency-domain and time-frequency approaches have also been developed, but typically these methods are only able to handle multiple (not multivariate) stationary time series (e.g., Cadonna *et al.*, 2019), or are methodologically adequate and flexible (e.g., Bruce *et al.*, 2018; Li and Krafty, 2018), but computationally unfeasible to jointly analyze more than a relatively small number of multivariate time series components.

In the univariate context Yang *et al.* (2016) consider a Bayesian lattice filter approach for analyzing a single time series which uses univariate dynamic linear models (DLMs) to describe the evolution of the forward and backward partial autocorrelation coefficients of such series. A key feature of this approach is its computational appeal. A DLM representation of a univariate time-varying autoregression requires a model with a state-parameter vector of dimension P, where P is the order of the autoregression (see e.g., West *et al.*, 1999; Prado and West, 2010). Therefore, filtering and smoothing in this setting requires the inversion of $P \times P$ matrices at each time t. Alternatively, the DLM formulation in the PARCOR domain requires fitting $2 \times P$ DLMs, P for the forward coefficients and P for the backward coefficients, where each DLM has a univariate state-space parameter, fully avoiding matrix inversions and resulting in computational savings for cases in which $P \geq 2$.

In this article, we extend the Bayesian lattice filter approach of Yang *et al.* (2016) to the multivariate case. Our proposed models offer several advantages over currently available multivariate approaches for non-stationary time series including computational feasibility for joint analysis of relatively large-dimensional multivariate time series, and interpretable time-frequency analysis in the multivariate context. In particular, the proposed framework leads to posterior estimates of the time-varying spectral densities of each individual time series, as well as posterior measurements of the time-frequency relationships across multiple time series over time, such as time-varying coherence and partial coherence. We note that extending the approach Yang *et al.* (2016) to the multivariate case is non-trivial, as the closed-form inference used in the univariate DLM formulation of the lattice filter is not available for the multivariate case considered here. Multivariate DLM theory (West and Harrison; Prado and West, 2010) allows for full posterior inference in closed-form only when the covariance matrices of the innovations at the observation level and those at the system level are known, which is rarely the case in practice. Full posterior inference via MCMC can be obtained for more general multivariate DLM settings, but such posterior sampling schemes are very computationally expensive, making them only feasible when dealing with a small number of time series of small/moderate time lengths, and low-order TV-VAR models. We address these challenges by approximating the covariance matrices of the innovations at the observational level for the multivariate dynamic forward and backward PARCOR models using the approach of Triantafyllopoulos (2007). In addition, we use discount factors to specify the structure of the covariance matrices at the system levels. Our framework casts the time-varying multivariate representation of the input–output relations between the vectorial forward and backward predictions of a multivariate time series process – and their corresponding forward and backward matrices of PARCOR coefficients – as a Bayesian multivariate state-space model. Once approximate posterior inference is obtained for the multivariate time-varying PARCOR coefficients, posterior estimates for the implied time-varying vector autoregressive (TV-VAR) coefficient matrices and innovations covariance matrices can be obtained via Whittle's algorithm (Zhou, 1992). Similarly, posterior estimates for any function of such matrices, such as the multivariate spectra and functions of the spectra, can also be obtained. A key feature of the proposed TV-VPARCOR representation is that it is more parsimonious and flexible than directly working with the TV-VAR state-space representation. We illustrate this in the analyses of simulated and real data presented in Sections 3 and 4. We also propose a method for selecting the number of stages in the TV-VPARCOR setting based on an approximate calculation of the deviance information criterion (DIC).

J. Time Ser. Anal. **41**: 759–784 (2020)
DOI: 10.1111/jtsa.12534

The article is organized as follows. Section 2 presents the models and discusses approximate posterior inference. Section 3 illustrates the performance of the proposed TV-VPARCOR models in simulation studies. Comparisons with results obtained from DLM representations of TV-VAR models are also provided. Section 4 presents analyses of two real multivariate temporal datasets. The first application considers joint analysis of multi-channel EEG data and the second one illustrates the analysis and forecasting of multi-location bivariate wind components. Finally, Section 5 presents a discussion and briefly describes potential future developments.

2. MODELS AND METHODS FOR POSTERIOR INFERENCE

2.1. TV-VAR Models and Lattice Filters

Let $\boldsymbol{x}_t$ be a $K \times 1$ vector time series for $t = 1, \ldots, T$. A time-varying vector autoregressive model of order P, referred to as TV-VAR(P), is given by

$$\boldsymbol{x}_t = \boldsymbol{A}_{t,1}^{(P)} \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{A}_{t,P}^{(P)} \boldsymbol{x}_{t-P} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t),$$

where $\boldsymbol{A}_{t,j}^{(P)}$ is the $K \times K$ matrix of time-varying coefficients at lag j, $j = 1, \ldots, P$, and $\boldsymbol{\Sigma}_t$ is the $K \times K$ innovations variance–covariance matrix at time t. The $\boldsymbol{\epsilon}_t$s are assumed to be independent over time.

Yang *et al.* (2016) consider a Bayesian lattice filter dynamic linear modeling approach for the case of univariate time-varying autoregressions (TVAR), that is, when $K = 1$ above. Such approach is based on a lattice structure formulation of the univariate Durbin–Levinson algorithm (see, e.g., Brockwell and Davis, 1991; Shumway and Stoffer, 2017) used in Kitagawa (2010). The idea is to obtain posterior estimation on the forward and backward time-varying PARCOR coefficients using a computationally efficient lattice filter representation. Once dynamic PARCOR estimation is obtained, estimates of the TVAR coefficients can be derived using the Durbin–Levinson recursion. The main advantage of using the dynamic PARCOR lattice filter representation instead of a dynamic linear model TVAR representation such as that used in West *et al.* (1999), is that the former avoids the inversion of $P \times P$ matrices required in the TVAR DLM filtering and smoothing equations. Instead, the PARCOR approach considers $2P$ dynamic linear models with univariate state parameters (e.g., P DLMs with univariate state parameters for the forward coefficients and P DLMs with univariate state parameters for the backward coefficients), completely avoiding matrix inversions. This is important for computational efficiency when considering models with $P > 2$ and large T. The PARCOR approach also offers additional modeling advantages due to the fact that considering $2P$ DLMs with univariate state parameters generally provides more flexibility than using a single DLM TVAR with P-dimensional state parameters.

We extend the approach of Yang *et al.* (2016) to consider multivariate non-stationary time series. More specifically, we consider Bayesian multivariate DLMs that use the multivariate Whittle algorithm (Zhou, 1992), also known as the multivariate Durbin–Levinson algorithm (Brockwell and Davis, 1991), to obtain a representation of the TV-VAR coefficient matrices in terms of time-varying PARCOR matrices as follows. Let $\boldsymbol{f}_t^{(P)}$ and $\boldsymbol{b}_t^{(P)}$ be the K-dimensional prediction error vectors at time t for the forward and backward TV-VAR(P) model respectively, where,

$$\boldsymbol{f}_t^{(P)} = \boldsymbol{x}_t - \sum_{j=1}^{P} \boldsymbol{A}_{t,j}^{(P)} \boldsymbol{x}_{t-j}, \quad \text{and} \quad \boldsymbol{b}_t^{(P)} = \boldsymbol{x}_t - \sum_{j=1}^{P} \boldsymbol{D}_{t,j}^{(P)} \boldsymbol{x}_{t+j}.$$

$\boldsymbol{A}_{t,j}^{(P)}$ and $\boldsymbol{D}_{t,j}^{(P)}$ denote, respectively, the $K \times K$ time-varying matrices of forward and backward TV-VAR(P) coefficients for $j = 1, \ldots, P$. Similarly, $\boldsymbol{A}_{t,j}^{(m)}$ and $\boldsymbol{D}_{t,j}^{(m)}$ denote the time-varying matrices of forward and backward TV-VAR(m) coefficients for $j = 1, \ldots, m$. Then, we write the m-stage of the lattice filter in terms of the pair of

DOI: 10.1111/jtsa.12534

input–output relations between the forward and backward K-dimensional vector predictions, as follows,

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{\Lambda}_{t,m}^{(m)} \boldsymbol{b}_{t-m}^{(m-1)} + \boldsymbol{f}_t^{(m)}, \qquad \boldsymbol{f}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t,f,m}), \tag{1}$$
$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{\Theta}_{t,m}^{(m)} \boldsymbol{f}_{t+m}^{(m-1)} + \boldsymbol{b}_t^{(m)}, \qquad \boldsymbol{b}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t,b,m}), \tag{2}$$

where $\boldsymbol{\Lambda}_{t,m}^{(m)}$ and $\boldsymbol{\Theta}_{t,m}^{(m)}$ are, respectively, the $K \times K$ matrices of time-varying forward and backward PARCOR coefficients for $m = 1, \ldots, P$. Note that for stationary AR(P), that is, models with $K = 1$ and static AR coefficients in the stationary region, the forward and backward PARCOR coefficients are equal, that is, $\lambda_m^{(m)} = \theta_m^{(m)}$ for all m. For general K and non-stationary processes the forward and backward PARCOR coefficients are not the same.

For each stage m of the lattice structure above, we obtain the forward and backward TV-VAR coefficient matrices, $\boldsymbol{A}_{t,m}^{(P)}$ and $\boldsymbol{D}_{t,m}^{(P)}$, from the time-varying forward and backward PARCOR coefficient matrices, $\boldsymbol{\Lambda}_{t,m}^{(m)}$ and $\boldsymbol{\Theta}_{t,m}^{(m)}$, using Whittle's algorithm (see, e.g., Zhou, 1992), that is,

$$\boldsymbol{A}_{t,j}^{(m)} = \boldsymbol{A}_{t,j}^{(m-1)} - \boldsymbol{A}_{t,m}^{(m)} \boldsymbol{D}_{t,m-j}^{(m-1)}, \tag{3}$$
$$\boldsymbol{D}_{t,j}^{(m)} = \boldsymbol{D}_{t,j}^{(m-1)} - \boldsymbol{D}_{t,m}^{(m)} \boldsymbol{A}_{t,m-j}^{(m-1)}, \quad j = 1, \ldots, m-1, \tag{4}$$

with $\boldsymbol{A}_{t,m}^{(m)} = \boldsymbol{\Lambda}_{t,m}^{(m)}$ and $\boldsymbol{D}_{t,m}^{(m)} = \boldsymbol{\Theta}_{t,m}^{(m)}$, for $m = 1, \ldots, P$.

2.2. Model Specification and Inference

Our proposed model specification uses Equations (1) and (2) as observational level equations of multivariate DLMs (West and Harrison; Prado and West, 2010) on the forward and backward PARCOR time-varying coefficients. These multivariate DLMs are specified as follows. For each t, let $\mathrm{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)})$ and $\mathrm{vec}(\boldsymbol{\Theta}_{t,m}^{(m)})$ be the vectorized forward and backward PARCOR coefficients, that is, these are K^2 vectors obtained by stacking the forward and backward PARCOR coefficient matrices at time t, $\boldsymbol{\Lambda}_{t,m}^{(m)}$ and $\boldsymbol{\Theta}_{t,m}^{(m)}$, by columns respectively. In addition, define the forward and backward $K \times K^2$ matrices $\boldsymbol{F}_{t+m}^{(m-1)} = (\boldsymbol{f}_{t+m}^{(m-1)}) \otimes \boldsymbol{I}_{K\times K}$ and $\boldsymbol{B}_{t-m}^{(m-1)} = (\boldsymbol{b}_{t-m}^{(m-1)}) \otimes \boldsymbol{I}_{K\times K}$, where $\boldsymbol{I}_{K\times K}$ denotes the $K \times K$ identity matrix and $\otimes$ denotes the Kronecker product. Then, Equations (1) and (2) can be rewritten as

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{B}_{t-m}^{(m-1)} \mathrm{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)}) + \boldsymbol{f}_t^{(m)}, \quad \boldsymbol{f}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t,f,m}), \tag{5}$$
$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{F}_{t+m}^{(m-1)} \mathrm{vec}(\boldsymbol{\Theta}_{t,m}^{(m)}) + \boldsymbol{b}_t^{(m)}, \quad \boldsymbol{b}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t,b,m}), \tag{6}$$

which correspond to the observational equations of two multivariate dynamic linear regressions on $\boldsymbol{f}_t^{(m-1)}$ and $\boldsymbol{b}_t^{(m-1)}$, with dynamic coefficients $\mathrm{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)})$ and $\mathrm{vec}(\boldsymbol{\Theta}_{t,m}^{(m)})$ respectively. To complete the MDLM structure we specify random walk evolution equations for $\mathrm{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)})$ and $\mathrm{vec}(\boldsymbol{\Theta}_{t,m}^{(m)})$ as follows,

$$\mathrm{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)}) = \mathrm{vec}(\boldsymbol{\Lambda}_{t-1,m}^{(m)}) + \boldsymbol{\epsilon}_{t,f,m}, \quad \boldsymbol{\epsilon}_{t,f,m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_{t,f,m}), \tag{7}$$
$$\mathrm{vec}(\boldsymbol{\Theta}_{t,m}^{(m)}) = \mathrm{vec}(\boldsymbol{\Theta}_{t-1,m}^{(m)}) + \boldsymbol{\epsilon}_{t,b,m}, \quad \boldsymbol{\epsilon}_{t,b,m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_{t,b,m}), \tag{8}$$

where $\boldsymbol{W}_{t,f,m}$ and $\boldsymbol{W}_{t,b,m}$ are time dependent system covariance matrices. Finally, we specify prior distributions for $\mathrm{vec}(\boldsymbol{\Lambda}_{0,m}^{(m)})$ and $\mathrm{vec}(\boldsymbol{\Theta}_{0,m}^{(m)})$ and all m. We use conjugate normal priors for these parameters, that is, we assume

$$\mathrm{vec}(\boldsymbol{\Lambda}_{0,m}^{(m)}) | \mathcal{D}_{0,f,m} \sim \mathcal{N}(\boldsymbol{m}_{0,f,m}, \boldsymbol{C}_{0,f,m}), \tag{9}$$
$$\mathrm{vec}(\boldsymbol{\Theta}_{0,m}^{(m)}) | \mathcal{D}_{0,b,m} \sim \mathcal{N}(\boldsymbol{m}_{0,b,m}, \boldsymbol{C}_{0,b,m}), \tag{10}$$

DOI: 10.1111/jtsa.12534

where $\mathcal{D}_{0,f,m}$ and $\mathcal{D}_{0,b,m}$ denote the information available at time $t = 0$ for the forward and backward state parameter vectors respectively.

Given $\boldsymbol{\Sigma}_{t,f,m}$, and $\boldsymbol{W}_{t,f,m}$ for all $t = 1, \dots, T$, and all $m = 1, \dots, P$, Equations (5), (7), and (9) define a normal MDLM (see, e.g., Prado and West, 2010, Chapter 10) for the forward time-varying PARCOR. Similarly, given $\boldsymbol{\Sigma}_{t,b,m}$ and $\boldsymbol{W}_{t,b,m}$ for all t and all m, (6), (8), and (10) define a normal MDLM for the backward time-varying PARCOR.

Note that posterior inference in the case of univariate models with $K = 1$ is available in closed form via the DLM filtering and smoothing equations. This is used in Yang *et al.* (2016) to obtain posterior inference in this univariate case. However, posterior inference in the general multivariate setting proposed here is not available in closed form when the observational and system covariance matrices are unknown, which is typically the case in practical settings. Therefore, as explained below, we use discount factors to specify $\boldsymbol{W}_{t,f,m}$, and $\boldsymbol{W}_{t,b,m}$. We also assume $\boldsymbol{\Sigma}_{t,f,m} = \boldsymbol{\Sigma}_{f,m}$ and $\boldsymbol{\Sigma}_{t,b,m} = \boldsymbol{\Sigma}_{b,m}$ for all t, and use the approach of Triantafyllopoulos (2007) to obtain estimates of $\boldsymbol{\Sigma}_{f,m}$ and $\boldsymbol{\Sigma}_{b,m}$, which allows us to get approximate posterior inference in the multivariate case.

We follow Ameen and Harrison (1985), and first define the $K^2 \times K^2$ system covariance matrices using discount factors by setting

$$\boldsymbol{\Delta}_{f,m} = diag(\boldsymbol{\delta}_{f,m,1}^{-1/2}, \dots, \boldsymbol{\delta}_{f,m,K}^{-1/2}), \quad \text{and} \quad \boldsymbol{\Delta}_{b,m} = diag(\boldsymbol{\delta}_{b,m,1}^{-1/2}, \dots, \boldsymbol{\delta}_{b,m,K}^{-1/2}),$$

where each component, $\boldsymbol{\delta}_{\cdot,m,i}$, is a K-dimensional vector that contains the discount factors for each of the K components at stage m. Although we can assume different discount factors for different elements of $\boldsymbol{\delta}_{\cdot,m,k}$ and also across different ks, in practice we usually set all the elements of $\boldsymbol{\delta}_{f,m,k}$ equal to $\delta_{f,m}$ and all the elements of $\boldsymbol{\delta}_{b,m,k}$ equal to $\delta_{b,m}$ for all $k = 1, \dots, K$, and then choose $\delta_{f,m}$ and $\delta_{b,m}$ optimally according to some criterion for each stage m (this is discussed in Section 2.3). This structure for $\boldsymbol{\Delta}_{f,m}$ and $\boldsymbol{\Delta}_{b,m}$ allows us to obtain closed form expressions for $\boldsymbol{W}_{t,f,m}$ and $\boldsymbol{W}_{t,b,m}$ sequentially over time.

We now describe the full algorithm for approximate posterior inference in the forward TV-VPARCOR model. The algorithm for the backward model is similar. Let $\mathcal{D}_{t,f,m}$ denote all the information available up to time t at stage m for the forward model, with $\mathcal{D}_{t,f,m} = \{\mathcal{D}_{t-1,f,m}, \boldsymbol{f}_t^{(m-1)}\}$. Consider the posterior expectation of $\boldsymbol{\Sigma}_{f,m}$ up to time t, that is, $E(\boldsymbol{\Sigma}_{f,m}|\mathcal{D}_{t,f,m})$, and assume that $\lim_{t\to\infty} E(\boldsymbol{\Sigma}_{f,m}|\mathcal{D}_{t,f,m}) = \boldsymbol{\Sigma}_{f,m}$. Let $n_{0,f,m}$ be a positive scalar and $\boldsymbol{S}_{0,f,m}$ be the prior expectation of $\boldsymbol{\Sigma}_{f,m}$. Assume that at time $t-1$, we have that $\text{vec}(\boldsymbol{\Lambda}_{t-1,m}^{(m)})|\mathcal{D}_{t-1,f,m}$ is approximately distributed as $N(\boldsymbol{m}_{t-1,f,m}, \boldsymbol{C}_{t-1,f,m})$, and so, $E(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{t-1,f,m})$ is approximated by $\boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{m}_{t-1,f,m}$ and $V(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{t-1,f,m})$ is approximated by $\boldsymbol{Q}_{t-1,f,m} = \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{R}_{t,f,m}(\boldsymbol{B}_{t-1}^{(m-1)})' + \boldsymbol{S}_{t-1,f,m}$, with $\boldsymbol{R}_{t,f,m} = \boldsymbol{C}_{t-1,f,m} + \boldsymbol{W}_{t,f,m}$, for some $\boldsymbol{S}_{t-1,f,m}$. Then, following Theorem 1 of Triantafyllopoulos (2007) we have that, if $\boldsymbol{\Sigma}_{f,m}$ is bounded, $\boldsymbol{S}_{t,f,m}$ will approximate $\boldsymbol{\Sigma}_{f,m}$ for t large, with

$$\boldsymbol{S}_{t,f,m} = \frac{1}{(n_{0,f,m} + t)} \left(n_{0,f,m}\boldsymbol{S}_{0,f,m} + \sum_{i=1}^{t} \boldsymbol{S}_{i-1,f,m}^{1/2}\boldsymbol{Q}_{i,f,m}^{-1/2}\boldsymbol{e}_{i,f,m}\boldsymbol{e}'_{i,f,m}\boldsymbol{Q}_{i,f,m}^{-1/2}\boldsymbol{S}_{i-1,f,m}^{1/2} \right), \tag{11}$$

where in our case $\boldsymbol{e}_{t,f,m} = \boldsymbol{f}_t^{(m-1)} - \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{m}_{t-1,f,m}$, and $\boldsymbol{S}_{i-1,f,m}^{1/2}, \boldsymbol{Q}_{i,f,m}^{-1/2}$ are symmetric square roots of the matrices $\boldsymbol{S}_{i-1,f,m}$ and $\boldsymbol{Q}_{i,f,m}^{-1}$ respectively, based on the spectral decomposition factorization of symmetric positive definite matrices for all $i = 1, \dots, t$.

Using the approximation above we obtain the filtering equations below for approximate inference in the forward TV-VPARCOR model.

- The one-step ahead forecast mean and covariance at time t are given by:

$$E(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{t-1,f,m}) \approx \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{m}_{t-1,f,m}.$$

  and

$$V(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{t-1,f,m}) \approx \boldsymbol{Q}_{t,f,m} = \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{R}_{t,f,m}(\boldsymbol{B}_{t-m}^{(m-1)})' + \boldsymbol{S}_{t-1,f,m},$$

DOI: 10.1111/jtsa.12534

where $\boldsymbol{R}_{t,f,m} = \boldsymbol{C}_{t-1,f,m} + \boldsymbol{W}_{t,f,m}$ and $\boldsymbol{W}_{t,f,m} = \Delta_{f,m}\boldsymbol{C}_{t-1,f,m}\Delta_{f,m} - \boldsymbol{C}_{t-1,f,m}$.

- The one-step forecast error vector is given by $\boldsymbol{e}_{t,f,m} = \boldsymbol{f}_t^{(m-1)} - \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{m}_{t-1,f,m}$.
- Using Bayes' theorem and the equations above we can obtain the approximate posterior distribution at time t as $\text{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)})|\mathcal{D}_{t,f,m} \approx \mathcal{N}(\boldsymbol{m}_{t,f,m}, \boldsymbol{C}_{t,f,m})$, where

$$\boldsymbol{m}_{t,f,m} = \boldsymbol{m}_{t-1,f,m} + \boldsymbol{U}_{t,f,m}\boldsymbol{e}_{t,f,m}, \tag{12}$$

$$\boldsymbol{C}_{t,f,m} = \boldsymbol{\Delta}_{f,m}\boldsymbol{C}_{t-1,f,m}\boldsymbol{\Delta}_{f,m} + \boldsymbol{U}_{t,f,m}\boldsymbol{Q}_{t,f,m}\boldsymbol{U}'_{t,f,m}, \tag{13}$$

$$\boldsymbol{U}_{t,f,m} = \boldsymbol{\Delta}_{f,m}\boldsymbol{C}_{t-1,f,m}\boldsymbol{\Delta}_{f,m}\boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{Q}_{t,f,m}^{-1}. \tag{14}$$

Approximate filtering and predictive distributions for $\text{vec}(\boldsymbol{\Lambda}_{t,m}^{m})|\mathcal{D}_{t,f,m}, \boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{t,f,m}$ and $\boldsymbol{f}_{t+h}^{(m-1)}|\mathcal{D}_{t,f,m}$ for a positive integer $h > 0$ can also be obtained by taking $\boldsymbol{\Sigma}_{f,m} = \boldsymbol{S}_{t,f,m}$.

After applying the filtering equations up to time T, it is possible to compute approximate smoothing distributions for the forward PARCOR model by setting $\boldsymbol{\Sigma}_{f,m} = \boldsymbol{S}_{T,f,m}$. This leads to approximate smoothing distributions

$$\text{vec}(\boldsymbol{\Lambda}_{t,m}^{(m)})|\mathcal{D}_T \approx \mathcal{N}(\boldsymbol{a}_{T,f,m}(t-T), \boldsymbol{R}_{T,f,m}(t-T)),$$

where the mean and covariance are computed recursively via

$$\boldsymbol{a}_{T,f,m}(t-T) = \boldsymbol{m}_{t,f,m} - \boldsymbol{J}_{t,f,m}(\boldsymbol{a}_{t+1,f,m} - \boldsymbol{a}_{T,f,m}(t-T+1)), \tag{15}$$

$$\boldsymbol{R}_{T,f,m}(t-T) = \boldsymbol{C}_{t,f,m} - \boldsymbol{J}_{t,f,m}(\boldsymbol{R}_{t+1,f,m} - \boldsymbol{R}_{T,f,m}(t-T+1)), \tag{16}$$

for $t = (T-1), \ldots, 1$, with $\boldsymbol{J}_{t,f,m} = \boldsymbol{C}_{t,f,m}\boldsymbol{R}_{t+1,f,m}^{-1}$, and starting values $\boldsymbol{a}_T(0) = \boldsymbol{m}_{T,f,m}$ and $\boldsymbol{R}_{T,f,m}(0) = \boldsymbol{C}_{T,f,m}$. Filtering and smoothing equations can be obtained for the backward PARCOR model in a similar manner. Finally, the algorithm for approximate posterior estimation is as follows.

Algorithm

1. Given hyperparameters $\{P, \boldsymbol{\Delta}_{f,m}, \boldsymbol{\Delta}_{b,m}; m = 1, \ldots, P\}$, set $\boldsymbol{f}_t^{(0)} = \boldsymbol{b}_t^{(0)} = \boldsymbol{x}_t,$ for $t = 1, \ldots, T$.
2. Use $\{\boldsymbol{f}_t^{(0)}\}$ and $\{\boldsymbol{b}_t^{(0)}\}$ as vectors of responses in the observational level Equations (1) and (2) respectively, which, combined with the random walk evolution Equations (7) and (8), and the priors (9) and (10), define the multivariate PARCOR forward and backward models. Then, use the sequential filtering Equations (12)–(14) to obtain the estimated $\{\boldsymbol{S}_{T,f,1}\}$ and $\{\boldsymbol{S}_{T,b,1}\}$. Use the sequential filtering Equations (12)–(14) along with the smoothing Equations (15) and (16) to obtain a series of estimated parameters $\{\text{vec}(\hat{\boldsymbol{\Lambda}}_{t,1}^{(1)})\}$, $\{\text{vec}(\hat{\boldsymbol{\Theta}}_{t,1}^{(1)})\}$ for $t = 1 : T$. These estimated parameters are set at the posterior means of the smoothing distributions, that is, the values in (16) for the forward case and a similar equation in the backward case.
3. Use the observational Equations (1) and (2) to obtain the new series of forward and backward prediction errors, $\{\boldsymbol{f}_t^{(1)}\}$ and $\{\boldsymbol{b}_t^{(1)}\}$, for $t = 1, \ldots, T$.
4. Repeat steps 2 and 3 above until $\{\text{vec}(\hat{\boldsymbol{\Lambda}}_{t,m}^{(m)})\}$, $\{\text{vec}(\hat{\boldsymbol{\Theta}}_{t,m}^{(m)})\}$, $\{\boldsymbol{S}_{T,f,m}\}$ and $\{\boldsymbol{S}_{T,b,m}\}$ have been obtained for all $m = 1, \ldots, P$.
5. Finally, use $\{\text{vec}(\hat{\boldsymbol{\Lambda}}_{t,m}^{(m)})\}$ and $\{\text{vec}(\hat{\boldsymbol{\Theta}}_{t,m}^{(m)})\}$, for $m = 1, \ldots, P$, and Equations (3) and (4) to obtain the forward and backward TV-VAR coefficient matrices via Whittle's algorithm.

2.3. Model Selection and Time-Frequency Representation

To select the optimal model order and discount factors, we begin by specifying a potential maximum value of P, say $P_{\max}$, for the model order. At level m we search for the optimal values of $\boldsymbol{\Delta}_{f,m}$ and $\boldsymbol{\Delta}_{b,m}$. In other words, at level $m = 1$ we search for the combination of values of $\boldsymbol{\Delta}_{f,1}$ and $\boldsymbol{\Delta}_{b,1}$ maximizing the log-likelihood resulting from (1)

DOI: 10.1111/jtsa.12534

with $m = 1$. Using the selected optimal $\mathbf{\Delta}_{f,1}$ and $\mathbf{\Delta}_{b,1}$, we can obtain the corresponding series $\{\boldsymbol{f}_t^{(2)}\}$ and $\{\boldsymbol{b}_t^{(2)}\}$, for $t = 1, \ldots, T$, as well as the maximum log-likelihood value $\mathcal{L}_{f,1}$. Then, we repeat the above search procedure for stage two, that is, $m = 2$, using the output $\{\boldsymbol{f}_t^{(2)}\}$ and $\{\boldsymbol{b}_t^{(2)}\}$ obtained from implementing the filtering and smoothing equations with the previously selected hyperparameters $\mathbf{\Delta}_{f,1}$ and $\mathbf{\Delta}_{b,1}$. We obtain optimal $\mathbf{\Delta}_{f,2}$, $\mathbf{\Delta}_{b,2}$ as well as $\{\boldsymbol{f}_t^{(3)}\}$ and $\{\boldsymbol{b}_t^{(3)}\}$, for $t = 1, \ldots, T$. We also obtain the value of the corresponding maximum log-likelihood $\mathcal{L}_{f,2}$. We repeat the procedure until the set $\{\mathbf{\Delta}_{f,m}, \mathbf{\Delta}_{b,m}, \mathcal{L}_{f,m}\}, m = 1, \ldots P_{\max}$, has been selected. We then consider two different methods for selecting the optimal model order as described below. Note that one can also obtain the optimal likelihood values from the backward model, $\mathcal{L}_{b,m}$, for $m = 1, \ldots P_{\max}$. For all the examples and real data analyses presented below we choose the optimal model orders based on the optimal likelihood values for the forward model. Similar results were obtained based on the optimal likelihood values for the backward models.

Method 1: Scree plots. This method was used by Yang *et al.* (2016) to select the model order visually by plotting $\mathcal{L}_{f,m}$ against the order m. The idea is that, when the observed vector of time series truly follows a TV-VAR model, the values of $\mathcal{L}_{f,m}$ will stop increasing after a specific lag and this lag is then chosen to be the model order. A numerical version of this method can also be implemented by computing the percent of change in the likelihood going from $\mathcal{L}_{f,m-1}$ to $\mathcal{L}_{f,m}$, however, here we use scree plots as a visualization tool and use the model selection criterion below to numerically find an optimal model order.

Method 2: DIC model selection criterion. We consider an approach based on the DIC to choose the model order (see Gelman *et al.*, 2014, and references therein). In general, for a model with parameters denoted as $\boldsymbol{\theta}$, the DIC is defined as

$$DIC = -2 \log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{DIC},$$

where $\boldsymbol{y}$ denotes the data, $\hat{\boldsymbol{\theta}}_{Bayes}$ is the Bayes estimator of $\boldsymbol{\theta}$ and p_{DIC} is the effective number of parameters. The effective number of parameters is given by

$$p_{DIC} = 2 \left[\log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{Bayes}) - E_{post} \left(\log p(\boldsymbol{y}|\boldsymbol{\theta}) \right) \right],$$

where the expectation in the second term is an average of $\boldsymbol{\theta}$ over its posterior distribution. The expression above is typically estimated using samples $\boldsymbol{\theta}^s, s = 1, \ldots, S$, from the posterior distribution as

$$\hat{p}_{DIC} = 2 \left[\log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{Bayes}) - \frac{1}{S} \sum_{s=1}^{S} \log p(\boldsymbol{y}|\boldsymbol{\theta}^s) \right].$$

Note, however, that in our case we do not have samples from the exact posterior distribution of the parameters since we are using approximate inference to avoid computationally costly exact inference via MCMC. Therefore, for a given model order m we compute the likelihood term in the DIC calculation approximately using the forward filtering distributions as explained below. Also, note that, fitting a PARCOR model at stage m requires fitting all the models of the previous $m - 1$ stages. Therefore, the effective number of parameters at stage m is computed by adding the estimated effective number of parameters of stage m plus the estimated effective number of parameters for the previous $m - 1$ stages. In other words, for each stage m :

- Compute the estimated implied log-likelihood from Equation (5) for $t = 1, \ldots, T$, using $\text{vec}(\hat{\mathbf{\Lambda}}_{t,m}^{(m)})$ and $\boldsymbol{S}_{T,m,f}$. In this way we obtain the first term in the calculation of the DIC for model order m.
- Obtain samples, $\text{vec}(\mathbf{\Lambda}_{t,m,s}^{(m)})$, for $s = 1, \ldots, S$, from the approximate sequential filtering equations with distributions $\mathcal{N}(\boldsymbol{m}_{t,f,m}, \boldsymbol{C}_{t,f,m})$, and use these samples to compute the estimated number of parameters related only to stage m which we denote as $\hat{p}_{DIC,m}^{m}$. Note that, as mentioned above, stage m requires fitting all the PARCOR models for the previous $(m - 1)$ stages and so, in the final DIC calculation at stage m the total estimated

DOI: 10.1111/jtsa.12534

effective number of parameters is computed as

$$\hat{p}^m_{DIC} = \sum_{l=1}^{m} \hat{p}^l_{DIC,l}.$$

We denote the final estimated DIC for model order m as $\widehat{DIC}_m$.

2.4. Posterior Summaries

Once an optimal TV-VPARCOR model is chosen we can obtain posterior summaries of any quantities associated to such model. For instance, we can obtain posterior summaries of the TV-VPARCOR coefficients over time at each stage, and consequently summaries of the corresponding TV-VAR coefficients over time.

Time-frequency representations are generally more useful in practice, and these can be obtained by computing the spectral density matrix, $\boldsymbol{g}(t, \omega)$, for any time t and frequency $\omega \in (0, 1/2)$, as well as measurements derived from this matrix such as coherence, partial coherence, or partial directed coherence (PDC). The spectral density matrix is estimated as

$$\hat{\boldsymbol{g}}(t, \omega) = \hat{\boldsymbol{\Phi}}^{-1}(t, \omega) \times \hat{\boldsymbol{\Sigma}} \times \hat{\boldsymbol{\Phi}}^{*}(t, \omega)^{-1}, \tag{17}$$

where $\hat{\boldsymbol{\Phi}}(t, \omega) = \boldsymbol{I} - \sum_{m=1}^{P} \hat{\boldsymbol{A}}^{(P)}_{t,m} \exp\{-2\pi i m \omega\}$, with $i = \sqrt{-1}$ (see e.g., Shumway and Stoffer, 2017, Chapter 4). $\hat{\boldsymbol{\Sigma}}$ can be set at $\mathbf{S}_{T,f,P}$. Note that the spectral density matrix $\boldsymbol{g}(t, \omega)$ consists of individual spectra $g_{j,j}(t, \omega)$ for each component $j = 1, \dots, K$ of $\boldsymbol{x}_t$, and the cross-spectra $g_{i,j}(t, \omega)$ between components i and j. From these we can compute the estimated squared coherence between components i and j as

$$\hat{\rho}^2_{i,j}(t, \omega) = \frac{|\hat{g}_{i,j}(t, \omega)|^2}{\{\hat{g}_{i,i}(t, \omega)\hat{g}_{j,j}(t, \omega)\}},$$

for all $i \neq j$. This measure is used to estimate the power transfer between two components of the time series. Similarly, the partial squared coherence between components i and j can be estimated as follows. Let $\boldsymbol{c}(t, \omega) = \boldsymbol{g}^{-1}(t, \omega)$ be the inverse of the spectral density matrix with elements $c_{i,j}(t, \omega)$ for $i, j = 1, \dots, K$. Then, the estimated squared partial coherence between components i and j is given by

$$\hat{\gamma}^2_{i,j}(t, \omega) = \frac{|\hat{c}_{i,j}(t, \omega)|^2}{\{\hat{c}_{i,i}(t, \omega)\hat{c}_{j,j}(t, \omega)\}}.$$

The squared partial coherence is essentially the frequency domain squared correlation coefficient between components i and j after the removal of the linear effects of all the remaining components of $\boldsymbol{x}_t$. Directional measures such as the PDC and the direct transfer function (DTF) can also be computed (see e.g., Baccala and Sameshima, 2001; Kuś *et al.*, 2004; Astolfi *et al.*, 2008; Blinowska, 2011; Milde *et al.*, 2009; Omidvarnia *et al.*, 2014). Such measures provide information of directionality in the interactions between signals in a Granger causality sense. The estimated PDC from signal j to signal i at time t and frequency ω is given by

$$\widehat{PDC}_{i,j}(t, \omega) = \frac{\hat{\boldsymbol{\Phi}}_{i,j}(t, \omega)}{\sqrt{\hat{\boldsymbol{\Phi}}^{*}_{\cdot,j}(t, \omega)\hat{\boldsymbol{\Phi}}_{\cdot,j}(t, \omega)}},$$

  J. Time Ser. Anal. **41**: 759–784 (2020)
DOI: 10.1111/jtsa.12534

with $\hat{\mathbf{\Phi}}_{\cdot,j}$ the jth column of the matrix $\hat{\mathbf{\Phi}}(t,\omega)$. Similarly, the estimated DTF from signal j to signal i at time t and frequency ω is given by

$$\widehat{DTF}_{i,j}(t,\omega) = \frac{\hat{\mathbf{\Phi}}^{-1}_{i,j}(t,\omega)}{\sqrt{[\hat{\mathbf{\Phi}}^{-1}_{i,\cdot}(t,\omega)]^{*}\hat{\mathbf{\Phi}}^{-1}_{i,\cdot}(t,\omega)}},$$

where $\hat{\mathbf{\Phi}}^{-1}_{i,j}(t,\omega)$ is the (i,j)th element of the matrix $\hat{\mathbf{\Phi}}^{-1}(t,\omega)$ and $\hat{\mathbf{\Phi}}^{-1}_{i,\cdot}(t,\omega)$ is the ith row of $\mathbf{\Phi}^{-1}(t,\omega)$, with $\mathbf{A}^{*}$ denoting the Hermitian matrix of $\mathbf{A}$. The DTF shows all direct and so called 'cascade flows', for example, in the case of 3 signals, all propagations of the form $1 \to 2 \to 3$ and $1 \to 3$ would be reflected in the DTF between signals 1 and 3. On the other hand, PDC shows only direct flows between signals, that is, indirect propagations like $1 \to 2 \to 3$ are not included.

Finally, uncertainty measures for the spectral density matrix, and any functions of this matrix, can be obtained from the approximate filtering and smoothing posterior distributions of the forward and backward TV-VPARCOR models. This is done by sampling from the approximate posterior distributions of the TV-VPARCOR parameters described in Section 2.2. Then, each posterior sample of the model parameters is transformed into the corresponding spectral density matrix, or any other function of this matrix, allowing us to obtain a posterior sample of such function. Uncertainty measures for these functions are computed based on the samples. This is illustrated in Section 4.2.

2.5. Forecasting

We show how to obtain h-steps ahead forecasts. To have a non-explosive behavior in the forecasts, we assume the series is locally stationary in the future, that is, $\mathbf{\Lambda}^{(m)}_{t,m} = \mathbf{\Theta}^{(m)}_{t,m}$ at time $t = T+1, \dots, T+h$. Then, the approximate h-steps ahead forecast posterior distribution of the PARCOR coefficients, with $h > 0$, is approximated as $(\mathbf{\Lambda}^{(m)}_{T+h,m} | \mathcal{D}_{T,f,m}) \approx \mathcal{N}(\boldsymbol{m}_{T,f,m}(h), \boldsymbol{C}_{T,f,m}(h))$, where

$$\boldsymbol{m}_{T,f,m}(h) = \boldsymbol{m}_{T,f,m}; \qquad \boldsymbol{C}_{T,f,m}(h) = \boldsymbol{C}_{T,f,m} + h \cdot \boldsymbol{W}_{T+1,f,m},$$

with $\boldsymbol{W}_{T+1,f,m} = \Delta_{f,m}\boldsymbol{C}_{T,f,m}\Delta_{f,m} - \boldsymbol{C}_{T,f,m}$, for $m = 1, \dots, P$. Then, we apply Whittle's algorithm to transform the PARCOR coefficients, $\mathbf{\Lambda}^{(P)}_{T+h,P}$, into TV-VAR coefficients $\boldsymbol{A}^{(P)}_{T+h,j}$ and $\boldsymbol{D}^{(P)}_{T+h,j}$, for $j = 1, \dots, P$. Finally, we obtain the h-steps ahead forecasts using

$$\hat{\boldsymbol{x}}_{T+h} = \sum_{i=1}^{P} \hat{\boldsymbol{A}}^{(P)}_{T+h,i}\hat{\boldsymbol{x}}_{T+h-i} + \hat{\boldsymbol{\epsilon}}^{(P)}_{T+h}, \qquad \hat{\boldsymbol{\epsilon}}^{(P)}_{T+h} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{S}_{T,f,P}).$$

3. SIMULATION STUDIES

We illustrate our proposed approach in the analysis of simulated data. The relative performances of the models considered here, including that of the proposed TV-VPARCOR, were assessed by computing the average squared error (ASE) between the estimated spectral density matrix and the true spectral density matrix.

3.1. Bivariate TV-VAR(2) Processes

We simulated 50 bivariate time series of length $T = 1034$ from the following TV-VAR(2) model:

$$\boldsymbol{x}_t = \mathbf{\Phi}_{1,t}\boldsymbol{x}_{t-1} + \mathbf{\Phi}_{2,t}\boldsymbol{x}_{t-2} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2),$$

DOI: 10.1111/jtsa.12534

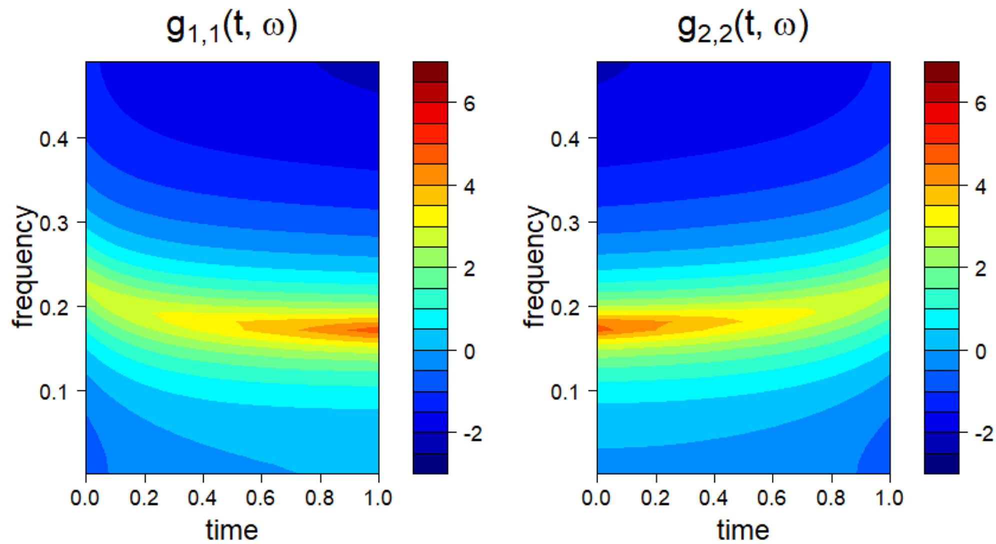

Figure 1. Case $\phi_{1,1,2} = 0$. Left: True log spectral density $g_{11}(t, \omega)$. Right: True log spectral density $g_{22}(t, \omega)$ [Color figure can be viewed at wileyonlinelibrary.com]

with

$$\boldsymbol{\Phi}_{1,t} = \begin{pmatrix} r_{1,t}\cos(\frac{2\pi}{\lambda_{1,t}}) & \phi_{1,1,2,t} \\ 0 & r_{2,t}\cos(\frac{2\pi}{\lambda_{2,t}}) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Phi}_{2,t} = \begin{pmatrix} -r_{1,t}^2 & \phi_{2,1,2,t} \\ 0 & -r_{2,t}^2 \end{pmatrix},$$

where $r_{1,t} = \frac{0.1}{T}t + 0.85$, $r_{2,t} = -\frac{0.1}{T}t + 0.95$, $r_{3,t} = \frac{0.2}{T} - 0.9$, $r_{4,t} = \frac{0.2}{T} + 0.7$, $\lambda_{1,t} = \frac{15}{T}t + 5$, and $\lambda_{2,t} = -\frac{10}{T}t + 15$. We also considered three different scenarios for the values of $\phi_{1,1,2,t}$, and $\phi_{2,1,2,t}$, namely (i) $\phi_{1,1,2,t} = \phi_{2,1,2,t} = 0$ for all t; (ii) $\phi_{1,1,2,t} = -0.8$ and $\phi_{2,1,2,t} = 0$ for all t; and (iii) $\phi_{1,1,2,t} = r_{3,t}$ and $\phi_{2,1,2,t} = r_{4,t}$.

The true 2×2 spectral matrix of this process is given by

$$\boldsymbol{g}(t, \omega) = \boldsymbol{\Phi}^{-1}(t, \omega) \times \boldsymbol{\Sigma} \times \boldsymbol{\Phi}^{*}(t, \omega)^{-1},$$

where $\boldsymbol{\Phi}(t, \omega) = \boldsymbol{I}_2 - \boldsymbol{\Phi}_{1,t}\exp\{-2\pi i\omega\} - \boldsymbol{\Phi}_{2,t}\exp\{-4\pi i\omega\}$, and $\boldsymbol{\Sigma} = \boldsymbol{I}_2$. The spectral matrix $\boldsymbol{g}(t, \omega)$ is symmetric, with corresponding components $g_{11}(t, \omega)$, $g_{12}(t, \omega)$, and $g_{22}(t, \omega)$, representing, respectively, the spectrum of the first component, the co-spectrum between the first and the second components, and the spectrum of the second component. The squared coherence between the first and second components is given by

$$\rho_{12}^2(t, \omega) = \frac{|g_{12}(t, \omega)|^2}{g_{11}(t, \omega)g_{22}(t, \omega)}.$$

Note that when $\phi_{1,1,2,t} = 0$ and $\phi_{2,1,2,t} = 0$ for all t (scenario (i)), the two processes are uncorrelated and $g_{1,2}(t, \omega) = \rho_{12}^2(t, \omega) = 0$ for all t and ω. Figure 1 shows the true log spectral densities $g_{11}(t, \omega)$ and $g_{22}(t, \omega)$ in this scenario. The true log spectral densities and square coherences for scenarios (ii) and (iii) are shown, respectively, in the top row plots of Figures 2 and 3.

We fit bivariate TV-VPARCOR models to each of the 50 simulated bivariate time series for t =1:1024 under cases (i), (ii), and (iii). We assess the forecasting performance of the model in all cases using the last 10 observations not included in the fit, that is, $t = 1025{:}1034$. We set a maximum of order $P_{\max} = 5$. The elements of the diagonal component of discount factor matrices $\boldsymbol{\Delta}_{f,m}$ and $\boldsymbol{\Delta}_{b,m}$, $\delta_{f,m}$, and $\delta_{b,m}$ respectively, were chosen from a grid of values in $(0.995, 1)$. We set the hyperparameters $n_{f,m,0} = n_{b,m,0} = 1$, $\boldsymbol{S}_{0,f,m} = \boldsymbol{S}_{0,b,m} = \boldsymbol{I}_2$, $\boldsymbol{m}_{0,f,m} = \boldsymbol{m}_{0,b,m} = (0, 0, 0, 0)'$

DOI: 10.1111/jtsa.12534

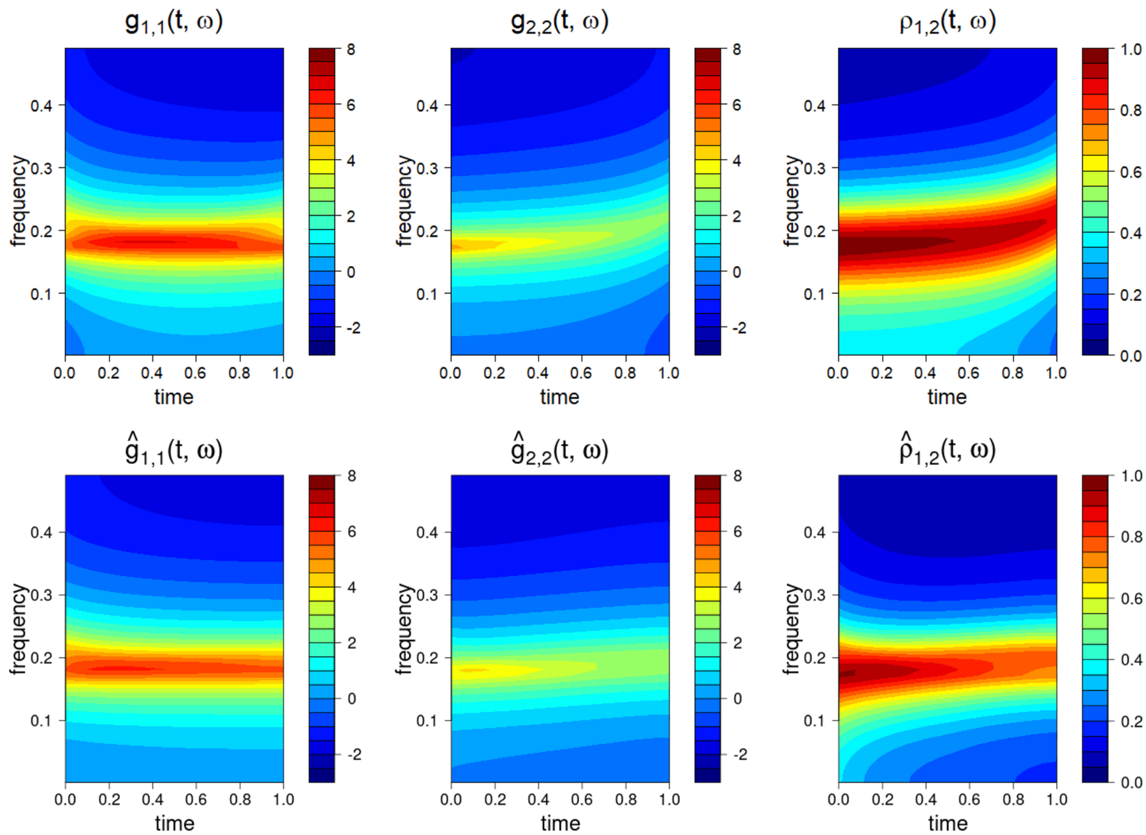

Figure 2. Case with $\phi_{1,1,2,t} = -0.8$ and $\phi_{2,1,2,t} = 0$ for all t. Top: True log spectral density $g_{11}(t, \omega)$ (left), true log spectral density $g_{22}(t, \omega)$ (middle), true squared coherence $\rho^2_{1,2}(t, \omega)$ (right). Bottom: Estimated $\hat{g}_{11}(t, \omega)$ (left), estimated $\hat{g}_{22}(t, \omega)$ (middle), estimated $\hat{\rho}^2_{1,2}(\omega, t)$ (right) [Color figure can be viewed at wileyonlinelibrary.com]

and $\boldsymbol{C}_{0,f,m} = \boldsymbol{C}_{0,b,m} = \boldsymbol{I}_4$. For comparison, we also fit TV-VAR models to the simulated bivariate data with model orders ranging from 1 to 5. Multivariate DLM representations of bivariate TV-VAR(m) processes were considered for each $m = 1, \dots, 5$. Each TV-VAR representation has an $4m$-dimensional state parameter vector. For each model order a single optimal discount factor, δ_m was chosen from a grid of values in (0.995, 1). Furthermore, to provide a similar model setting to the one we used in our TV-VPARCOR approach, the covariance matrix at the observational level in the DLM formulation for each TV-VAR(m) was also specified following the approach of Triantafyllopoulos (2007).

Figure 4 shows the BLF-scree plots obtained from the PARCOR approach for each of the 50 datasets under the three scenarios for model orders $m = 1, \dots, 5$. We see that in all scenarios the BLF-scree plots indicate that the optimal model order is $P = 2$. We also computed the DIC as explained in the previous section for each model order $m = 1, \dots, 5$ and each dataset under the three scenarios. The bottom right plot in Figure 4 shows the distributions of the optimal model orders chosen by the TV-VPARCOR and TV-VAR approaches for scenario (ii). We see that the TV-VPARCOR and TV-VAR approaches lead to very similar results and model order 2 is adequately chosen as the optimal model order in this scenario for most of the 50 datasets. Similar results were obtained for scenarios (i) and (iii).

Figures 2, 3, and 5 summarize posterior inference obtained from the TV-VPARCOR approach using a model order of 2 for the three scenarios. Estimated spectral densities were obtained from the posterior means of the approximate smoothing distributions of the forward and backward PARCOR coefficient matrices over time. The

DOI: 10.1111/jtsa.12534

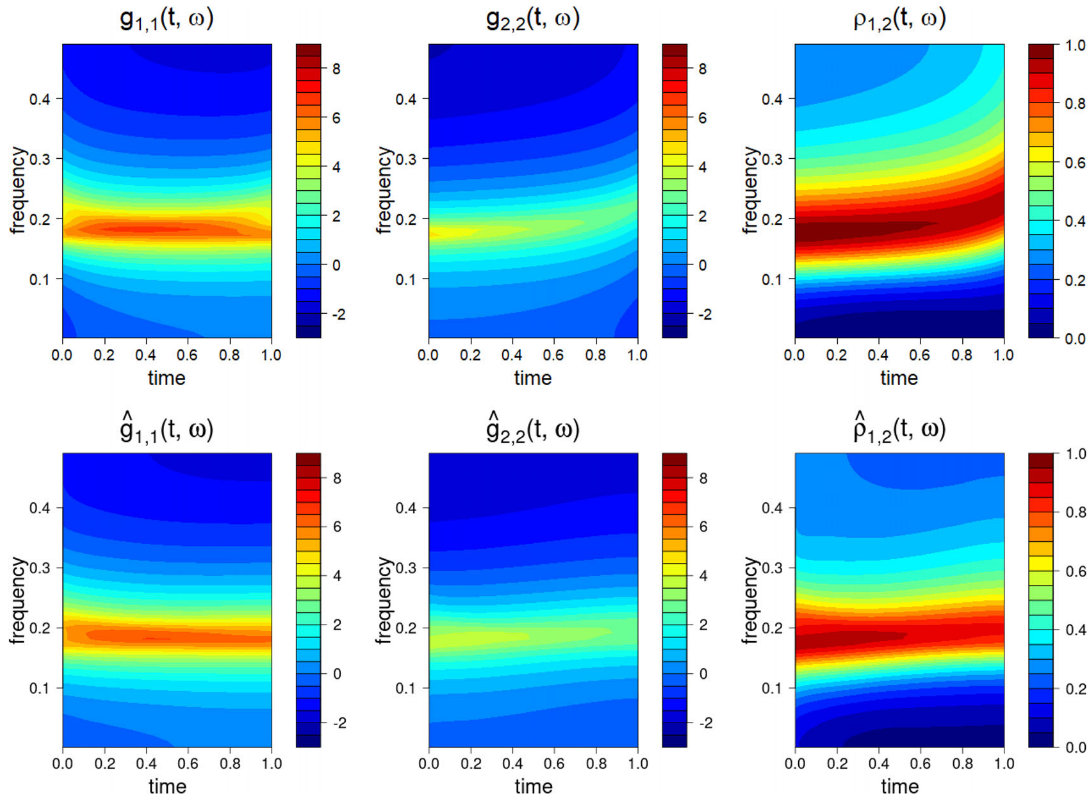

Figure 3. Case with $\phi_{1,1,2,t} = r_{3,t}$ and $\phi_{2,1,2,t} = r_{4,t}$. Top: True log spectral density $g_{11}(t, \omega)$ (left), true log spectral density $g_{22}(t, \omega)$ (middle), true squared coherence $\rho^2_{1,2}(t, \omega)$ (right). Bottom: Estimated $\hat{g}_{11}(t, \omega)$ (left), estimated $\hat{g}_{22}(t, \omega)$ (middle), estimated $\hat{\rho}^2_{1,2}(\omega, t)$ (right) [Color figure can be viewed at wileyonlinelibrary.com]

estimated log spectral densities displayed in the figures were obtained by averaging over the 50 simulated datasets. The bivariate TV-VPARCOR model is able to adequately capture the structure of the individual spectral densities and also that of the squared coherences. From these figures we also see that in scenarios (ii) and (iii) the second series has stronger impact on the first one and therefore their coherence is stronger. The TV-VPARCOR model is able to adapt and adequately capture this feature in the case in which the off-diagonal coefficients in the VAR process are non-zero and constant over time (scenario (ii)), and also when these coefficients are non-zero and time-varying (scenario (iii)).

To compare the performance of the TV-VPARCOR and TV-VAR models in estimating the various time-frequency representations, we computed the mean and standard deviations of the ASE for each of the models in each of the three simulation scenarios. The ASE is defined as follows (Ombao *et al.*, 2001)

$$\mathrm{ASE}_n = (TL)^{-1} \sum_{t=1}^{T} \sum_{l=1}^{L} \left(\log \hat{\boldsymbol{g}}(t, \omega_l) - \log \boldsymbol{g}(t, \omega_l) \right)^2, \tag{18}$$

where $\omega_l = 0, 0.001, 0.011, \ldots, 0.5$. Note that we have $n = 50$ simulated datasets for each of the three scenarios. Table I summarizes the mean and standard deviations of the ASE based on ASE_n for the three scenarios. Note that the simulated data are actually generated from TV-VAR models, not from TV-VPARCOR models, so we expect TV-VAR models to do better in terms of ASE for this specific simulation study. Nevertheless the proposed TV-VPARCOR approach has comparable performance in terms of estimating the time-frequency characteristics

DOI: 10.1111/jtsa.12534

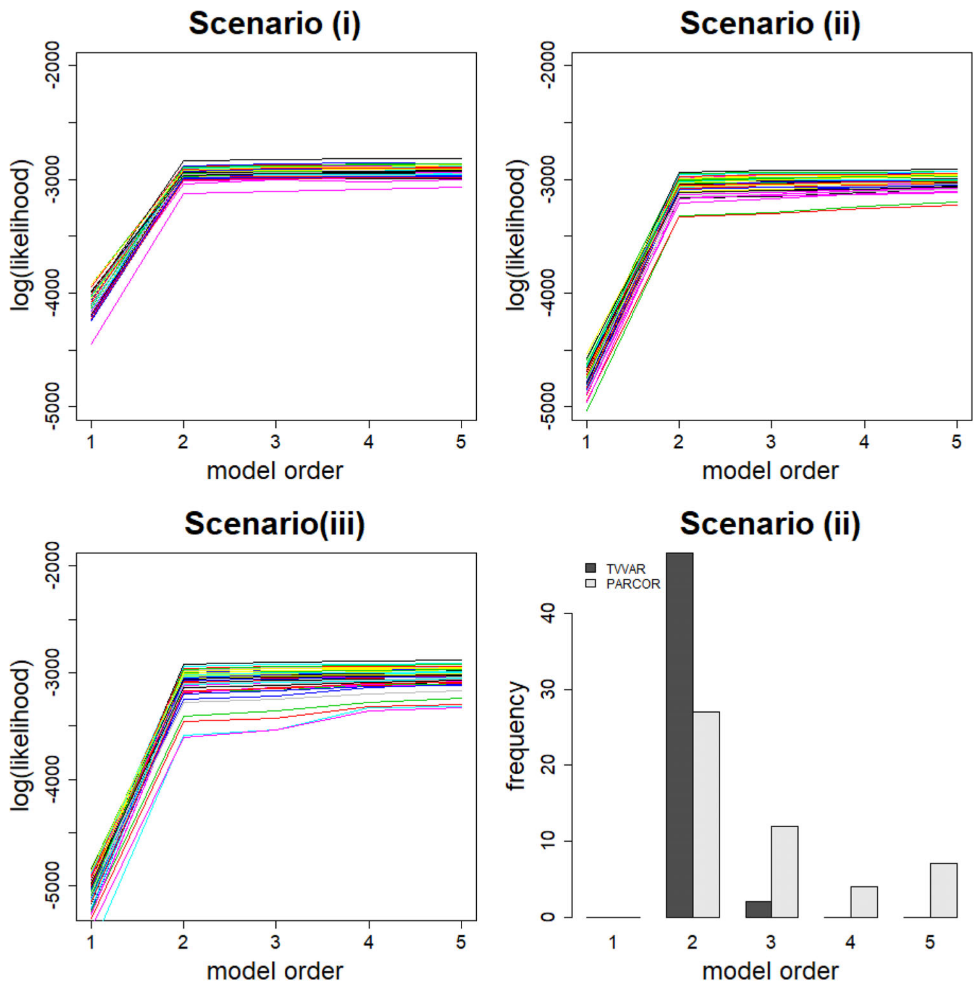

Figure 4. Top: BLF-scree plots of the 50 realizations of the for scenarios (i) and (ii). Bottom: BLF-scree plot for scenario (iii) and optimal model orders for scenario (ii) [Color figure can be viewed at wileyonlinelibrary.com]

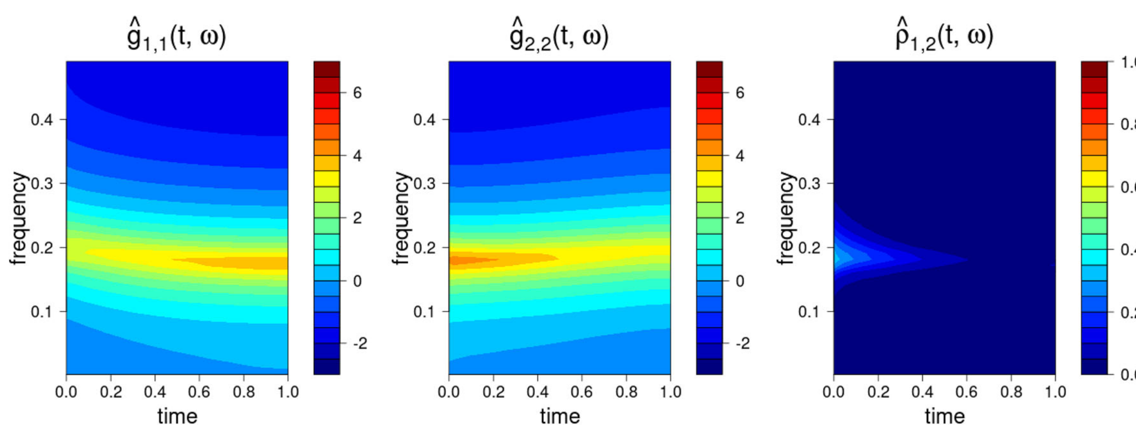

Figure 5. Case with $\phi_{1,1,2,t} = \phi_{2,1,2,t} = 0$. Left: Estimated average log spectral density of the first component. Middle: Estimated average log spectral density of the second component. Right: Estimated average squared coherence [Color figure can be viewed at wileyonlinelibrary.com]

Table I. Mean ASE values and corresponding standard deviations (in parentheses) for the log-spectral densities and log squared coherences obtained from TV-VPARCOR and TV-VAR models of order 2 for the TV-VAR(2) simulated data for $t = 1{:}1024$

	Case (i): $\phi_{1,1,2,t} = \phi_{2,1,2,t} = 0$		
Model	g_{11}	g_{22}	ρ^2_{12}
TV-VPARCOR	0.0246(0.0183)	0.0255(0.0147)	0.0008(0.0006)
TV-VAR	0.0171(0.0068)	0.0186(0.0080)	0.0009(0.0005)
	Case (ii): $\phi_{1,1,2,t} = -0.8, \phi_{2,1,2,t} = 0$		
Model	g_{11}	g_{22}	ρ^2_{12}
TV-VPARCOR	0.0284(0.0118)	0.0238(0.0086)	0.0027(0.0023)
TV-VAR	0.0254(0.0073)	0.0253(0.0081)	0.0023(0.0011)
	Case (iii): $\phi_{1,1,2,t} = r_{3,t}, \phi_{2,1,2,t} = r_{4,t}$		
Model	g_{11}	g_{22}	ρ^2_{12}
TV-VPARCOR	0.1227 (0.0418)	0.3289 (0.0732)	0.0281 (0.0189)
TV-VAR	0.1001 (0.0258)	0.3747 (0.0729)	0.0188 (0.0062)

Table II. Computation times (in seconds) for TV-VPARCOR and TV-VAR models

Model	Case (i)	Case (ii)	Case (iii)
TV-VPARCOR	2.54 seconds	2.48 seconds	2.71 seconds
TV-VAR	8.98 seconds	8.95 seconds	8.36 seconds

of the original process while being computationally more efficient. In fact, Table II presents the computation times for both models averaging over the 50 realizations in each case. We see that even for this example with only two time series components and a model order of 2, the TV-VPARCOR models require almost a quarter of the computation time required by the TV-VAR models. As the model order and the number of time series components increase, differences in computational time will be more pronounced, making the TV-VPARCOR approach more efficient for modeling large temporal datasets.

Finally, Table III shows the MSE values for the 10-steps ahead forecasts ($t = 1025{:}1034$) and corresponding standard deviations for the TV-VPARCOR and the TV-VAR models for the 3 scenarios. The MSE values for both models are comparable, with the TV-VPARCOR MSE being smaller than that for the TV-VAR in case (iii), which corresponds to the case in which some of the off-diagonal parameters are non-zero and varying over time.

Table III. MSE values for the 10-steps ahead forecast ($t = 1025{:}1034$) and corresponding standard deviations (in parentheses) obtained from TV-VPARCOR and TV-VAR models for the TV-VAR(2) simulated data

Model	Case (i)	Case (ii)	Case (iii)
TV-VPARCOR	2.556	5.624	6.378
TV-VAR	2.548	5.408	6.594

DOI: 10.1111/jtsa.12534

3.2. 20-Dimensional TV-VAR(1)

We analyze data simulated from a 20-dimensional non-stationary TV-VAR(1) process with $T = 300$ in which the (i, j) elements of the matrix of VAR coefficients at time t, $\mathbf{\Phi}_t$, are given as follows:

$$\mathbf{\Phi}_t(i,j) = \begin{cases} 0.7 + \frac{0.2}{299} \times t & \text{for all} \quad i = j, \quad i = 1, \dots, 10, \\ -0.95 + \frac{0.2}{299} \times t & \text{for all} \quad i = j, \quad i = 11, \dots, 20, \\ 0.9 & \text{for} \quad (i,j) \in \{(1,5),(2,15)\}, \\ -0.9 & \text{for} \quad (i,j) \in \{(6,12),(15,20)\}, \\ 0 & \text{otherwise.} \end{cases}$$

for $t = 1, \dots, 300$. In addition, we assume $\mathbf{\Sigma} = 0.1\boldsymbol{I}_{20}$.

We fit TV-VPARCOR models considering $P_{\max} = 3$. Note that the PARCOR approach with $P_{\max} = 3$ requires fitting 6 multivariate DLMs with state-space parameter vectors of dimension 400. Alternatively, working directly with TV-VAR representations with $P_{\max} = 3$ requires fitting 3 multivariate DLMs with state-space parameter vectors of dimension 400 for model order 1, 800 for model order 2, and 1200 for model order 3. The TV-VAR model representation leads to a rapid increase of the dimension of the state-space vector with the model order, which significantly reduces the computational efficiency, particularly for large and even moderate T. The TV-VPARCOR approach requires fitting more multivariate DLMs, but the dimensionality of the state-space vectors remains constant with the model order. This is an important advantage of the TV-VPARCOR approach. In fact, the TV-VPARCOR model required 585 seconds of computation time for $P_{\max} = 3$, while the TV-VAR model required 3379 seconds with the same $P_{\max} = 3$ value. Posterior computations were completed in both cases using a MacBookPro13 with Intel Core i5, with 2 GHz (1 Processor). Note also that, for a given model order the PARCOR approach can be further optimized in terms of computational efficiency, as the forward and backward DLMs can run in parallel.

We assumed prior hyperparameters $\boldsymbol{m}_{0,\cdot,m} = \mathbf{0}$ and $\boldsymbol{C}_{0,\cdot,m} = \boldsymbol{I}_{400}$ for the forward and backward PARCOR models. The elements of the diagonal component of discount factor matrices, $\delta_{f,m}$ and $\delta_{b,m}$, were chosen from a grid of values in $(0.99, 1)$. As mentioned above we also fit TV-VAR models with model orders going from 1 to 3 using similar prior hyperparameters and discount factors. For both types of models the DIC picked model order 1 as the optimal model order, which is the corresponding true model order in this case. Both types of models led to similar posterior inference of the time-frequency spectra.

Here we only show the results from the TV-VPARCOR approach. Figure 6 shows the true and estimated log spectral densities from the TV-VPARCOR model for 4 components of the 20-dimensional time series, namely, components 1, 2, 8, and 15. Figure 7 shows the true and estimated coherences between components 1 and 5, components 2 and 15, components 5 and 12, and components 15 and 20. Overall we see that the TV-VPARCOR approach adequately captures the space–time characteristics of the original multivariate non-stationary time series process. Furthermore, the TV-VPARCOR approach led to similar posterior estimates of the VAR coefficients over time to those obtained from using a DLM representation of a TV-VAR (see Figure 1 in the Supporting information).

3.3. Additional Simulation Studies

Here we consider two additional simulation studies with higher model orders to highlight the performance of the TV-VPARCOR in multivariate cases that require a much larger number of parameters.

We first evaluate the impact on model performance in terms of the number of time series for models with model order $P = 10$. We simulated data from multivariate non-stationary TV-VAR(10) models with a number of series increasing from 2 to 5. We simulated the 2-dimensional time series as follows. We took 2 of the EEG channels analyzed in Section 4.1 and fitted a 2-dimensional TV-VAR(10) to such series. We then simulated a 2-dimensional dataset using the estimated TV-VAR(10) parameters for these EEG series. Similarly, we then generated a 3-dimensional time series dataset by using the estimated parameters obtained from fitting a TV-VAR(10)

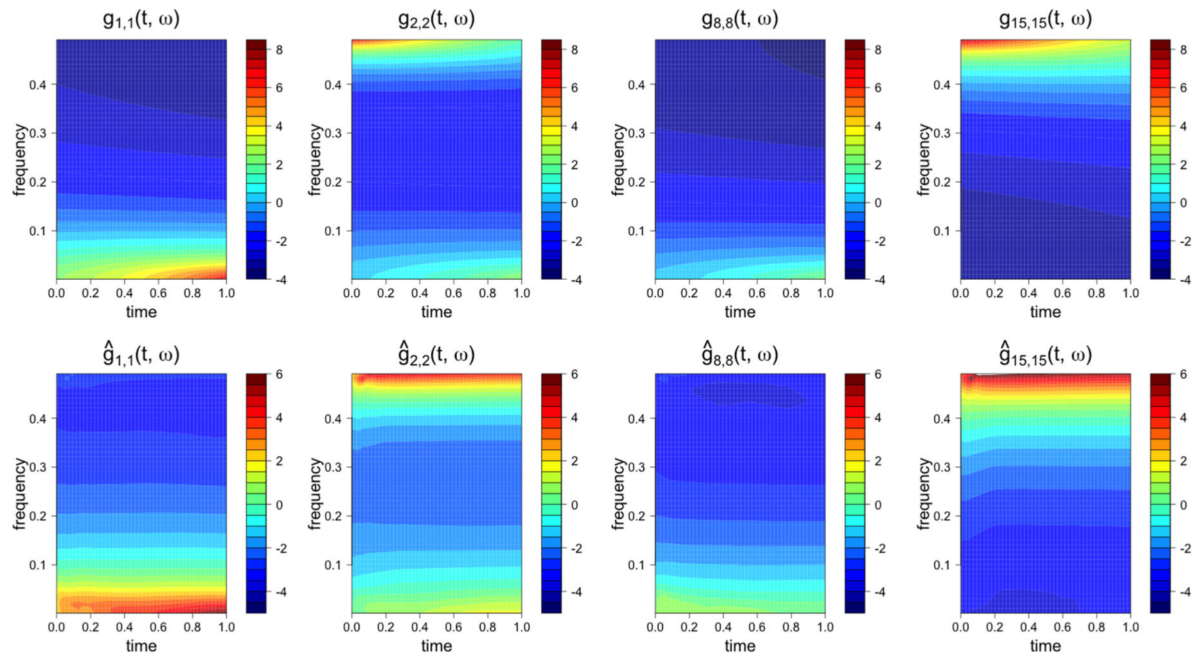

Figure 6. Top: True log spectral densities of time series components 1, 2, 8, and 15. Bottom: estimated log spectral densities of the same components obtained from the PARCOR approach with model order 1 [Color figure can be viewed at wileyonlinelibrary.com]

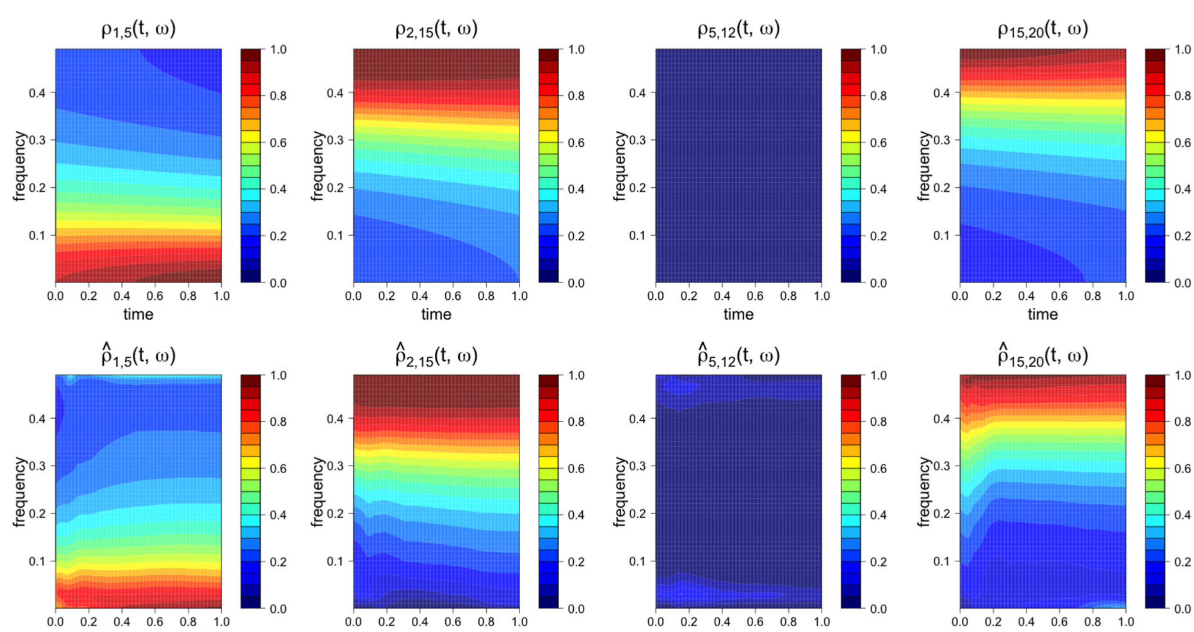

Figure 7. Top: True coherence between components 1 and 5, 2 and 15, 5 and 12, and 15 and 20. Bottom: Corresponding estimated coherences obtained from the PARCOR model [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12534

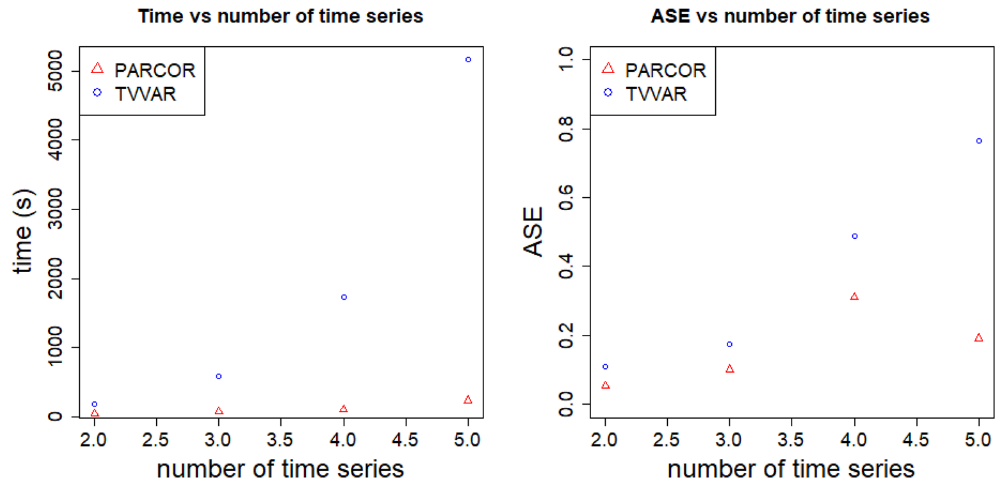

Figure 8. Left plot: Running time against number of time series. Right plot: estimated ASE against number of time series [Color figure can be viewed at wileyonlinelibrary.com]

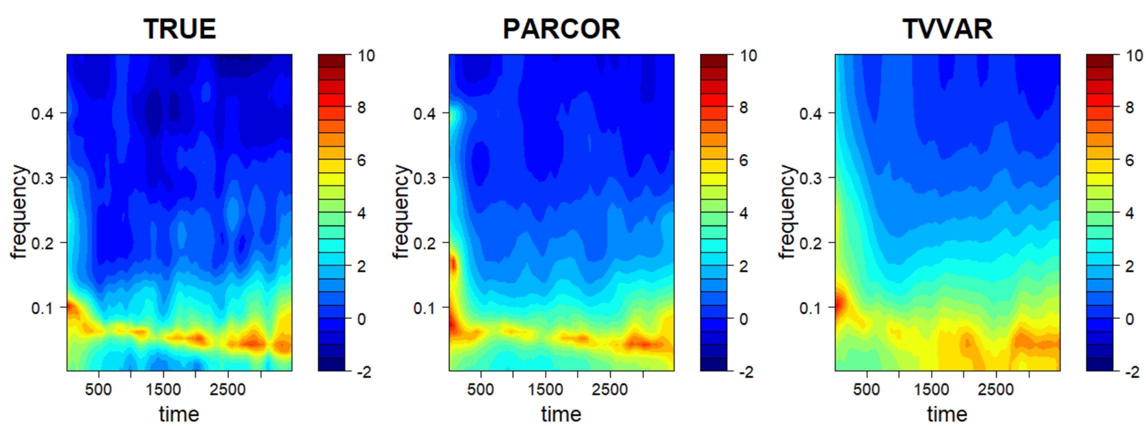

Figure 9. Left: True spectral density of the first time series in the first simulation. Middle: PARCOR estimated spectral density of the first time series. Right: TV-VAR estimated spectral density of the first time series [Color figure can be viewed at wileyonlinelibrary.com]

to 3 EEG channels (including the previous 2 channels). We repeated this procedure to obtain 4-dimensional and 5-dimensional datasets, adding one EEG time series at the time. We then fit TV-VAR(10) and TV-VPARCOR(10) to the 4 simulated datasets of dimensions 2, 3, 4, and 5. Figure 8 compares the performance of the two approaches in terms of the running time and the ASE as the number of time series increases. We see that in both cases the TV-VPARCOR approach leads to much smaller running times and also smaller ASE values as the number of time series increases. Figure 9 shows the true and estimated spectral density estimates obtained from the TV-VAR(10) and TV-VPARCOR(10) for the first time series component obtained from the 5-dimensional models that considered 5 channels. We see that the TV-VPARCOR leads to more accurate estimates of the spectral density.

We then consider another simulated scenario to evaluate the performance of the TV-VPARCOR approach in terms of the model order. For this we simulated data from 6-dimensional TV-VAR models with model orders ranging from 1 to 10. Again we used the EEG data to simulate these data by first fitting TV-VAR models of orders 1–10 to the EEG data and then using the estimated parameters from these models to simulate the data. Figure 10 shows a graph of the ASE values obtained from fitting TV-VAR and TV-VPARCOR models to the different datasets simulated under different models orders. The plot shows that the TV-VPARCOR approach leads to lower ASE

DOI: 10.1111/jtsa.12534

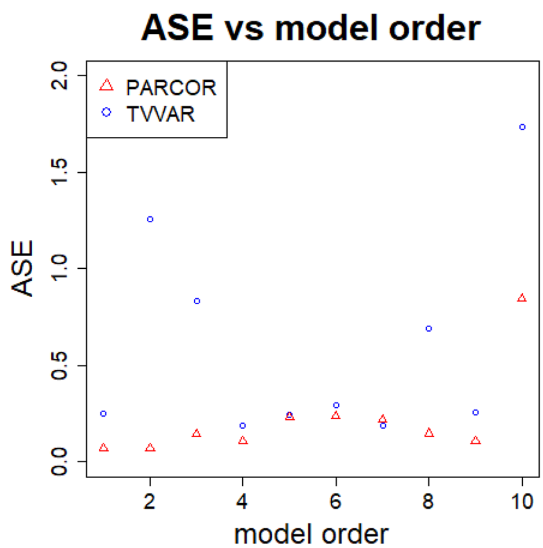

Figure 10. ASE values against model order [Color figure can be viewed at wileyonlinelibrary.com]

values, or comparable values, to those obtained from the TV-VAR models for all the model orders. Note that ASE values for different model orders are not comparable, as they are based on different datasets.

Finally, we also considered a study in which the data were simulated from a piecewise time series process. Once again the proposed TV-VPARCOR approach outperformed the TV-VAR approach. The results of this study are included in the Supporting information.

4. CASE STUDIES

4.1. Analysis of Multi-Channel EEG Data

We analyze multi-channel EEG data recorded on a patient that received electroconvulsive therapy (ECT) as a treatment for major depression. These data are part of a larger dataset, code named `Ictal19`, that corresponds to recordings of 19 EEG channels from one subject during ECT. As an illustration, we use our multivariate TV-VPARCOR model to analyze 9 of the channels, specifically channels $F_3, F_z, F_4, C_3, C_z, C_4, P_3, P_z, P_4$ shown in Figure 11. We chose these channels because they are closely located and because based on previous analyses we expect strong similarities in their temporal structure over time. The full multi-channel dataset was analyzed in West *et al.* (1999) and Prado *et al.* (2001) using univariate TVARs separately for each channel, and also using dynamic regression models. The original recordings of about 26,000 observations per channel were subsampled every sixth observation from the highest amplitude portion of the seizure, leading to a set of series of 3600 observations (corresponding to 83.72 seconds) per channel (Prado *et al.*, 2001).

We analyzed the $K = 9$ series listed above jointly using a multivariate TV-VPARCOR model. We considered a maximum model order of $P_{\max} = 20$ and discount factor values on a grid in the (0.99, 1] range (with equal spacing of 0.001). We further assumed that the discount factor values were the same across channels. This assumption was based on previous analyses of the individual channels using univariate TVAR models that showed similar optimal discount values for the different channels. We set $n_{0,f,m} = n_{0,b,m} = 1$, and $S_{0,f,m} = S_{0,b,m} = 2000\boldsymbol{I}_9$ for all m. In addition, we set the same initial prior parameters $\boldsymbol{m}_{0,f,m} = \boldsymbol{m}_{0,b,m} = \boldsymbol{0}$ and $\boldsymbol{C}_{0,f,m} = \boldsymbol{C}_{0,b,m} = 1000\boldsymbol{I}_{81}$. The computation time to run the search for the optimal model with $P_{\max} = 20$ in this dataset was 1142 seconds in an Inter(R) Xeon(R) server with CPU E5-4650 with 2 cores and 2.70 GHz. The optimal model order was found to be 5 (see Figure 4 in the Supporting information) and so, the results presented here correspond to a TV-VPARCOR

DOI: 10.1111/jtsa.12534

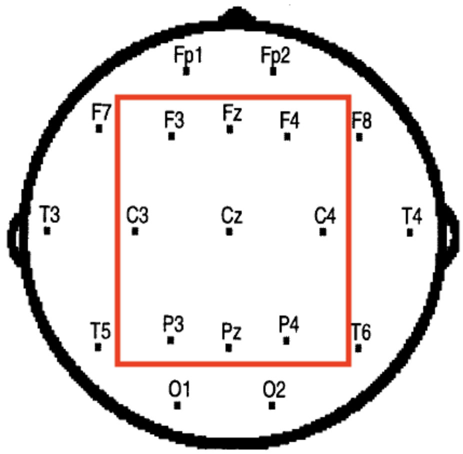

Figure 11. Representation of the `lctal19` electrode placement. Here we focus on the nine channels in the region highlighted [Color figure can be viewed at wileyonlinelibrary.com]

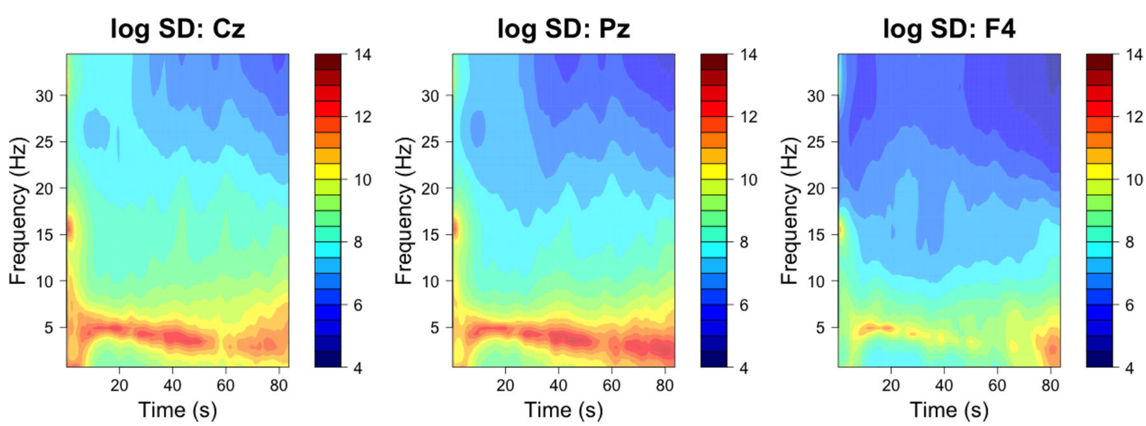

Figure 12. Estimated log-spectral densities for channels Cz, Pz, and F4 [Color figure can be viewed at wileyonlinelibrary.com]

model with this order. Higher-order models were also fitted leading to similar but slightly smoother results in terms of the estimated spectral density, coherence, and partial coherence.

Figure 12 displays estimated log spectral densities of channels Cz, Pz, and F4. We note that the multi-channel EEG data are dominated by frequency components in the lower frequency band (below 18 Hz). Furthermore, each EEG channel shows a decrease in the dominant frequency over time, starting around 5 Hz and ending around approximately 3 Hz. This decrease in the dominant frequency was also found in West *et al.* (1999). Channels Cz and Pz are more similar to each other than to channel F4 in terms of their log-spectral densities. The three channels show the largest power around the same frequencies; however, channel F4 displays smaller values in the power log-spectra than those for channels Cz and Pz. The remaining channels also show similarities in their spectral content (not shown).

Figure 13 shows estimated squared coherences (top) and estimated squared partial coherences (bottom) between channels Pz and Cz, F4 and Cz, and F4 and Pz. Channels Pz and Cz show a very strong coherence over time across almost all the frequency bands under 35 Hz. On the other hand, channel F4 shows strong coherence with channels Pz and Cz across frequencies below 15–18 Hz at the beginning of the seizure. After the initial 10—15 seconds, and approximately until about 50 seconds, there is a strong coherence between F4 and Pz and Cz only at the dominant frequency of 3–5 Hz that dissipates towards the end of the seizure. The partial coherence across pairs of

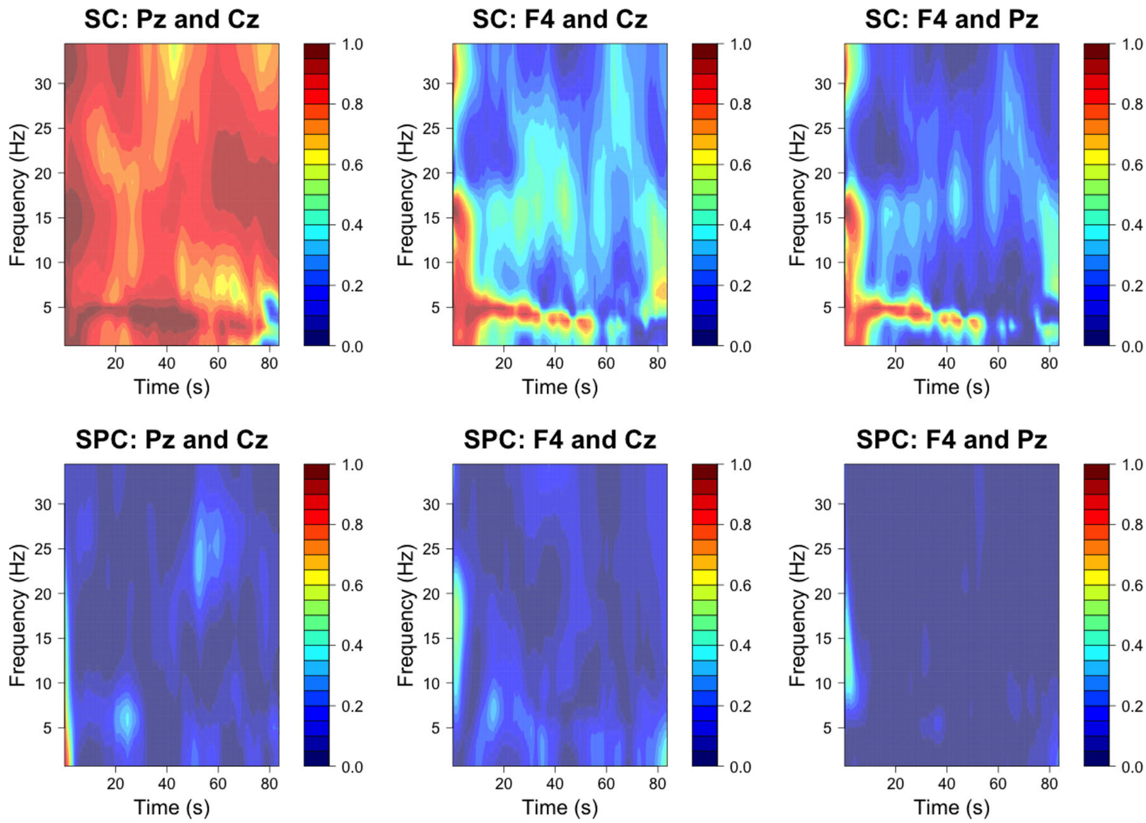

Figure 13. Top plots: Squared coherence between Pz and Cz, F4 and Cz, and F4 and Pz respectively. Bottom plots: Squared partial coherence between Pz and Cz, F4 and Cz, and F4 and Pz [Color figure can be viewed at wileyonlinelibrary.com]

channels is the frequency domain version of the squared correlation coefficient between relationship between pairs of components after the removal of the effects of all the other components. Figure 13 shows that the estimated squared partial coherences between Pz and Cz, F4 and Cz, and F4 and Pz are essentially negligible for most frequency bands over the seizure course. This makes sense due to the fact that most of the 9 EEG channels are so strongly coherent across different frequency bands over the entire period of recording. The estimated squared partial coherence between channels Pz and Cz is large for frequencies below 5 Hz only at the very beginning of the seizure. These findings are consistent with results from the analysis of these data in West *et al.* (1999) and Prado *et al.* (2001).

Finally, we also estimated DTFs and PDC between channels as explained in Section 2.4. Figure 14 shows the estimated time-varying PDC among channels Pz, Cz, and F4. Channel Pz is located in the parietal region, channel Cz is a central channel, and F4 is a frontal right channel. From the PDC and DTF (not shown) results we see that channel Pz has the largest directed and cascade flow towards channels Cz and F4. There is also some PDC activity flow between channels Cz and Pz.

4.2. Analysis of Multi-Location Wind Data

We analyze wind component data derived from median wind speed and direction measurements taken every 4 hours from 1 June 2010 to 15 August in 3 stations in Northern California. These data were obtained from the Iowa Environmental Mesonet Automated Surface Observing System (ASOS) Network, a publicly available database

DOI: 10.1111/jtsa.12534

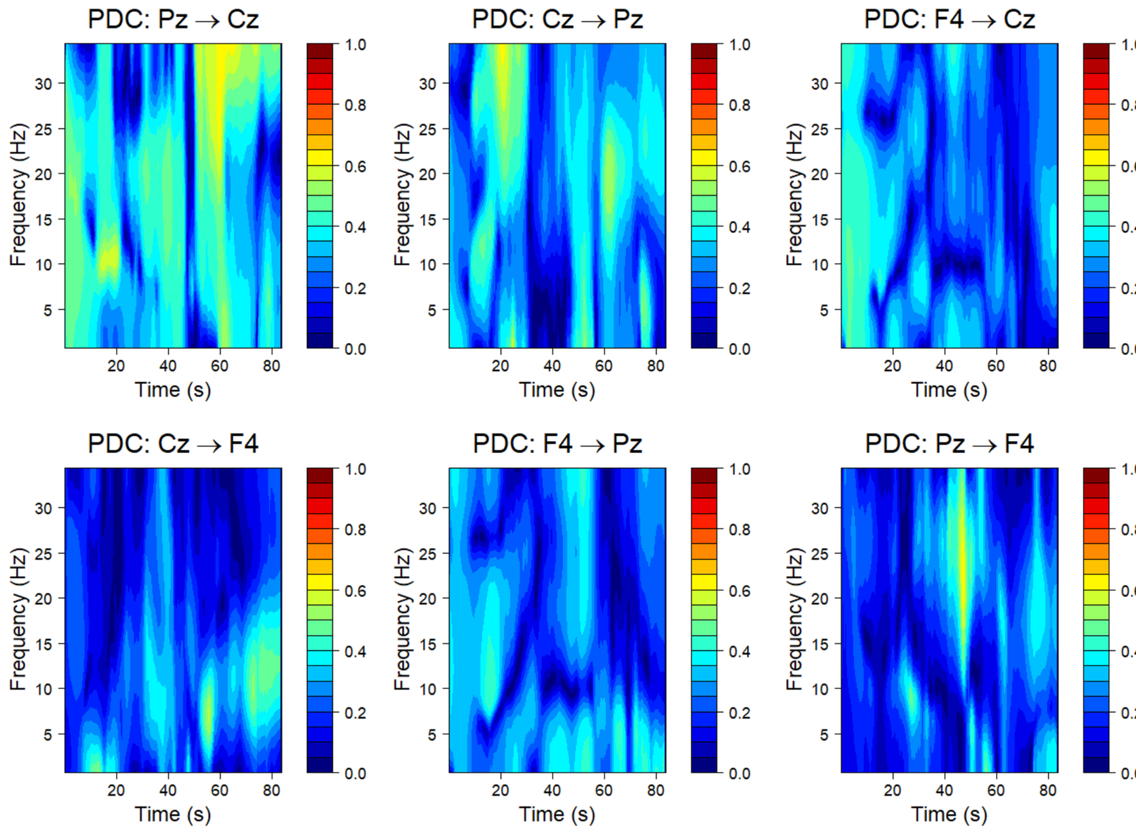

Figure 14. Estimated partial directed coherence among channels Pz, Cz, and F4 [Color figure can be viewed at wileyonlinelibrary.com]

(see `http://mesonet.agron.iastate.edu/ASOS/`). ASOS stations are located at airports and take observations and basic reports from the National Weather Service, the Federal Aviation Administration, and the Department of Defense. For additional information about the ASOS measurements see NOAA (1998). Here we analyze time series data from Monterey, Salinas, and Watsonville, 3 stations located near the Monterey Bay.

We use the TV-VPARCOR approach for joint analysis of the six-dimensional time series corresponding to the wind time series components for the 3 stations. We set $P_{\max} = 10$ and consider discount factor values on a grid in the $(0.9, 1]$ range. We assume that discount factor values were the same across components for the 3 stations. We set the prior hyperparameters as follows: $n_{0,f,m} = n_{0,b,m} = 1$, and $S_{0,f,m} = S_{0,b,m} = 5\boldsymbol{I}_6$, $\boldsymbol{m}_{0,f,m} = \boldsymbol{m}_{0,b,m} = \boldsymbol{0}$ and $\boldsymbol{C}_{0,f,m} = \boldsymbol{C}_{0,b,m} = 10\boldsymbol{I}_{36}$ for all m. The computation time to run the search for the optimal model with $P_{\max} = 10$ in this dataset was 35.72 seconds in an Inter(R) Xeon(R) server with CPU E5-4650 with 2 cores and 2.70 GHz. The optimal model order chosen by the approximate DIC calculation is $P = 3$ (see Figure 5 in the Supporting information). For this model order we found that the optimal discount factors were 0.97, 0.97, and 0.99, respectively, for each of the 3 levels of the forward PARCOR model, and 0.98, 0.98, and 0.99 for each of the 3 levels of the backward PARCOR model.

Figure 15 shows the estimated log spectral densities of the east–west component (X component) and the north–south component (Y component) for each location. We can observe that there is a dominant quasi-periodic behavior around the 24-hour period for the east–west (X) components in Monterey and Salinas, as well as the north–south (Y) component in Watsonville. This quasi-periodic behavior is also present, although is less persistent over time, in the east–west component in Watsonville and the north–south components in Monterey and Salinas. The observed quasi-periodic pattern observed in the estimated log-spectral for these three locations is consistent

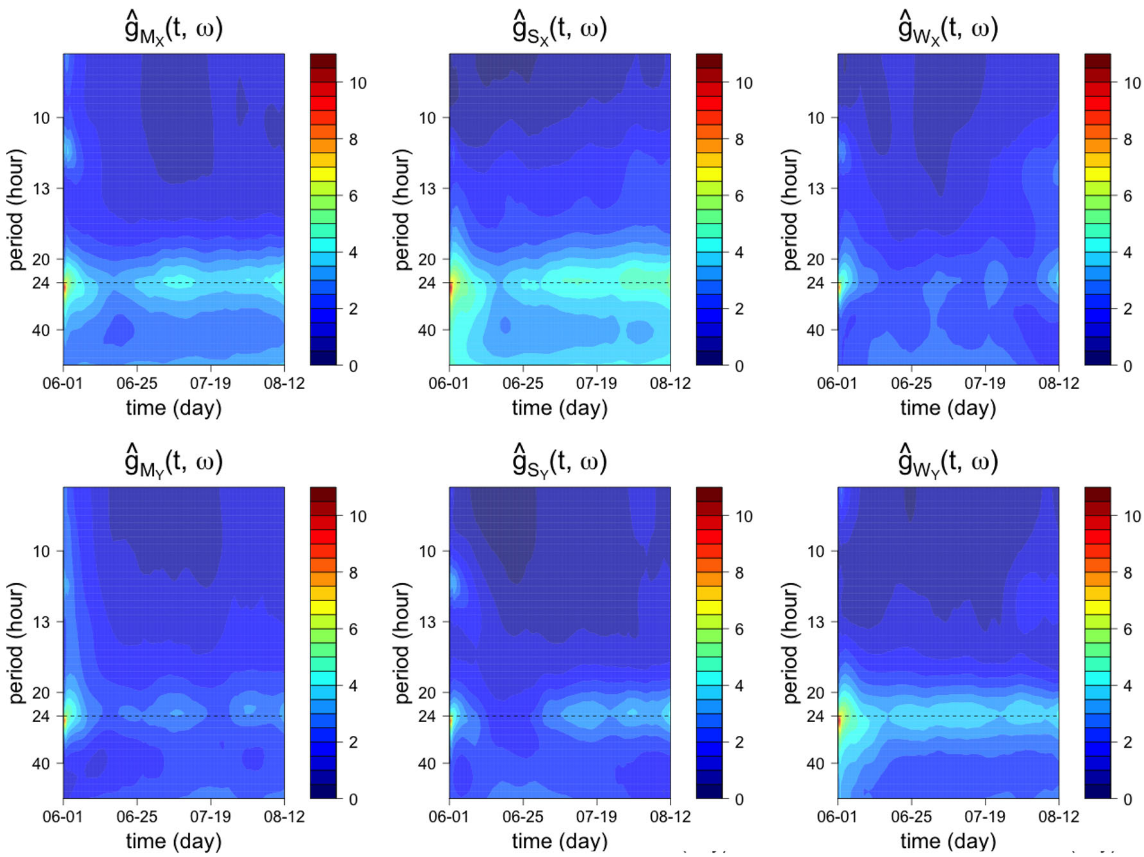

Figure 15. Top row: Estimated log-spectral densities of the east–west (X) components for Monterey, Salinas, and Watsonville. Middle row: Estimated log-spectral densities of the north–south (Y) components for Monterey, Salinas, and Watsonville [Color figure can be viewed at wileyonlinelibrary.com]

with the fact that stronger winds are usually observed in the afternoons/evenings during the summer in these locations, while calmer winds are observed during the rest of the day. Note also that the quasi-periodic daily behavior is more persistent over the entire set of summer months for the north–south component than the east–west component in Watsonville, while the quasi-periodic behavior is more persistent in the east–west component than in the north–south component in Monterey and Salinas.

Approximate uncertainty quantification for the spectral density matrix estimates, or any functions of this matrix, can also be obtained by sampling from the approximate posterior distributions of the TV-VPARCOR model parameters as illustrated in Figure 16. The figure provides approximate 95% posterior bounds for the log-spectral density of the north–south wind component in Watsonville. The dominant quasi-periodic behavior around the 24 hours period also appears in the lower and upper uncertainty bands, indicating that there is less uncertainty around this frequency band than around, say, higher periods (low-frequency) bands that display a much larger uncertainty. There is also very low power estimated at relatively low periods (higher frequencies) of 14 hours and below and these estimates also show very low uncertainty.

Figure 17 shows the estimated squared coherences between each pair of wind components across the three locations. There is a very strong coherence between Monterey and Salinas in the east–west (X) components for periods above 15 hours, with the strongest relationship observed around 24 hours. We also observe that in general, there is a strong coherence between all the components around the 24 hours period. This coherence relationship tends to be more marked across some locations during the month of June (e.g., between the north–south components

DOI: 10.1111/jtsa.12534

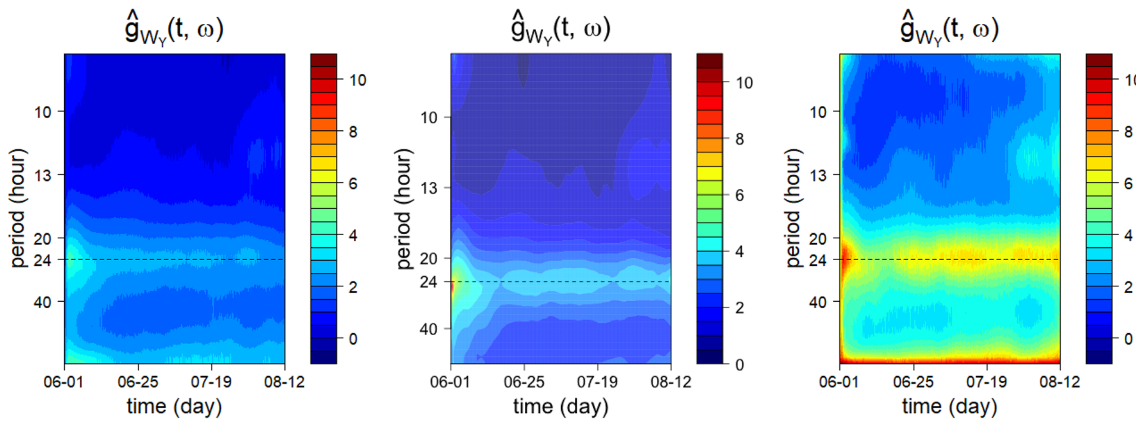

Figure 16. Left: Lower bound of a 95% posterior interval of the log-spectral density of the north–south (Y) component in Watsonville. Middle: Estimated mean log-spectral density of the north–south (Y) component in Watsonville. Right: Upper bound of 95% posterior interval of the log-spectral density of the north–south (Y) component in Watsonville [Color figure can be viewed at wileyonlinelibrary.com]

of Monterey and Salinas). Furthermore, the estimated squared partial coherence (see Figure 6 in the Supporting information) between the east–west components of Monterey and Salinas also shows that there is a relatively large linear relationship between these components for periods above 13 hours even after removing of the effect of all the other components for these locations and also after removing the effect of the wind components in Watsonville.

The TV-VPARCOR model can also be used for forecasting as described in Section 2.5. Figure 18 shows 72 hours forecasts obtained from the TV-VPARCOR model for the north–south wind component in Monterey. We see that the model adequately captures the general future behavior of this time series component.

5. DISCUSSION

We present a computationally efficient approach for analysis and forecasting of non-stationary multivariate time series. We propose a multivariate dynamic linear modeling framework to describe the evolution of the PARCOR coefficients of a multivariate time series process over time. We use approximations in this multivariate TV-VPARCOR setting to obtain computationally efficient and stable inference and forecasting in the time and time-frequency domains. The approximate posterior distributions derived from our approach are all of standard form. We also provide a method to choose the optimal number of stages in the TV-VPARCOR model based on an approximate DIC calculation. In addition, our model can provide reliable short term forecasting.

The proposed framework provides computational efficiency and excellent performance in terms of the ASE between the true and estimated time-varying spectral densities as shown in extensive simulation studies and in the analysis of two multivariate time series datasets. The TV-VPARCOR model representations also lead to very significant reduction in computational time when compared to TV-VAR model representations, particularly for cases in which we have model orders larger than 2–3 and more than a handful of time series components.

In addition to simulation studies we have shown that the TV-VPARCOR approach can be successfully used to analyze real multivariate non-stationary time series data. We presented the analysis of non-stationary multi-channel EEG data and also the analysis and forecasting of multi-location wind data. In the EEG case, our model was able to adequately detect the main time-frequency characteristics of individual EEG channels as well as the relationships across multiple channels over time. For the multi-location wind component data, our model detected a quasi-periodic pattern through the estimated spectral densities of each time series component which is consistent with the expected behavior of these components during the summer for locations near the Monterey Bay area. The model was also able to describe the time-varying relationships across multiple components and locations and led to reasonable short term forecasting.

DOI: 10.1111/jtsa.12534

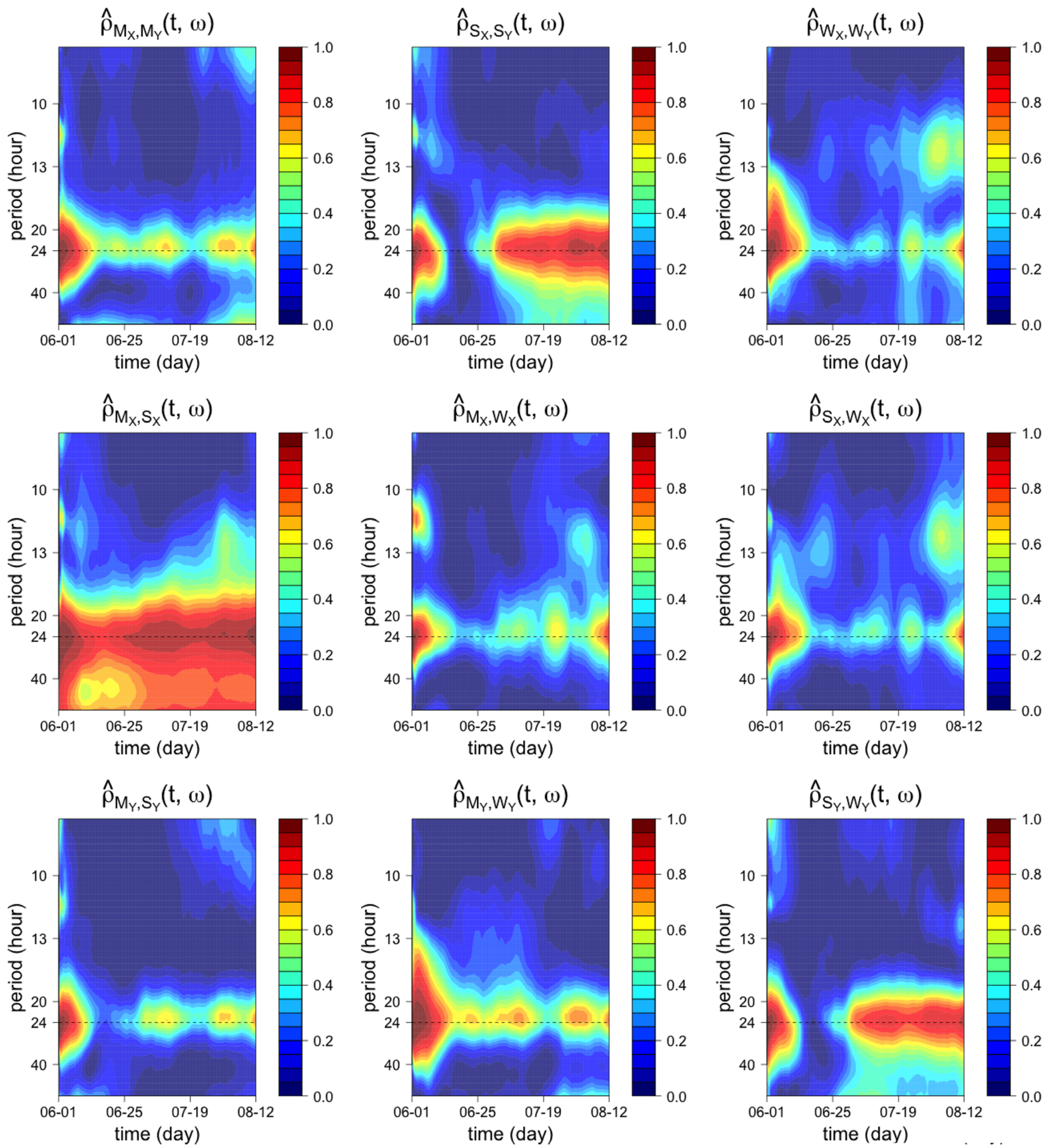

Figure 17. Top row: Estimated squared coherences between the east–west (X) component and north–south (Y) component in Monterey, Salinas, and Watsonville. Middle row: Estimated squared coherences between the east–west (X) components of Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville. Bottom row: Estimated squared coherences between the north–south components in Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville [Color figure can be viewed at wileyonlinelibrary.com]

The proposed dynamic multivariate TV-VPARCOR approach is computationally efficient when compared to state-space representations TV-VAR models. However, in many practical settings we may expect sparsity in the model parameters or situations in which some parameters change over time and others do not. Future work will

DOI: 10.1111/jtsa.12534

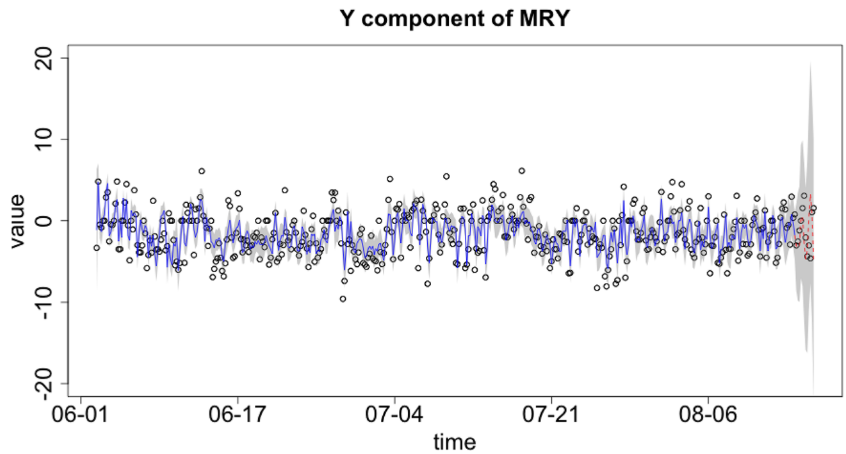

Figure 18. Observed Monterey north–south wind component (dots); smoothed estimates obtained from the posterior mean values of the TV-VPARCOR model (solid red line) and corresponding 90% bands (gray shade); 72 hours forecast (dotted red line) and corresponding 90% bands (gray shade) [Color figure can be viewed at wileyonlinelibrary.com]

explore inducing, possibly time-varying, sparsity and dimension reduction in these multivariate TV-VPARCOR models while maintaining computational efficiency and accuracy in inference and forecasting.

ACKNOWLEDGMENT

The authors were partially supported by National Science Foundation award SES-1853210.

DATA AVAILABILITY STATEMENT

The wind component data analyzed in this article are publicly available from the website cited in the article. The multi-channel EEG data and the simulated data are available upon request from the authors. No newly collected data have been analyzed in this article.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Ameen JR, Harrison PJ. 1985. Normal discount Bayesian models. In *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting, September 6/10, 1983*, Vol. 2, Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds.). North Holland; 271.

Astolfi L, Cincotti F, Mattia D, De Vico Fallani F, Tocci A, Colosimo A, Salinari S, Marciani MG, Hesse W, Witte H, Ursino M, Zavaglia M, Babiloni F. 2008. Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators. *IEEE Transactions on Biomedical Engineering* **55**: 902–913.

Baccala LA, Sameshima K. 2001. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics* **84**: 463–474.

Blinowska KJ. 2011. Review of the methods of determination of directed connectivity from multichannel data. *Medical and Biological Engineering and Computing* **49**: 521–529.

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

Bruce SA, Hall MH, Buysse DJ, Krafty RT. 2018. Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series. *Biometrics* **74**: 260–269.

Cadonna A, Kottas A, Prado R. 2019. Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association* **114**: 1838-1853.

DOI: 10.1111/jtsa.12534

Cheung BLP, Riedner BA, Tononi G, Van Veen BD. 2010. Estimation of cortical connectivity from EEG using state-space models. *IEEE Transactions on Biomedical Engineering* **57**: 2122–2134.
Chiang S, Guindani M, Yeh HJ, Haneef Z, Stern JM, Vannucci M. 2017. Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping* **38**: 1311–1332.
Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014. *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: CRC Press.
Kitagawa G. 2010. *Introduction to Time Series Modeling*. Boca Raton, FL: Chapman & Hall.
Kuś R, Kamiński M, Blinowska KJ. 2004. Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE Transactions of Biomedical Engineering* **51**: 1501–1510.
Li Z, Krafty RT. 2018. Adaptive Bayesian time-frequency analysis of multivariate time series. *Journal of the American Statistical Association* **114**: 453–465, 10.1080/01621459.2017.1415908.
Milde T, Leistritz L, Astolfi L, Miltner WHR, Weiss T, Babiloni F, Witte H. 2009. A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application analysis of laser-evoked brain potentials. *NeuroImage* **50**: 960–969.
Nakajima J, West M. 2017. Dynamics and sparsity in latent threshold factor models: a study in multivariate EEG signal processing. *Brazilian Journal of Probability and Statistics* **31**: 701–731.
NOAA, DoD, FAA, USNavy. 1998. *Automated Surface Observing System (ASOS) User's Guide*.
Ombao H, Raz JA, von Sachs R, Malow BA. 2001. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association* **96**: 543–560.
Omidvarnia A, Azemi G, Boashash B, O'Toole JM, Colditz PB, Vanhatalo S. 2014. Measuring time-varying information flow in scalp EEG signals: orthogonalized partial directed coherence. *IEEE Transactions on Biomedical Engineering* **61**: 680–693.
Prado R, West M. 2010. *Time Series: Modeling, Computation, and Inference*. Taylor & Francis: Chapman & Hall/CRC Texts in Statistical Science. ISBN: 9781420093360.
Prado R, West M, Krystal AD. 2001. Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag-lead structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**: 95–109, 10.1111/1467-9876.00222. ISSN: 1467-9876.
Schmidt C, Pester B, Schmid-Hertel N, Wittle H, Wismuller A, Leistritz L. 2016. A multivariate Granger causality concept towards full brain functional connectivity. *PLOS ONE* **11**: 1–25.
Shumway RH, Stoffer DS. 2017. Time series analysis and its applications: with R examples. In *Springer Texts in statistics* New York: Springer. ISBN: 9780387293172.
Ting C, Ombao H, Balqis Samdin S, Hussain Salleh S. 2017. Estimating dynamic connectivity using regime-switching factor models. *IEEE Transactions on Medical Imaging* **37**: 1011–1023.
Triantafyllopoulos K. 2007. Covariance estimation for multivariate conditionally Gaussian dynamic linear models. *Journal of Forecasting* **26**: 551–569.
Tsay R. 2013. Multivariate time series analysis: with R and financial applications. In *Wiley Series in Probability and Statistics*.
West M, Harrison J. Bayesian forecasting and dynamic models. In *Springer Series in Statistics* New York: Springer, 1997. ISBN: 9780387947259.
West M, Prado R, Krystal AD. 1999. Evaluation and comparison of EEG traces: latent structure in nonstationary time series. *Journal of the American Statistical Association* **94**: 375–387.
Yang WH, Holan SH, Wikle CK. 2016. Bayesian lattice filters for time-varying autoregression and time–frequency analysis. *Bayesian Analysis* **11**: 977–1003.
Yu Z, Prado R, Burke Quinlan E, Cramer SC, Ombao H. 2016. Understanding the impact of stroke on brain motor function: a hierarchical Bayesian approach. *Journal of the American Statistical Association* **111**: 549–563, 10.1080/01621459.2015.1133425.
Zhang Z. 2017. *Multivariate Time Series Analysis in Climate and Environmental Research*. New York: Springer.
Zhou G. 1992. Algorithms for estimation of possibly nonstationary vector time series. *Journal of Time Series Analysis* **13**: 171–188.

DOI: 10.1111/jtsa.12534