Title

Efficient Analysis of Multiple and Multivariate Non-stationary Time Series in the Partial Autocorrelation Domain

Author

Zhao, Wenjie

Publication Date

2022

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

EFFICIENT ANALYSIS OF MULTIPLE AND MULTIVARIATE NON-STATIONARY TIME SERIES IN THE PARTIAL AUTOCORRELATION DOMAIN

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

Wenjie Zhao

March 2022

The Dissertation of Wenjie Zhao
is approved:

Raquel Prado, Chair

Abel Rodriguez

Bruno Sansó

Peter F. Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Efficient Analysis of Multiple and Multivariate Non-stationary Time Series in the Partial Autocorrelation Domain

by

Wenjie Zhao

Recent advances in information technology have made high-dimensional non-stationary signals increasingly common in many areas. We develop a suite of models and computationally fast methods for analysis and forecasting of multiple and multivariate non-stationary time series. These approaches are based on dynamic model representations in the partial autocorrelation domain.

Chapter 1 introduces some background and discusses the limitations of current models and methods for analyzing high-dimensional non-stationary time series. In order to obtain fast and accurate modeling and inference such high-dimensional dynamic settings, a system of Bayesian lattice filtering and smoothing approaches in the PARCOR domain are proposed in this thesis. This PARCOR framework leads to lower dimensional representations, and consequently computationally faster inference, than those required by models in the time and/or frequency domains, such as state-space representations of time-varying autoregressive and vector autoregressive models, which are commonly used in practice.

Chapter 2 proposes an efficient hierarchical dynamic PARCOR model to describe the time-varying behavior of multiple time series, and develops procedures to infer the latent structure underlying multiple non-stationary time series. The performance of the proposed models and methods is tested in the context of analyzing multiple brain signals recorded simultaneously during specific experimental settings or clinical studies. The proposed approach improves the efficiency in ob-

taining posterior summaries of the time-frequency characteristics of the multiple time series, as well as those summarizing their common underlying structure.

Chapter 3 proposes a set of multivariate dynamic linear models (MDLMs) on the forward and backward predictions errors in the PARCOR domain along with computationally efficient methods for filtering and smoothing methods within this modeling class. 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. Computationally expensive schemes for posterior inference on the multivariate dynamic PARCOR model are avoided using approximate inference. The performance of the TV-VPARCOR methods is illustrated in simulation studies and in the analysis of multivariate non-stationary temporal data arising in neuroscience and in 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.

Chapter 4 extends the models and methods of Chapter 3 by considering shrinkage priors on the TV-VPARCOR parameters in order to reduce overfitting. Such priors allow shrinkage of time-varying parameters to static ones, as well as shrinkage to zero for those parameters that are not statistically significant. A Markov chain Monte Carlo (MCMC) algorithm for full posterior inference is proposed. In addition, an importance sampling variational Bayes (ISVB) approach is also developed and implemented for fast and reliable approximate inference, making the dynamic TV-VPARCOR modeling and inference framework feasible for analysis of large-dimensional time series. The performance of the proposed models and methods is examined in extensive simulation studies and also in a case study involving the analysis of wind component data from several locations in Northern California.

Chapter 5 summarizes this thesis and provides some ideas for possible future work.

Acknowledgments

I would first like to thank my major advisor Professor Raquel Prado for her help, support, and guidance throughout my Ph.D. program. Her kindness, enthusiasm and insight make working with her full of joy. Besides Raquel, this thesis also owes its completion to committee members, Professor Abel Rodriguez and Professor Bruno Sansó for their valuable advice on my research. I am grateful to the faculty at UCSC statistics department for their excellent teaching. And I am also grateful to the students I became friends with, in particular, Xingchen Yu and Bohan Liu, for their help in my studies, and entertaining me outside research. I would also like to thank my family. My grandparents Dongquan Xu and Zhaodi Dai, and my parents Yang Zhao and Min Xu are always supportive and understanding. I particularly want to thank my wife Yiqing Yao for her love, patience and encouragement.

The contents of Chapters 2 and 3 are an adaptation of the articles:

- Wenjie Zhao and Raquel Prado. Efficient Bayesian PARCOR Approaches for Dynamic Modeling of Multivariate Time Series. *Journal of Time Series Analysis*, 41:759-784, 2020.
- Wenjie Zhao and Raquel Prado. Hierarchical Dynamic PARCOR Models for Analysis of Multiple Brain Signals. *Statistics and Its Interface*, in press.

Thanks go to the editors and reviewers of *Journal of Time Series Analysis* and *Statistics and Its Interface*, whose comments and suggestion strengthened the chapters.

Chapter 1

Introduction and Background

Recent advances in information technology have made high-dimensional non-stationary signals increasingly common in many areas. For example, multi-channel electroencephalography (EEG) recordings from a number of locations over a subject's scalp result in high-dimensional time series (see e.g., Prado et al., 2001, Peterson and Ferris, 2019). These recordings are also non-stationary due to the significant changes observed over time in their frequency and amplitude structure. Another area where multi-dimensional non-stationary time series arise is in environmental sciences. For example, in this thesis we show the analysis of wind component data recorded in different locations over a period of time.

In some applied settings we may be interested in analyzing multiple time series to better understand their underlying structure. For example, simultaneous brain signals such as EEG, can be recorded in subjects undergoing a treatment during repeated trials under the same experimental conditions. A typical goal in such settings is not only to infer the time-frequency features of each individual time series, but also the underlying structure common to the multiple time series. Hierarchical time-domain models such as autoregressions (ARs) and vector autoregressions (VARs) –as well as versions of these models that allow for changes in the parameters over time to capture non-stationary behavior, such as time-varying ARs (TVARs) and time-varying VARs (TV-VARs)– have been used to infer latent structure from multiple brain signals (e.g., Gorrostieta et al., 2013, Hu et al., 2020, Nakajima and West, 2013, 2017, Prado and West, 1997, Prado et al., 2001). Some of these approaches consider a hierarchical structure in the AR or VAR coefficients, while others consider latent factor models within a Bayesian framework coupled with sophisticated and flexible prior structures. Other modeling frameworks such as those based on factor models, focus on discovering the latent structure underlying multiple time series (e.g., Nakajima and West, 2017, Prado et al., 2001). Alternative frequency domain and time-frequency domain hierarchical approaches are also available to analyze multiple time series. For example, Cadonna et al. (2019) proposed using a structured mixture of Normal distributions under the Whittle likelihood approximation for analysis of multiple time series. Krafty et al. (2017) proposed a Bayesian tensor-product spline model of the Cholesky components of outcome-dependent power spectra of multiple time series. The sophisticated modeling approaches mentioned above are powerful and have been successfully used in practice to model multiple time series. However, they are usually very computationally expensive, often requiring simulation-based methods, e.g., Markov chain Monte Carlo (MCMC), for inference, which limits their use in practical settings that involve simultaneous modeling of a large number of time series.

In other applied settings, we want to infer the temporal and mutual dependence structure among different time series components. 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, Cheung et al., 2010, Chiang et al., 2017, Milde et al., 2009, Omidvarnia et al., 2014, Schmidt et al., 2016, Ting et al., 2017, Yu et al., 2016, 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, VARs are often used in the analysis of multi-channel 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 in order 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., Nakajima and West, 2017, Prado et al., 2001, West et al., 1999), are either not available for multivariate time series, or they are highly computationally intensive, requiring MCMC sampling for posterior inference. In the time-frequency domain, Li and Krafty (2018) and Bruce et al. (2018) apply adaptive Bayesian analysis with splines for estimating the time-varying spectrum of a multivariate time series. However, these methods are computationally unfeasible to jointly analyze more than a relative small number of multivariate time series components. Furthermore, they cannot be used for forecasting purposes.

Yang et al. (2016) proposed a computationally efficient and fully Bayesian time-varying autoregressive (TVAR) framework for univariate time series analysis by using a dynamic linear model representation in the partial autocorrelation (PARCOR) domain that assumes random walk evolutions on PARCOR coefficients. Compared to the standard dynamic linear model (DLM) representation of a TVAR model in the time-domain, the lattice structure on the PARCOR domain is more flexible and computationally appealing.

This thesis extends the approach of Yang et al. (2016) in different directions. First, we consider hierarchical extensions that allow us to jointly model multiple, rather than multivariate, non-stationary time series. These developments are presented and illustrated in Chapter 2. Then, the approach is extended to the multivariate case, allowing us to jointly analyze several non-stationary time series jointly, and consequently explore their individual spectral characteristics over time and also the time-frequency relationship across several components as shown in Chapter 3. Finally, we consider model extensions that deal with challenges related to overfitting by considering shrinkage priors on the time-varying PARCOR coefficients. This is presented in Chapter 4.

Since all the models proposed in this thesis are developed in the PARCOR domain, we introduce some notation and concepts and present a summary of the univariate dynamic PARCOR model of Yang et al. (2016).

1.1 Time-Varying Autoregressions, Their State-Space and Time-Frequency Representations

A time varying autoregressive (TVAR) model of order P for a non-stationary univariate time series $x_t, t = 1, \ldots, T$, can be expressed as

$$x_t = \sum_{m=1}^{P} a_{m,t}^{(P)} x_{t-m} + \epsilon_t, \tag{1.1}$$

where $a_{m,t}^{(P)}$ denotes the TVAR coefficient associated with time lag m, and ϵ_t is the innovation at time t, respectively. Typically, the innovations are assumed to be uncorrelated mean-zero Gaussian random variables with variance σ_t^2. Therefore, the TVAR model corresponds to a non-stationary AR model with AR coefficients and variances evolving through time.

Assuming that σ_t^2 is known, an additional equation is needed to model the evolution of the AR coefficients over time. Often, the vector of coefficients $\mathbf{a}_t^{(P)} = (a_{1,t}^{(P)}, \ldots, a_{P,t}^{(P)})'$ is assumed to follow a random walk evolution given by

$$\mathbf{a}_t^{(P)} = \mathbf{a}_{t-1}^{(P)} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t). \tag{1.2}$$

Combining equations (1.1) and (1.2) leads to a dynamic linear model (DLM) or state-space model with state-parameter vector of dimension P. Posterior inference can be obtained in this model setting assuming a particular prior structures on the model parameters as discussed in West and Harrison (1997) and Prado et al. (2021). In particular, under a conjugate prior structure filtering and smoothing can be achieved in closed form and requires operations with matrices of dimension $P \times P$ at each time t, given σ_t^2. Approaches for obtaining inference under conjugate prior distributions in cases where $\sigma_t^2 = \sigma^2$ for all t and also in cases in which σ_t^2 is specified via discount factors are discussed and illustrated in West and Harrison (1997) and Prado et al. (2021).

Similarly, a time-varying vector autoregressive model of order P for a K-dimensional time series $\boldsymbol{x}_t$, referred to as TV-VAR(P), is given by

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}^{(P)} \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{A}_{P,t}^{(P)} \boldsymbol{x}_{t-P} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t), \tag{1.3}$$

with $\boldsymbol{\epsilon}_t$ assumed independent over time. Note that in this case the $\boldsymbol{A}_{j,t}$ contain the matrices of VAR coefficients at each time t.

Similar to the evolution specified in equation (1.2) for the TVAR coefficients, the TV-VAR coefficients $\boldsymbol{A}_t = (\boldsymbol{A}_{1,t}^{(P)}, \ldots, \boldsymbol{A}_{P,t}^{(P)})$ are assumed to follow a random walk process given by

$$\text{vec}(\boldsymbol{A}_t) = \text{vec}(\boldsymbol{A}_{t-1}) + \boldsymbol{\omega}_t^{\boldsymbol{A}}, \;\; \boldsymbol{\omega}_t^{\boldsymbol{A}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W}_t^{\boldsymbol{A}}), \tag{1.4}$$

where vec$(\cdot)$ denotes the column stacking operator of a portion of a matrix. The dimension of state parameter $\boldsymbol{A}_t$ is K^2P. Conditional on $\boldsymbol{\Omega}_t$, filtering and smoothing can be achieved in closed form under a conjugate prior structure, requiring operations with matrices of dimension $K^2P \times K^2P$ at each time t, given variance matrices $\boldsymbol{\Omega}_t$.

1.1.1 TVAR and TV-VAR Time-Frequency Representations

For $t = 1 : T$, the time-frequency representation associated with a TVAR(P) model can be obtained in terms of the spectral density at time t and frequency ω

given by the following equation

$$g(t,\omega) = \frac{\sigma_t^2}{|1 - \sum_{m=1}^{P} a_{m,t}^{(P)} \exp\{-2\pi i m\omega\}|^2}, \quad 0 \le \omega \le 1/2, \tag{1.5}$$

where $i = \sqrt{-1}$ (see, e.g. Kitagawa and Gersch, 1996).

Similarly, the time-varying spectral density matrix associated with a TV-VAR(P) model is given by:

$$\boldsymbol{g}(t,\omega) = \boldsymbol{\Phi}^{-1}(t,\omega) \times \boldsymbol{\Omega}_t \times \boldsymbol{\Phi}^{*}(t,\omega)^{-1}, \quad 0 \le \omega \le 1/2, \tag{1.6}$$

where $\boldsymbol{\Phi}(t,\omega) = \boldsymbol{I}_K - \sum_{m=1}^{P} \boldsymbol{A}_{m,t}^{(P)} \exp\{-2\pi i m\omega\}$, with $i = \sqrt{-1}$ (see e.g., Shumway and Stoffer, 2017, Chapter 4). 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 the squared coherence between components i and j can be defined as follows:

$$\rho_{i,j}^2(t,\omega) = |g_{i,j}(t,\omega)|^2 / \{g_{i,i}(t,\omega) g_{j,j}(t,\omega)\},$$

for all $i \neq j$. This measure is used to estimate the power transfer between two components of the time series.

The squared partial coherence between two time series components can also be obtained. 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 squared partial coherence between components i and j is given by:

$$\gamma_{i,j}^2(t,\omega) = |c_{i,j}(t,\omega)|^2 / \{c_{i,i}(t,\omega) c_{j,j}(t,\omega)\}.$$

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 partial directed coherence (PDC) and the direct tranfer function (DTF) can also be computed (see e.g., Astolfi et al., 2008, Baccalá and Sameshima, 2001, Blinowska, 2011, Kuś et al., 2004, 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 PDC from signal j to signal i at time t and frequency ω is given by

$$PDC_{i,j}(t,\omega) \;=\; \frac{\mathbf{\Phi}_{i,j}(t,\omega)}{\sqrt{\mathbf{\Phi}^{*}_{\cdot,j}(t,\omega)\mathbf{\Phi}_{\cdot,j}(t,\omega)}},$$

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

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

where $\mathbf{\Phi}^{-1}_{i,j}(t,\omega)$ is the (i,j)th element of the matrix $\mathbf{\Phi}^{-1}(t,\omega)$ and $\mathbf{\Phi}^{-1}_{i,\cdot}(t,\omega)$ 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”, e.g., 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, i.e., indirect propagations like $1 \to 2 \to 3$ are not included.

1.2 Dynamic PARCOR models for Analysis of Univariate Non-stationary Time Series

In this section we review the dynamic PARCOR approach of Yang et al. (2016) for univariate analysis of non-stationary time series. As mentioned before we extend this approach for analysis of multiple (Chapter 2) and multivariate (Chapters 3 and 4) non-stationary time series. More specifically, Yang et al. (2016) proposes an approach to obtain posterior inference in a TVAR model on the PARCOR domain as follows. Let $f_t^{(P)}$ and $b_t^{(P)}$ denote the prediction error at time t for the forward and backward TVAR(P) model, respectively, where

$$f_t^{(P)} = x_t - \sum_{m=1}^{P} a_{m,t}^{(P)} x_{t-m}, \quad \text{and } b_t^{(P)} = x_t - \sum_{m=1}^{P} d_{m,t}^{(P)} x_{t+m}.$$

Then, the m-th stage of the lattice filter for a TVAR(P) (Hayes, 1996) can be characterized by the pair of input-output relations between the forward and backward prediction errors as follows:

$$f_t^{(m-1)} = \alpha_{m,t}^{(m)} b_{t-m}^{(m-1)} + f_t^{(m)}, \tag{1.7}$$

$$b_t^{(m-1)} = \beta_{m,t}^{(m)} f_{t+m}^{(m-1)} + b_t^{(m)}, \qquad m = 1, 2, \ldots, P, \tag{1.8}$$

with initial conditions $f_t^{(0)} = b_t^{(0)} = x_t$, and where $\alpha_{m,t}^{(m)}$ and $\beta_{m,t}^{(m)}$ are the lag m forward and backward PARCOR coefficients, respectively. In addition, it is assumed that $f_t^{(m)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{f,m,t}^2)$, and $b_t^{(m)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{b,m,t}^2)$. When the true process is a TVAR(P), the variance $\sigma_{f,P,t}^2 = \sigma_t^2$.

The main advantage of using the above dynamic PARCOR lattice filter representation instead of the DLM TVAR representation discussed in equation (1.1) in Section 1.1, 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.

1.2.1 Model Specification and Bayesian Inference

Yang et al. (2016) complete the model specification described in Section 1.2 by using an additive random walk model for the evolution of the forward and backward PARCOR coefficients, and a multiplicative random walk model for the variances of the innovations for the forward and backward PARCOR coefficients, i.e.,

$$\alpha_{m,t}^{(m)} = \alpha_{m,t-1}^{(m)} + \omega_{f,m,t}, \quad \omega_{f,m,t} \sim \mathcal{N}(0, W_{f,m,t}), \tag{1.9}$$

$$\beta_{m,t}^{(m)} = \beta_{m,t-1}^{(m)} + \omega_{b,m,t}, \quad \omega_{b,m,t} \sim \mathcal{N}(0, W_{b,m,t}), \tag{1.10}$$

$$\sigma_{f,m,t}^2 = \sigma_{f,m,t-1}^2(\delta_{f,m}/\eta_{f,m,t}), \quad \eta_{f,m,t} \sim \mathcal{B}(g_{f,m,t}, h_{f,m,t}), \tag{1.11}$$

$$\sigma_{b,m,t}^2 = \sigma_{b,m,t-1}^2(\delta_{b,m}/\eta_{b,m,t}), \quad \eta_{b,m,t} \sim \mathcal{B}(g_{b,m,t}, h_{b,m,t}), \tag{1.12}$$

where observational variances $\sigma_{f,m,t}^2$ and $\sigma_{b,m,t}^2$ are specified via discount factors $\delta_{f,m}$ and $\delta_{b,m}$, and system variances $\omega_{f,m,t}$ and $\omega_{b,m,t}$ are controlled by the different discount factors $\gamma_{f,m}$ and $\gamma_{b,m}$, respectively. The all discount factors $\delta_{\cdot,m}$ and $\omega_{\cdot,m}$ are in $(0, 1]$. The series of stochastic error terms $\omega_{\cdot,m,t}$ and $\eta_{\cdot,m,t}$ are assumed mutually independent, and also independent of the forward and backward innovations, $f_t^{(m)}$ and $b_t^{(m)}$ of equations (1.7) and (1.8). The conjugate initial priors for $\alpha_{m,0}^{(m)}, \beta_{m,0}^{(m)}, \sigma_{f,m,0}^2$, and $\sigma_{b,m,0}^2$ at each stage m are specified as follows

$$\alpha_{m,0}^{(m)}|D_{f,m,0}, \sigma_{f,m,0}^2 \sim \mathcal{N}(m_{f,m,0}, c_{f,m,0}), \quad \beta_{m,0}^{(m)}|D_{b,m,0}, \sigma_{b,m,0}^2 \sim \mathcal{N}(m_{f,m,0}, c_{f,m,0}),$$
$$\sigma_{f,m,0}^2|D_{f,m,0} \sim \mathcal{IG}(\nu_{f,m,0}/2, \kappa_{f,m,0}/2), \quad \sigma_{b,m,0}^2|D_{b,m,0} \sim \mathcal{IG}(\nu_{b,m,0}/2, \kappa_{b,m,0}/2),$$

where $D_{\cdot,m,0}$ denote the information sets at the initial time $t = 0$ for the forward and backward cases.

Following West and Harrison (1997) and West et al. (1999), the joint posterior forward filtering distributions of $\alpha_{m,t}^{(m)}, \sigma_{f,m,t}^2|D_{f,m,t}$ over $t = 1 : T$ can be sequentially updated, where $D_{f,m,t} = \{D_{f,m,t-1}, x_t\}$. Because of using the conjugate Normal-Inverse Gamma form, the posterior distribution of $\alpha_{m,t}^{(m)}, \sigma_{f,m,t}^2|D_{f,m,t}$ also has a Normal-Inverse Gamma form. Then, the marginal posterior distribution of $\alpha_{m,t}^{(m)}$ is a student T-distribution, i.e., $\alpha_{m,t}^{(m)}|D_{f,m,t} \sim \mathcal{T}_{\nu_{f,m,t}}(m_{f,m,t}, c_{f,m,t})$, with degrees of freedom $\nu_{f,m,t}$, location parameter $m_{f,m,t}$, and scale parameter $c_{f,m,t}$. The marginal posterior distribution of $\sigma_{f,m,t}^2|D_{f,m,t}$ is an Inverse Gamma distribution, i.e. $\sigma_{f,m,t}^2|D_{f,m,t} \sim \mathcal{IG}(\nu_{f,m,t}/2, \kappa_{f,m,t}/2)$, with shape parameter $\nu_{f,m,t}/2$ and rate parameter $\kappa_{f,m,t}/2$. After the sequential updating process, a retrospective approach can be applied to specify the forward smoothing distribution of $\alpha_{m,t}^{(m)}|D_{f,m,T}$, which is approximately a student T-distribution with the degrees of freedom $\nu_{f,m,t|T}$, location parameter $m_{f,m,t|T}$, and scale parameter $c_{f,m,t|T}$. In addition, $\sigma_{f,m,t}^2|D_{f,m,T}$ follows approximately an Inverse Gamma distribution with shape parameter $\nu_{f,m,t|T}/2$ and rate parameter $\kappa_{fm,t|T}/2$. Details on the posterior filtering and smoothing equations for the forward model are provided in Yang et al. (2016). In addition, the posterior filtering and smoothing equations can be similarly obtained for the backward time-varying PARCOR model and their corresponding variances.

1.2.2 Model Selection

Yang et al. (2016) select the model order and discount factor values $\{P, \gamma_m, \delta_m; m = 1, \ldots, P\}$ as follows. First, they specify a potential maximum value of P and a set of grid values for $\{\gamma_m, \delta_m\}$ for each stage m. Given P, they search for the combination of $\{\gamma_1, \delta_1\}$ maximizing the log likelihood derived from lattice filters equation (1.7) at stage $m = 1$. Then, they use the selected γ_1 and δ_1 to obtain the corresponding series $\{f_t^{(2)}\}$ and $\{b_t^{(2)}\}$, for $t = 1, \ldots, T$, and the corresponding log maximum likelihood $\mathcal{L}_1$. They repeat above procedure until the set of $\{\gamma_m, \delta_m, \mathcal{L}_m\}, m = 1, \ldots, P$, has been obtained.

Yang et al. (2016) suggest to select the model order visually by plotting $\mathcal{L}_{f,m}$ against the order m. The idea is that, when the observed time series truly follows a model of order P, the values of $\mathcal{L}_{f,m}$ will stop increasing after $m = P$, appropriately indicating the model order. They also provide a numerical approach of order selection based on calculating the relative change from $\mathcal{L}_{m-1}$ to $\mathcal{L}_m$ with respect to m,

$$|(\mathcal{L}_m - \mathcal{L}_{m-1})/\mathcal{L}_{m-1}| \times 100 < \tau.$$

They suggest to choose $\tau = 0.5$, which provides an effective cutoff to choose the order.

1.2.3 Durbin-Levinson Algorithm

The Durbin-Levinson Algorithm provides an unique correspondence between the PARCOR coefficients and AR coefficients (Brockwell and Davis, 1991, Shumway and Stoffer, 2017). The forward and backward TVAR coefficients $a_{m,t}^{(m)}$

and $d_{m,t}^{(m)}$ can be transformed by the PARCOR coefficients, $\alpha_{m,t}^{(m)}$ and $\beta_{m,t}^{(m)}$ as follows

$$a_{j,t}^{(m)} = a_{j,t}^{(m-1)} - a_{m,t}^{(m)} d_{m-j,t}^{(m-1)}, \tag{1.13}$$

$$d_{j,t}^{(m)} = d_{j,t}^{(m-1)} - d_{m,t}^{(m)} a_{m-j,t}^{(m-1)}, \quad j = 1, \ldots, m-1, \tag{1.14}$$

where $a_{m,t}^{(m)} = \alpha_{m,t}^{(m)}$ and $d_{m,t}^{(m)} = \beta_{m,t}^{(m)}$.

The idea of Yang et al. (2016) 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.

In addition, posterior summaries in the time-frequency domain can be obtained using the relationship between the TVAR coefficients and the spectral density given by equation (1.5).

1.3 Outline of the thesis: PARCOR Framework in Multiple and Multivariate Scenarios

In this thesis we extend the dynamic lattice filter PARCOR approach of Yang et al. (2016) to the case of multiple time series, by using the hierarchical dynamic linear structure of Gamerman and Migon (1993). The proposed hierarchical approach described in Chapter 2, once again describes the time-varying behavior of multiple time series in the partial autocorrelation domain, which results in lower dimensional representations, and consequently computationally faster inference, than those required by models in the time and/or frequency domains, such as hierarchical TVAR models, which are commonly used in practice. We illustrate the performance of the hierarchical TV-PARCOR models and the corresponding Bayesian inferential procedures in the context of analyzing multiple brain signals recorded simultaneously during specific experimental settings or clinical studies. The proposed approach allows us to efficiently obtain posterior summaries of the time-frequency characteristics of the multiple time series, as well as those summarizing their common underlying structure.

We then develop and implement a dynamic multivariate PARCOR model with a multivariate DLM structure (West and Harrison, 1997) in Chapter 3. This approach offers computational feasibility and interpretable time-frequency analysis in the multivariate context. Our 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 DLMs on the forward and backward time-varying vector partial autocorrelation coefficients (TV-VPARCOR). Computationally expensive schemes for posterior inference on this multivariate dynamic PARCOR model are avoided using approximations for filtering and smoothing, which leads to approximate inference on the corresponding TV-VAR coefficients obtained via Whittle's algorithm (Zhou, 1992). 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 representations offered by the TV-VPARCOR models have a huge advantage in terms of flexibility in modeling gradual temporal changes, however in spite of being more parsimonious than state-space representations of TV-VAR, still present challenges related to overfitting that increases with the growing number of coefficients. In Chapter 3, we developed a DIC-based model selection criterion to determine the model order P as a way to limit the negative effects of overfitting. However, in some large data sets scenarios, lower order models may be incorrectly chosen by the DIC because of an over-penalization on the number of parameters. Moreover, the assumption that all the model parameters are time-varying through entire time period may not be realistic and needed. It is reasonable to expect some parameters to remain unchanged over time. Therefore, considering shrinkage priors on the PARCOR coefficients is a good option to overcome the overfitting issues. Such priors may induce shrinkage towards zero in PARCOR coefficients with relatively small effects and can also shrink time-varying coefficients to static ones when needed.

There are many approaches in the time series literature designed to overcome the challenge of over-parameterization in time-varying parameter models using Bayesian methods. Nakajima and West (2015) propose a latent threshold model that regularizes all parameters within a certain threshold to zero. Rockova and McAlinn (2021) develop dynamic spike-and-slab priors which extend stochastic search variable selection (SSVS) priors. Frühwirth-Schnatter and Wagner (2010) propose a non-centered parameterization of the state-space model to deal with variable selection. Many shrinkage priors can be applied on this model structure, e.g., the Bayesian lasso prior (Park and Casella, 2008), the Normal-Gamma prior (Bitto and Frühwirth-Schnatter, 2019, Griffin and Brown, 2010), the Horseshoe prior (Carvalho et al., 2010), the Dirichlet-Laplace prior (Bhattacharya et al., 2015) and the Normal-mixture of inverse Gamma prior (Ishwaran and Rao, 2005). In Chapter 4, the non-centered parameterization of regression state-space models (Bitto and Frühwirth-Schnatter, 2019, Frühwirth-Schnatter and Wagner, 2010) is incorporated into the dynamic multivariate PARCOR framework. The Bayesian Lasso priors (Belmonte et al., 2014) and Normal-Gamma priors (Griffin and

Brown, 2010) are imposed on the model parameters in order to shrink time-varying PARCOR coefficients to static ones (including zero), when the model is overfitting. A MCMC algorithm is proposed for full posterior inference in this setting. In addition, an importance sampling variational Bayes (ISVB) approach is developed for fast and reliable approximate inference, which makes the dynamic PARCOR modeling and inference framework feasible for analysis of large-dimensional time series. We compare the results between TV-VPARCOR model with conjugate priors but no shrinkage developed in Chapter 3 and the TV-VPARCOR model with shrinkage priors in extensive simulation studies and a case study involving the analysis of wind component data from several locations in northern California.

Finally, Chapter 5 summarizes the results of this thesis and discusses some future directions.

Chapter 2

Hierarchical Dynamic PARCOR Models

We extend the univariate dynamic PARCOR model (Yang et al., 2016) to a hierarchical time-varying PARCOR model for analysis of multiple non-stationary time series. In particular, we develop a fast algorithm for posterior inference in this hierarchical PARCOR setting. A simulation study illustrates the advantages of the hierarchical dynamic PARCOR model over the standard TVAR models in terms of computational efficiency and functionality. We also show the performance of our proposed model in the analysis of two datasets: a dataset corresponding to multiple brain signals recorded simultaneously on an individual under a particular experimental condition, and a clinical multi-channel EEG dataset.

2.1 Introduction

In neuroscience studies a given subject may undergo repeated trials under a given experimental condition, resulting in multiple brain recordings. In such cases, researchers may need to summarize the information provided by the subject-specific repeated measurements, taken under a given experimental condition, for comparison with results for other subjects who were on the same experimental condition. Often the brain signals are averaged over trials for a given subject and a given condition, which is typically problematic, as this results in over-smoothing and information loss in the time-frequency domain.

Here we extend the univariate dynamic partial autocorrelation (PARCOR) model of (Yang et al., 2016) explained in Section 1.2 of Chapter 1 to obtain a hierarchical dynamic PARCOR model that is able to simultaneously describe the time-frequency behavior of multiple related time series and characterize their common underlying features, while also having the advantages of the PARCOR representation in terms of dimension reduction and computational efficiency, particularly when compared to other time or frequency-domain hierarchical models used in practice, such as those based on AR, TVAR, VAR, and TV-VAR representations. Our approach allows us to jointly analyze multiple brain signals recorded on a subject using a single model that is able to infer the common time-frequency characteristics underlying signals (Section 2.4.1 illustrates this in a non-clinical EEG study). The proposed hierarchical PARCOR approach can also be used to efficiently infer the latent structure underlying multiple signals recorded at different locations on a subject during a single trial (Section 2.4.2 illustrates this with data from a clinical EEG study).

2.2 Hierarchical PARCOR models

Assume we observe a set of n time series $\{x_{it}\}$ for $i = 1, \ldots, n$, and $t = 1, \ldots, T$, where i is the index of the time series. A TVAR model of order P

(Section 1.1 of Chapter 1) for time series i is given by

$$x_{it} = \sum_{j=1}^{P} a_{i,j,t}^{(P)} x_{i,t-j} + \epsilon_{it},$$

where $a_{i,j,t}^{(P)}$ denotes the ith series-specific TVAR coefficient associated with lag j at time t, and ϵ_{it} is the corresponding innovation. Typically, the innovations are assumed to be independent mean-zero Gaussian random variables.

As mentioned in Section 1.2 of Chapter 1, the univariate PARCOR representation of TVARs leads to lower dimensional models. This advantage of the PARCOR representations is increased in multiple time series analysis. For instance, in the case of a hierarchical TVAR DLM for joint modeling of n time series, inversion of $nP \times nP$ matrices is required at each step of the filtering and smoothing algorithms. Instead, the dynamic hierarchical PARCOR model proposed here requires the inversion of $n \times n$ matrices of the filtering and smoothing algorithms for P stages for the forward and backward coefficients, which results in a significant reduction in computation time, particularly when P is moderate or relatively large. In order to proceed with the hierarchical dynamic PARCOR model specification, let $f_{it}^{(P)}$ and $b_{it}^{(P)}$ be the prediction error of the ith time series at time t for the forward and backward TVAR(P) model respectively, where,

$$f_{it}^{(P)} = x_{it} - \sum_{j=1}^{P} a_{i,j,t}^{(P)} x_{i,t-j}, \tag{2.1}$$

$$b_{it}^{(P)} = x_{it} - \sum_{j=1}^{P} d_{i,j,t}^{(P)} x_{i,t+j}. \tag{2.2}$$

$a_{i,j,t}^{(P)}$ and $d_{i,j,t}^{(P)}$ denote, respectively, the time-varying forward and backward TVAR coefficients for jth lag, where $j = 1, \ldots, P$ and $i = 1, \ldots, n$. Similarly, $a_{i,j,t}^{(m)}$ and $d_{i,j,t}^{(m)}$ denote the time-varying forward and backward TVAR(m) coefficients for

$j = 1, \dots, m$. Then, the m-stage of the PARCOR lattice filter can be written in terms of the pair of input-output relations between the forward and backward prediction errors, as follows,

$$f_{it}^{(m-1)} = \alpha_{i,m,t}^{(m)} b_{i,t-m}^{(m-1)} + f_{it}^{(m)}, \quad f_{it}^{(m-1)} \sim \mathcal{N}(0, \sigma_{f,i,m,t}^2), \tag{2.3}$$

$$b_{it}^{(m-1)} = \beta_{i,m,t}^{(m)} f_{i,t+m}^{(m-1)} + b_{it}^{(m)}, \quad b_{it}^{(m-1)} \sim \mathcal{N}(0, \sigma_{b,i,m,t}^2), \tag{2.4}$$

where $\alpha_{i,m,t}^{(m)}$ and $\beta_{i,m,t}^{(m)}$ are, respectively, time-varying forward and backward PARCOR coefficients at stage m, with $m = 1, \dots, P$. Note that for stationary AR(P) models, the forward and backward PARCOR coefficients are constant over time and equal, that is $\alpha_{i,m}^{(m)} = \beta_{i,m}^{(m)}$ for all m.

Then, at each stage m of the lattice structure above, we can obtain the forward and backward TVAR coefficients of each time series i, $a_{i,j,t}^{(m)}$ and $d_{i,j,t}^{(m)}$, from the PARCOR coefficients, $\alpha_{i,m,t}^{(m)}$ and $\beta_{i,m,t}^{(m)}$ using Durbin-Levinson algorithm illustrated in Section 1.2.3 of Chapter 1, applied separately to obtain the TVAR coefficients for each time series i.

2.2.1 Model Specification and Inference

The first level of our proposed hierarchical model specification uses equations (2.3) and (2.4) as observational level equations of univariate DLMs (Prado et al., 2021, West and Harrison, 1997) on the forward and backward PARCOR time-varying coefficients.

The next level of the hierarchical model requires specifying the structural equations (Gamerman and Migon, 1993). For this level we assume that the forward and backward PARCOR coefficients at lag m and time t for each series i are decomposed as follows

$$\alpha_{i,m,t}^{(m)} = \mu_{f,m,t}^{(m)} + \gamma_{f,i,m,t}^{(m)} + \nu_{f,i,m,t}, \tag{2.5}$$

$$\beta_{i,m,t}^{(m)} = \mu_{b,m,t}^{(m)} + \gamma_{b,i,m,t}^{(m)} + \nu_{b,i,m,t}, \tag{2.6}$$

where the structural innovations $\nu_{f,i,m,t}$ and $\nu_{b,i,m,t}$ follow zero-mean normal distributions. Further assumptions regarding these distributions are provided below. $\mu_{f,m,t}^{(m)}$ and $\mu_{b,m,t}^{(m)}$ denote the common underlying forward and backward effects across all the time series, respectively. In addition, $\gamma_{f,i,m,t}^{(m)}$ and $\gamma_{b,i,m,t}^{(m)}$ respectively denote forward and backward effects that are specific to time series i. To avoid identifiability issues, we add restrictions on these parameters for both, forward and backward coefficients, i.e., we assume $\sum_{i=1}^{n} \gamma_{\cdot,i,m,t}^{(m)} = 0$ for all t and m.

The final level of the hierarchical model requires specification of the system equations that describe the variation of the parameters over time. We specify random walk system equations for both forward and backward common effects $\mu_{\cdot,m,t}^{(m)}$ and also for the series-specific effects $\gamma_{\cdot,i,m,t}^{(m)}$ as follows,

$$\mu_{f,m,t}^{(m)} = \mu_{f,m,t-1}^{(m)} + w_{\mu,f,m,t}, \tag{2.7}$$

$$\mu_{b,m,t}^{(m)} = \mu_{b,m,t-1}^{(m)} + w_{\mu,b,m,t}, \tag{2.8}$$

$$\gamma_{f,i,m,t}^{(m)} = \gamma_{f,i,m,t-1}^{(m)} + w_{\gamma,f,i,m,t},\ i = 1, \ldots, n-1, \tag{2.9}$$

$$\gamma_{b,i,m,t}^{(m)} = \gamma_{b,i,m,t-1}^{(m)} + w_{\gamma,b,i,m,t},\ i = 1, \ldots, n-1, \tag{2.10}$$

where the forward and backward system innovations $w_{\mu,\cdot,m,t}$ and $w_{\gamma,\cdot,i,m,t}$ follow normal zero-mean distributions. In addition, conjugate normal priors are assumed for $\mu_{f,m,0}^{(m)}$, $\mu_{b,m,0}^{(m)}$, $\gamma_{f,i,m,0}^{(m)}$ and $\gamma_{b,i,m,0}^{(m)}$ for all m and i. We further discuss these distributions below when we summarize the model in matrix form.

In the case of a normal DLM, posterior inference is available in closed form via the DLM filtering and smoothing equations (West and Harrison, 1997), however, this is not the case in hierarchical DLMs when the observational, structural and system variance are unknown. A conjugate model structure is available Gamerman and Migon (1993), West and Harrison (1997), if the observational, structural and system variances are scaled by a single observational variance, and the scaling factors are assumed to be known. In our model we assume that the forward prediction errors of the time series at each stage m in equation (2.3) independently follow normal distributions with the observational innovation variance scaled by the parameter $\sigma^2_{f,m}$ at the different stage m. Similarly, the backward prediction errors of all time series (2.4) are assumed to independently follow normal distributions with the observational innovation variances scaled by $\sigma^2_{b,m}$. In other words we set $\sigma^2_{f,i,m,t} = \sigma^2_{f,m}$ and $\sigma^2_{b,i,m,t} = \sigma^2_{b,m}$ for each stage m and all t, and further assume conjugate prior distributions of $\sigma^2_{f,m}$ and $\sigma^2_{b,m}$ as follows:

$$\sigma^2_{f,m} | \boldsymbol{D}_{f,m,0} \sim \mathcal{IG}\left(\frac{n_{f,0}}{2}, \frac{h_{f,0}}{2}\right), \tag{2.11}$$

$$\sigma^2_{b,m} | \boldsymbol{D}_{b,m,0} \sim \mathcal{IG}\left(\frac{n_{b,0}}{2}, \frac{h_{b,0}}{2}\right). \tag{2.12}$$

Here $\boldsymbol{D}_{f,m,t}$ and $\boldsymbol{D}_{b,m,t}$ denote, respectively, all the information available for the forward and backward models at stage m and time t.

We can then rewrite the hierarchical forward and backward PARCOR models above in matrix form as below.

- Observation equations:

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{F}_{1,f,m,t}^{(m-1)} \boldsymbol{\theta}_{1,f,m,t}^{(m)} + \boldsymbol{f}_t^{(m)}, \boldsymbol{f}_t^{(m)} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{f,m} \boldsymbol{I}_n\right), \tag{2.13}$$

$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{F}_{1,b,m,t}^{(m-1)} \boldsymbol{\theta}_{1,b,m,t}^{(m)} + \boldsymbol{b}_t^{(m)}, \boldsymbol{b}_t^{(m)} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{b,m} \boldsymbol{I}_n\right), \tag{2.14}$$

where $\boldsymbol{f}_t^{(m-1)} = (f_{1,t}^{(m-1)}, \dots, f_{n,t}^{(m-1)})'$, $\boldsymbol{F}_{1,f,m,t}^{(m-1)} = \text{diag}(b_{1,t-m}^{(m-1)}, \dots, b_{n,t-m}^{(m-1)})$, $\boldsymbol{\theta}_{1,f,m,t}^{(m)} = (\alpha_{1,m,t}^{(m)}, \dots, \alpha_{n,m,t}^{(m)})'$, $\boldsymbol{b}_t^{(m-1)} = (b_{1,t}^{(m-1)}, \dots, b_{n,t}^{(m-1)})'$, $\boldsymbol{F}_{1,b,m,t}^{(m-1)} = \text{diag}(f_{1,t+m}^{(m-1)}, \dots, f_{n,t+m}^{(m-1)})$, $\boldsymbol{\theta}_{1,b,m,t}^{(m)} = (\beta_{1,m,t}^{(m)}, \dots, \beta_{n,m,t}^{(m)})'$.

- Structural equations:

$$\boldsymbol{\theta}_{1,f,m,t}^{(m)} = \boldsymbol{F}_2 \boldsymbol{\theta}_{2,f,m,t}^{(m)} + \boldsymbol{\nu}_{f,m,t}^{(m)}, \boldsymbol{\nu}_{f,m,t} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}_{2,f,m,t}\right), \tag{2.15}$$

$$\boldsymbol{\theta}_{1,b,m,t}^{(m-1)} = \boldsymbol{F}_2 \boldsymbol{\theta}_{2,b,m,t}^{(m)} + \boldsymbol{\nu}_{b,m,t}^{(m)}, \boldsymbol{\nu}_{b,m,t} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}_{2,b,m,t}\right), \tag{2.16}$$

where

$$\boldsymbol{F}_2 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & -1 & -1 & \dots & -1 \end{pmatrix},$$

$\boldsymbol{\theta}_{2,\cdot,m,t}^{(m)} = (\mu_{\cdot,m,t}^{(m)}, \gamma_{\cdot,1,m,t}^{(m)}, \dots, \gamma_{\cdot,n-1,m,t}^{(m)})'$, $\boldsymbol{\nu}_{\cdot,m,t}^{(m)} = (\nu_{\cdot,1,m,t}, \dots, \nu_{\cdot,n,m,t})'$, $\boldsymbol{V}_{2,\cdot,m,t} = \sigma_{\cdot,m}^2 \boldsymbol{V}_{2,\cdot,m,t}^*$. The scale-free structural innovation variance-covariance matrices $\boldsymbol{V}_{2,\cdot,m,t}^*$ are controlled by structural discount factors $\delta_{1,\cdot,m}$ for each stage m. Discount factors are widely used in practice to specify variance-covariance matrix West and Harrison (1997). Note that here we imposed the restriction $\sum_{i=1}^{n} \gamma_{\cdot,i,m,t}^{(m)} = 0$ so that we have the matrix $\boldsymbol{F}_2$ above that appears in equations (2.15) and (2.16).

- System equations:

$$\boldsymbol{\theta}_{2,f,m,t}^{(m)} = \boldsymbol{\theta}_{2,f,m,t-1}^{(m)} + \boldsymbol{w}_{f,m,t}, \boldsymbol{w}_{f,m,t} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{W}_{f,m,t}\right), \tag{2.17}$$

$$\boldsymbol{\theta}_{2,b,m,t}^{(m)} = \boldsymbol{\theta}_{2,b,m,t-1}^{(m)} + \boldsymbol{w}_{b,m,t}, \boldsymbol{w}_{b,m,t} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{W}_{b,m,t}\right), \tag{2.18}$$

where $\boldsymbol{w}_{\cdot,m,t} = (w_{\mu,\cdot,m,t}, w_{\gamma,\cdot,1,m,t}, \ldots, w_{\gamma,\cdot,n-1,m,t})'$, $\boldsymbol{W}_{\cdot,m,t} = \sigma^2_{\cdot,m}\boldsymbol{W}^*_{\cdot,m,t}$. The scale-free system innovation variance-covariance matrices $\boldsymbol{W}^*_{\cdot,m,t}$ are controlled by system discount factors $\delta_{2,\cdot,m}$ for each stage m.

- Finally, in addition to the priors (2.11) and (2.12) on the innovation variances, we also assume

$$\begin{aligned}
\boldsymbol{\theta}^{(m)}_{2,f,m,0}|\boldsymbol{D}_{f,m,0} &\sim \mathcal{N}\left(\boldsymbol{m}_{f,m,0}, \sigma^2_{f,m}\boldsymbol{C}^*_{f,m,0}\right), \\
\boldsymbol{\theta}^{(m)}_{2,b,m,0}|\boldsymbol{D}_{b,m,0} &\sim \mathcal{N}\left(\boldsymbol{m}_{b,m,0}, \sigma^2_{b,m}\boldsymbol{C}^*_{b,m,0}\right).
\end{aligned}$$

Given the model structure above we obtain the filtering equations below for closed-form inference in the hierarchical TV-PARCOR model as follows.

- Prior distributions conditional on $\sigma^2_{\cdot,m}$ at time t,

$$\boldsymbol{\theta}^{(m)}_{k,\cdot,m,t}|\boldsymbol{D}_{\cdot,m,t-1}, \sigma^2_{\cdot,m} \sim \mathcal{N}\left(\boldsymbol{a}_{k,\cdot,m,t}, \sigma^2_{\cdot,m}\boldsymbol{R}^*_{k,\cdot,m,t}\right), k = 1, 2,$$

where

$$\begin{aligned}
\boldsymbol{a}_{2,\cdot,m,t} &= \boldsymbol{m}_{2,\cdot,m,t-1}, \\
\boldsymbol{R}^*_{2,\cdot,m,t} &= \boldsymbol{C}^*_{2,\cdot,m,t-1} + \boldsymbol{W}^*_{\cdot,m,t}, \\
\boldsymbol{W}^*_{\cdot,m,t} &= \frac{1-\delta_{2,\cdot,m}}{\delta_{2,\cdot,m}}\boldsymbol{C}^*_{2,\cdot,m,t-1}, \\
\boldsymbol{a}_{1,\cdot,m,t} &= \boldsymbol{F}_2\boldsymbol{a}_{2,\cdot,m,t}, \\
\boldsymbol{R}^*_{1,\cdot,m,t} &= \boldsymbol{F}_2\boldsymbol{R}^*_{2,\cdot,m,t}\boldsymbol{F}'_2 + \boldsymbol{V}^*_{2,\cdot,m,t}, \\
\boldsymbol{V}^*_{2,\cdot,m,t} &= \frac{1-\delta_{1,\cdot,m}}{\delta_{1,\cdot,m}}\boldsymbol{F}_2\boldsymbol{R}^*_{2,\cdot,m,t}\boldsymbol{F}'_2.
\end{aligned}$$

The values of forward and backward structural discount factors, $\delta_{1,\cdot,m}$, and the system discount factors, $\delta_{2,\cdot,m}$, are determined by maximizing loglikelihoods resulting from equations (2.13) and (2.14). Details about the selection of discount factors are discussed in Section 2.2.2.

- One-step ahead predictive distributions conditional on $\sigma^2_{\cdot,m}$

$$\begin{aligned}
\boldsymbol{f}_t^{(m-1)}|\boldsymbol{D}_{f,m,t-1}, \sigma^2_{f,m} &\sim \mathcal{N}\left(\boldsymbol{g}_{f,m,t}, \sigma^2_{f,m}\boldsymbol{Q}^*_{f,m,t}\right),\\
\boldsymbol{b}_t^{(m-1)}|\boldsymbol{D}_{b,m,t-1}, \sigma^2_{b,m} &\sim \mathcal{N}\left(\boldsymbol{g}_{b,m,t}, \sigma^2_{b,m}\boldsymbol{Q}^*_{b,m,t}\right),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{g}_{\cdot,m,t} &= \boldsymbol{F}^{(m-1)}_{1,\cdot,m,t}\boldsymbol{a}_{1,\cdot,m,t},\\
\boldsymbol{Q}^*_{\cdot,m,t} &= \boldsymbol{I}_n + \boldsymbol{F}^{(m-1)}_{1,\cdot,m,t}\boldsymbol{R}^*_{1,\cdot,m,t}\boldsymbol{F}^{(m-1)}_{1,\cdot,m,t}.
\end{aligned}$$

- Posterior distributions conditional on $\sigma^2_{\cdot,m}$ at time t for $k = 1, 2$,

$$\boldsymbol{\theta}^{(m)}_{k,\cdot,m,t}|\boldsymbol{D}_{\cdot,m,t}, \sigma^2_{\cdot,m} \sim \mathcal{N}\left(\boldsymbol{m}_{k,\cdot,m,t}, \sigma^2_{\cdot,m}\boldsymbol{C}^*_{k,\cdot,m,t}\right),$$

where

$$\begin{aligned}
\boldsymbol{m}_{2,f,m,t} &= \boldsymbol{a}_{2,f,m,t} + \boldsymbol{S}_{2,f,m,t}\boldsymbol{Q}^{*,-1}_{f,m,t}\left(\boldsymbol{f}_t^{(m-1)} - \boldsymbol{g}_{f,m,t}\right),\\
\boldsymbol{m}_{2,b,m,t} &= \boldsymbol{a}_{2,b,m,t} + \boldsymbol{S}_{2,b,m,t}\boldsymbol{Q}^{*,-1}_{b,m,t}\left(\boldsymbol{b}_t^{(m-1)} - \boldsymbol{g}_{b,m,t}\right),\\
\boldsymbol{C}^*_{2,\cdot,m,t} &= \boldsymbol{R}^*_{2,\cdot,m,t} - \boldsymbol{S}_{2,\cdot,m,t}\boldsymbol{Q}^{*,-1}_{\cdot,m,t}\boldsymbol{S}_{2,\cdot,m,t},\\
\boldsymbol{S}_{2,\cdot,m,t} &= \boldsymbol{R}^*_{2,\cdot,m,t}(\boldsymbol{F}^{(m-1)}_{1,\cdot,m,t}\boldsymbol{F}_2)',\\
\boldsymbol{m}_{1,f,m,t} &= \boldsymbol{a}_{1,f,m,t} + \boldsymbol{S}_{1,f,m,t}\boldsymbol{Q}^{*,-1}_{f,m,t}\left(\boldsymbol{f}_t^{(m-1)} - \boldsymbol{g}_{f,m,t}\right),\\
\boldsymbol{m}_{1,b,m,t} &= \boldsymbol{a}_{1,b,m,t} + \boldsymbol{S}_{1,b,m,t}\boldsymbol{Q}^{*,-1}_{b,m,t}\left(\boldsymbol{b}_t^{(m-1)} - \boldsymbol{g}_{b,m,t}\right),\\
\boldsymbol{C}^*_{1,\cdot,m,t} &= \boldsymbol{R}^*_{1,\cdot,m,t} - \boldsymbol{S}_{1,\cdot,m,t}\boldsymbol{Q}^{*,-1}_{\cdot,m,t}\boldsymbol{S}_{1,\cdot,m,t},\\
\boldsymbol{S}_{1,\cdot,m,t} &= \boldsymbol{R}^*_{1,\cdot,m,t}\left(\boldsymbol{F}^{(m-1)}_{1,\cdot,m,t}\right)'.
\end{aligned}$$

- Innovation variances $\sigma^2_{\cdot,m}$:

$$\sigma^2_{\cdot,m}|\boldsymbol{D}_{\cdot,m,t-1} \sim \mathcal{IG}\left(\frac{n_{\cdot,t-1}}{2}, \frac{h_{\cdot,t-1}}{2}\right),$$
$$\sigma^2_{\cdot,m}|\boldsymbol{D}_{\cdot,m,t} \sim \mathcal{IG}\left(\frac{n_{\cdot,t}}{2}, \frac{h_{\cdot,t}}{2}\right),$$

  where $n_t = n_{t-1} + n$, $h_{f,t} = h_{f,t-1} + (\boldsymbol{f}_t^{(m-1)} - \boldsymbol{g}_{f,m,t})'\boldsymbol{Q}^{-1,*}_{f,m,t}(\boldsymbol{f}_t^{(m-1)} - \boldsymbol{g}_{f,m,t})$ and $h_{b,t} = h_{b,t-1} + (\boldsymbol{b}_t^{(m-1)} - \boldsymbol{g}_{b,m,t})'\boldsymbol{Q}^{-1,*}_{b,m,t}(\boldsymbol{b}_t^{(m-1)} - \boldsymbol{g}_{b,m,t})$.

- Unconditional on $\sigma^2_{\cdot,m}$ we obtain Student-t distributions for $\boldsymbol{\theta}^{(m)}_{k,\cdot,m,t}|\boldsymbol{D}_{\cdot,m,t}$, $\boldsymbol{f}_t^{(m-1)}|\boldsymbol{D}_{f,m,t-1}$, and $\boldsymbol{b}_t^{(m-1)}|\boldsymbol{D}_{b,m,t-1}$.

After computing the filtering equations up to time T, it is possible to obtain the closed-form smoothing distributions for the forward and backward models as follows:

$$\boldsymbol{\theta}^{(m)}_{k,\cdot,m,t}|\boldsymbol{D}_{\cdot,m,T} \sim \mathcal{T}_{n_T}\left(\boldsymbol{m}_{k,\cdot,m,t|T}, \boldsymbol{C}_{k,\cdot,m,t|T}\right), k = 1, 2,$$

where

$$\begin{aligned}
\boldsymbol{m}_{k,\cdot,m,t|T} &= \boldsymbol{m}_{k,\cdot,m,t} + \boldsymbol{B}_{k,\cdot,m,t}\left(\boldsymbol{m}_{k,\cdot,m,t|T} - \boldsymbol{a}_{k,\cdot,m,t}\right),\\
\boldsymbol{C}_{k,\cdot,m,t|T} &= \frac{d_T n_t}{n_T d_t}\left(\boldsymbol{C}_{k,\cdot,m,t} - \boldsymbol{E}_{k,\cdot,m,t}\right),\\
\boldsymbol{E}_{k,\cdot,m,t} &= \boldsymbol{B}_{k,\cdot,m,t}\left(\boldsymbol{R}_{k,\cdot,m,t+1} - \frac{d_t n_T \boldsymbol{C}_{k,\cdot,m,t+1|T}}{n_t d_T}\right)\boldsymbol{B}'_{k,\cdot,m,t},\\
\boldsymbol{B}_{k,\cdot,m,t} &= \boldsymbol{A}'_{k,\cdot,m,t}\boldsymbol{R}^{-1}_{k,\cdot,m,t+1},\\
\boldsymbol{A}_{2,\cdot,m,t} &= \boldsymbol{C}_{2,\cdot,m,t}, \quad \boldsymbol{A}_{1,\cdot,m,t} = \boldsymbol{F}_2\boldsymbol{C}_{2,\cdot,m,t}\boldsymbol{G}_{\cdot,m,t},\\
\boldsymbol{G}_{\cdot,m,t} &= \left\{\left(\boldsymbol{I}_n - \boldsymbol{V}_{2,\cdot,m,t}\boldsymbol{F}'_{1,\cdot,m,t}\boldsymbol{V}^{\Delta,-1}_{2,\cdot,m,t}\boldsymbol{F}_{1,\cdot,m,t}\right)\boldsymbol{F}_2\right\}',\\
\boldsymbol{V}^{\Delta}_{2,\cdot,m,t} &= \boldsymbol{I}_n + \boldsymbol{F}_{1,\cdot,m,t}\boldsymbol{V}^*_{2,\cdot,m,t}\boldsymbol{F}'_{1,\cdot,m,t},
\end{aligned}$$

and initialized at $t = T$ with $\boldsymbol{m}_{k,\cdot,m,t|T} = \boldsymbol{m}_{k,\cdot,m,t}$ and $\boldsymbol{C}_{k,\cdot,m,t|T} = \boldsymbol{C}_{k,\cdot,m,t}$. Finally, the algorithm for posterior estimation is as follows.

Algorithm

1. Given hyperparameters $\{P, \delta_{k,f,m}, \delta_{k,b,m}\}$ for $m = 1, \ldots, P$, $k = 1, 2$, set $\boldsymbol{f}_t^{(0)} = \boldsymbol{b}_t^{(0)} = (x_{1t}, \ldots, x_{nt})'$, 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 (2.13) and (2.14), respectively, which, combined with structural equations (2.15) and (2.16), the random walk system equations (2.17) and (2.18) and the priors, define the PARCOR forward and backward hierarchical models. Then, use the sequential filtering equations along with the smoothing equations to obtain a series of estimated parameters $\{\hat{\boldsymbol{\theta}}_{k,f,1,t}^{(1)}\}$, $\{\hat{\boldsymbol{\theta}}_{k,b,1,t}^{(1)}\}$ for $t = 1 : T$. In addition, use the sequential filtering equations to obtain estimated $\hat{\sigma}^2_{\cdot,1}$. These estimated parameters are set at the posterior means of the smoothing distributions.

3. Use the observational equations (2.13) and (2.14) 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 2-3 above until $\{\hat{\boldsymbol{\theta}}_{k,f,m,t}^{(m)}\}$, $\{\hat{\boldsymbol{\theta}}_{k,b,m,t}^{(m)}\}$, $\{\hat{\sigma}^2_{f,m}\}$ and $\{\hat{\sigma}^2_{b,m}\}$ have been obtained for all $m = 1, \ldots, P$.

5. Finally, use $\{\hat{\boldsymbol{\theta}}_{1,f,m,t}^{(m)}\}$ and $\{\hat{\boldsymbol{\theta}}_{1,b,m,t}^{(m)}\}$, for $m = 1, \ldots, P$, as well as equations (1.13) and (1.14) in Chapter 1 Section 1.2.3 to obtain estimates of the forward and backward TVAR coefficients $\{\hat{a}_{i,j,t}^{(m)}\}$ and $\{\hat{d}_{i,j,t}^{(m)}\}$, for $i = 1, \ldots, n$, $j = 1, \ldots, m-1$, and $t = 1, \ldots, T$ via the Durbin-Levinson algorithm (Section 1.2.3 of Chapter 1).

6. (Optional) Use the estimated $\hat{\mu}_{f,m,t}^{(m)}$ and $\hat{\mu}_{b,m,t}^{(m)}$, for $m = 1, \ldots, P$, $t = 1, \ldots, T$, and equations (1.13) and (1.14) in Chapter 1 Section 1.2.3 to obtain the

forward and backward baseline TVAR coefficients $\{\bar{a}_{j,t}^{(P)}\}$ and $\{\bar{d}_{j,t}^{(P)}\}$, for $j = 1, \ldots, P$, and $t = 1, \ldots, T$ using the Durbin-Levinson algorithm (Section 1.2.3 of Chapter 1).

2.2.2 Model Selection

We develop an approach to select discount factors by maximizing log-likelihood functions derived from the one-step ahead predictive density functions for hiearchical PARCOR models. We start with a potential maximum value of P, say $P_{\max}$, for the model order. At level m we search for the optimal values of the forward discount factors $\delta_{k,f,m}$ and the backward discount factors $\delta_{k,b,m}$, for $k = 1, 2$. At level $m = 1$ we search for the combination of values of $\delta_{1,f,1}$ and $\delta_{2,f,1}$ maximizing the log-likelihood resulting from (2.13) with $m = 1$. Similarly, we can obtain the optimal combination of values of $\delta_{1,b,1}$ and $\delta_{2,b,1}$ by maximizing the log-likelihood resulting from (2.14). Using the selected optimal $\delta_{k,f,1}$ and $\delta_{k,b,1}$, 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 of the one-step ahead predictive density function for the forward model, which we denote as $\mathcal{L}_{f,1}$. 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 previously selected hyperparameters $\delta_{k,f,1}$ and $\delta_{k,b,1}$. We obtain optimal $\delta_{k,f,2}$, $\delta_{k,b,2}$, 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 $\{\delta_{k,f,m}, \delta_{k,b,m}, \mathcal{L}_{f,m}\}$, $m = 1, \ldots, P_{\max}$, has been selected. We then consider 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 $m = 1, \ldots, P_{\max}$.

Method 1: Scree plots. Yang et al. (2016) suggest to select the model order visually or numerically based on the values of $\mathcal{L}_{f,m}$ (Section 1.2.2 of Chapter 1). However, these methods do not work well in hierarchical settings, since they do not include a penalization for the number of model parameters.

Method 2: DIC model selection criterion. We consider an approach based on the deviance information criterion (DIC) to choose the model order (see Spiegelhalter et al., 2002, and references therein). In general, the deviance of a model given parameters $\boldsymbol{\theta}$ is defined as $D(\boldsymbol{\theta}) = -2\log p(\boldsymbol{y}|\boldsymbol{\theta})$, where $\boldsymbol{y}$ denotes the data. DIC is computed as $DIC = D(\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{DIC}$, where $\hat{\boldsymbol{\theta}}_{Bayes}$ is the Bayes estimator of $\boldsymbol{\theta}$ and p_{DIC} is the effective number of parameters. The effective number of parameters is given by $p_{DIC} = E_{post}\left(D\left(\boldsymbol{\theta}\right)\right) - D(\hat{\boldsymbol{\theta}}_{Bayes})$, where the expectation in the first 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

$$\hat{p}_{DIC} = \frac{1}{S}\sum_{s=1}^{S} D\left(\boldsymbol{\theta}^s\right) - D(\hat{\boldsymbol{\theta}}_{Bayes}).$$

In DLM settings, it is computationally intensive to compute the conventional deviance directly. Therefore, following Millar and McKechnie (2014), we compute the one-step-ahead DIC, which uses a pseudo deviance that conditions on the state at the previous time point, i.e.,

$$D^*\left(\boldsymbol{\theta}\right) = -2\log\prod_{t=1}^{T} p\left(\boldsymbol{y}_t|\boldsymbol{\theta}_{t-1}\right).$$

For a given model order m we compute the one-step ahead pseudo deviance 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 all previous $m-1$ stages. In other words, for each stage m :

- Compute the estimated implied log-likelihood from equation (2.13) for $t = 1, \dots, T$, using $\hat{\boldsymbol{\theta}}^{(m)}_{1,f,m,t}$ and $\hat{\sigma}^2_{f,m}$. In this way we obtain estimated $D^*(\hat{\boldsymbol{\theta}}_{Bayes})$ for model order m.

- Obtain samples, $\boldsymbol{\theta}^{(m)}_{2,f,m,t,s}$, for $s = 1, \dots, S$, from the sequential filtering equations with distributions and use these samples to compute the estimated number of parameters related only to stage m which we denote as $\hat{p}^m_{DIC,m}$. 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 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 $\widehat{DIC}_m$.

2.2.3 Posterior Summaries

Time-frequency representations summarized by estimates of the spectral densities can be obtained as follows. For each time series l and time t, the spectral density $g_l(t, \omega)$ can be estimated by plugging estimates $\hat{a}^{(P)}_{l,j,t}$ and $\hat{\sigma}^2_{f,P}$ into equation (1.5) in Section 1.1.1. Note that we can also compute the underlying baseline spectral density, which we denote by $\bar{g}(t, \omega)$, using estimates $\bar{a}^{(P)}_{j,t}, j = 1, \dots, P$ and $\hat{\sigma}^2_{f,P}$, benefiting from the hierarchical model structure.

Uncertainty measures for the estimates of the spectral density of each time series can be obtained by sampling from the filtering and smoothing posterior distributions of the forward and backward hierarchical PARCOR models. Each posterior sample of the model parameters is transformed via equation (1.5), leading to a posterior sample of the spectral density for time series l at time t. Uncertainty measures for these functions are then computed based on the samples.

2.3 Simulation Study

We illustrate our proposed approach in the analysis of simulated data. More specifically, we simulated 51 data sets, each with 5 time series of length $T = 1024$ from the following TVAR(2) model:

$$\begin{aligned}
y_{it} &= \phi_{i,1,t} y_{i,t-1} + \phi_{i,2,t} y_{i,t-2} + \epsilon_{it}, \epsilon_{it} \sim \mathcal{N}(0, 0.64), \\
\phi_{i,1,t} &= 2 r_t \cos\left(\frac{2\pi}{\lambda_{it}}\right), \quad \phi_{i,2,t} = -r_t^2, \quad r_t^2 = 0.9 \text{ for all } t, \\
\lambda_{it} &= \lambda_t + \gamma_i + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, 0.01), \quad \lambda_t = \frac{15t}{T} + 5,
\end{aligned}$$

where $\gamma_1 = \gamma_2 = \gamma_3 = 0$, $\gamma_4 = 1$, $\gamma_5 = 5$.

We fit hierarchical TV-PARCOR models to each of the 51 simulated data sets. We set a maximum model order $P_{\max} = 5$. The discount factors $\delta_{i,f,m}$ and $\delta_{i,b,m}$, for $i = 1, 2$ are chosen from a grid of values in $(0.99, 0.999)$. We set the prior hyperparameters to be $n_{f,0} = n_{b,0} = 1$, $h_{f,0} = h_{b,0} = 10$, $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = (0, 0, 0, 0, 0)'$ and $\boldsymbol{C}^*_{f,m,0} = \boldsymbol{C}^*_{b,m,0} = 10\boldsymbol{I}_5$. The left column of Figure 2.3 shows the true log spectral densities for 3 of the series, namely, $f_1(t, \omega)$, $f_4(t, \omega)$ and $f_5(t, \omega)$.

Figure 2.1 shows the BLF-scree plots obtained from the hierarchical PARCOR approach for each of the 51 data sets for model orders $m = 1, \dots, 5$. We can see that model order 2 is adequately chosen as the optimal model order, as after model order 2 the relative change of $\mathcal{L}_{f,m}$ is quite small. We also computed the

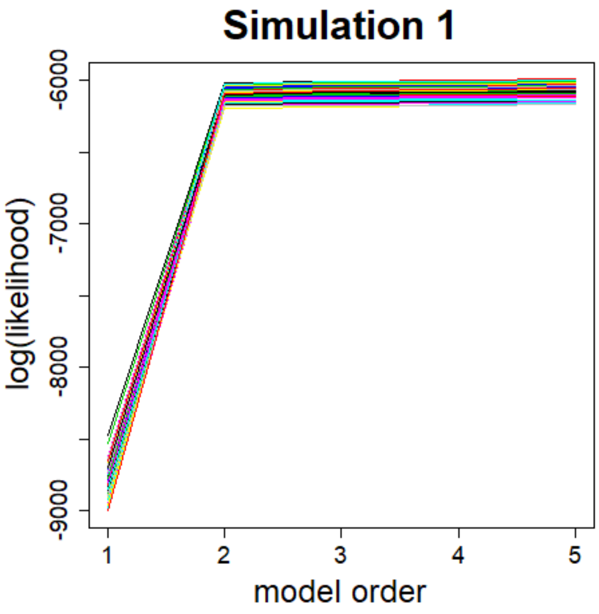

Figure 2.1: BLF-scree plots of the 51 realizations of simulation

DIC as explained in the previous section for each model order $m = 1, \ldots, 5$. DIC computations (not shown) also adequately identify 2 as the optimal model order.

To illustrate the performance of the hierarchical TV-PARCOR model, we also fit TVAR(2) models to each univariate time series in each data set to obtain some benchmark results. We computed the mean and stardard deviations of the average squared error (ASE) for each of the models and each of the five time series averaging over the 51 data sets. The ASE for each time series l is defined as follows (Ombao et al., 2001)

$$ASE_l = (TK)^{-1} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\log \hat{g}_{l,k}(t, \omega) - \log g_{l,k}(t, \omega) \right)^2 , \tag{2.19}$$

where $\omega \in [0, 0.5]$. Note that we have $K = 51$ data sets.

Table 2.1 summarizes the mean and standard deivations of the ASE based on ASE_l for each simulated time series. Our proposed model outperforms TVAR

Table 2.1: Mean ASE values and corresponding standard deviations (in parentheses) for the log-spectral densities obtained from TV-VPARCOR and TVAR models of order 2 for the TVAR(2) simulated data for $t = 1 : 1024$.

	Model	
Time Series (l)	TV-HPARCOR	TVAR
1	0.0857(0.0135)	0.1217(0.0279)
2	0.0882(0.0124)	0.1209(0.0293)
3	0.0899(0.0135)	0.1135(0.0304)
4	0.0779(0.0110)	0.1163(0.0319)
5	0.0649(0.0101)	0.1067(0.0395)

models for estimating log spectral densities.

Figure 2.3 summarizes posterior inference obtained from the hierarchical TV-PARCOR approach using model order of 2. Estimated spectral densities were obtained from the posterior means of the smoothing distributions of the forward and backward PARCOR coefficients over time. The estimated log-spectral densities displayed in the figures correspond to those that led to the median ASE. The hierarchical TV-PARCOR model clearly captures the structure of the individual spectral densities. In addition, a key feature of the hierarchical model is that it allows us to infer the latent/baseline log spectral density for all time series and compare it to the true baseline process used to generate the data sets. In this case the true baseline TVAR process has coefficients $\bar{\phi}_{1,t} = 2r_t \cos\left(\frac{2\pi}{\lambda_t}\right)$ and $\bar{\phi}_{2,t} = -r_t^2, r_t^2 = 0.9$ for all t. We can also obtain measures that quantify the uncertainty around the model estimates. Figure 2.2 shows the posterior inference obtained from the hierarchical TV-PARCOR model using the common underlying forward and backward effects across all the time series from a single data set. Once again, we see that the model adequately captures the baseline structure underlying the five simulated time series. Figure 2.4 shows the estimated 95% posterior interval of the log-spectral density of the first time series. The main structure of

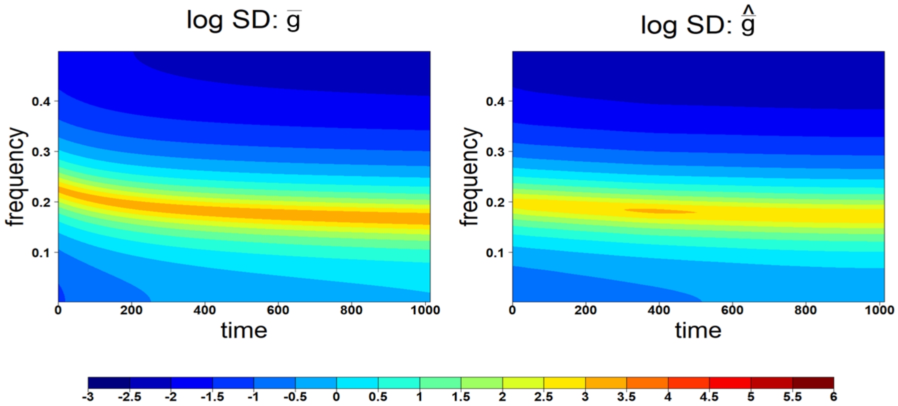

Figure 2.2: Left: True baseline of log-spectral density $\bar{g}(t,\omega)$. Right: Estimated baseline of log-spectral density $\hat{\bar{g}}(t,\omega)$.

the individual spectral density can be captured in the lower and upper bound. There is more uncertainty at the beginning of process.

2.4 Case Studies

2.4.1 Analysis of Group-level EEG Data

A key feature of the proposed hierarchical PARCOR model is that it can be used to detect a common underlying structure of multiple times series recorded in a setting that involves repeated trials. Here we analyze multiple EEG data recorded from subjects walking at or standing on a wide balance beam mounted to a treadmill. During the experiment, subjects were perturbed physically or visually. There were 30 healthy, young adults (15 females and 15 males, age 22.5±4.8 years) performing this experiment. Details regarding the data collection methodology and analysis via autoregressive models are available in Peterson and Ferris (2019). The data set used here is also publicly available (see references in Peterson and Ferris, 2019).

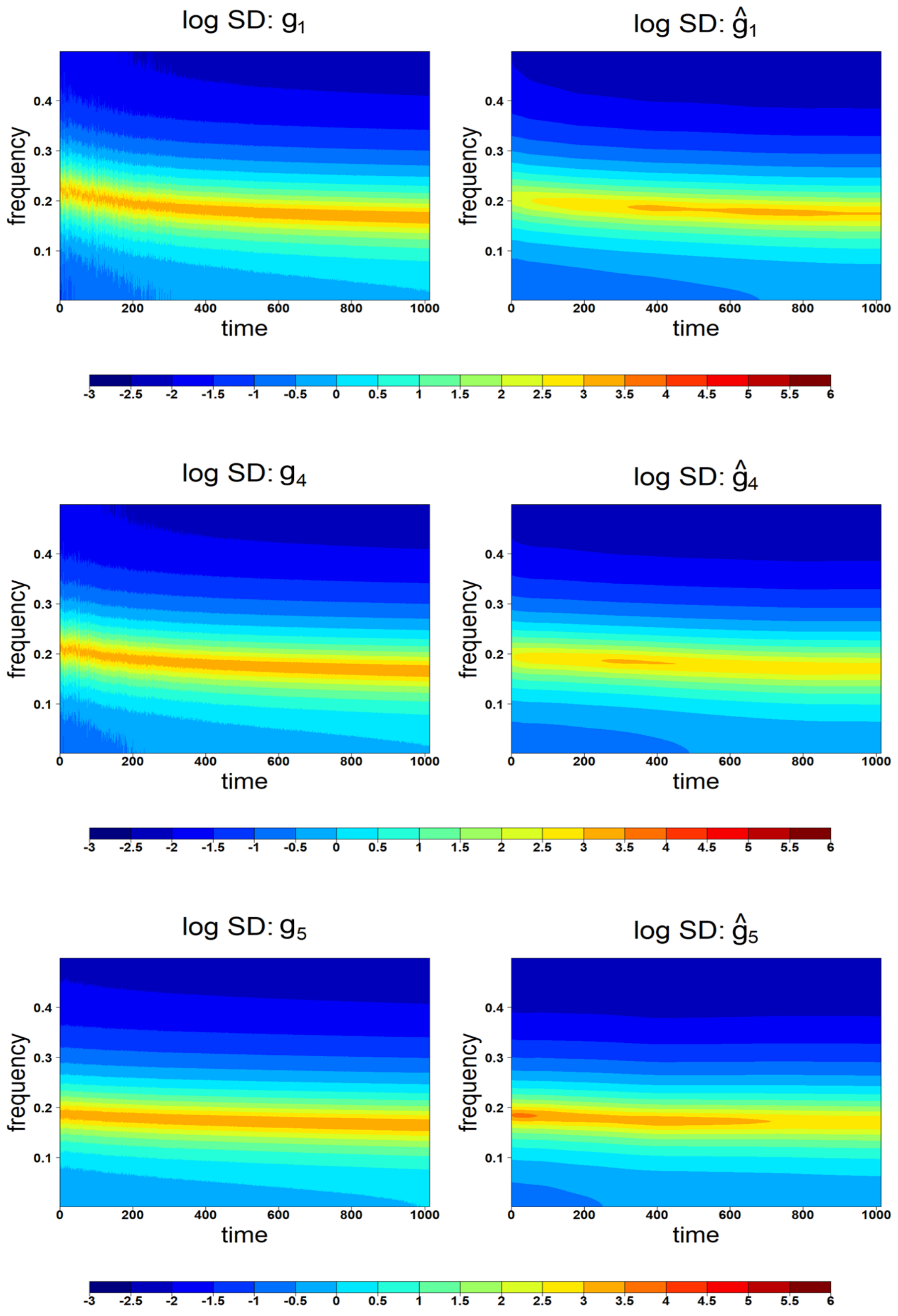

Figure 2.3: Left: True log-spectral densities $g_1(t,\omega)$, (top) $g_4(t,\omega)$ (middle) and $g_5(t,\omega)$ (bottom). Right: estimated log-spectral densities $\hat{g}_1(t,\omega)$, (top) $\hat{g}_4(t,\omega)$, (middle) and $\hat{g}_5(t,\omega)$ (bottom).

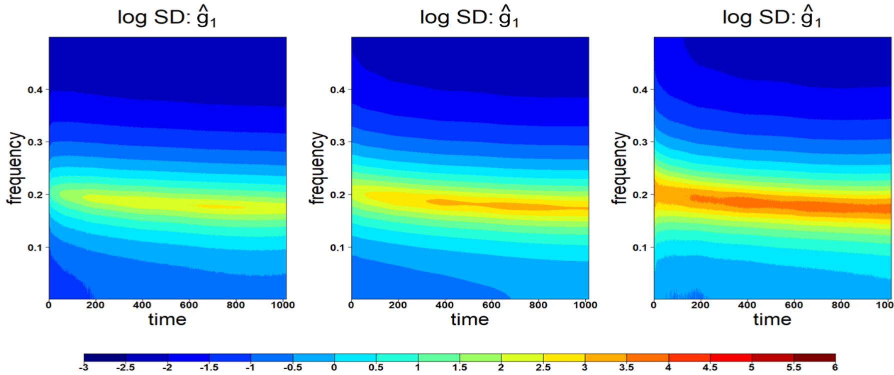

Figure 2.4: Left: Lower bound of a 95% posterior interval of the log-spectral densities $g_1(t, \omega)$. Middle: Estimated mean of the log-spectral densities $g_1(t, \omega)$. Right: Upper bound of a 95% posterior interval of the log-spectral densities $g_1(t, \omega)$,

The data contains 136-channel EEG recordings per subject with a sample rate of 512 Hz. After data pre-processing and independent component analysis, the information provided by the 136-channel EEG was summarized in terms of 8 cortical clusters. For each subject, each of the cortical regions, each experimental condition, and each trial within such condition, the data set contains time series that correspond to epochs going from -1s to 2s, centered around perturbation onset, leading to an average of 146 ± 1 epochs for stand pull, 145 ± 5 epochs for walk pull, 144 ± 9 epochs for stand rotate, and 146 ± 1 epochs for walk rotate (mean ± SD) for each subject.

We use the hierarchical PARCOR model to analyze data from Subject 25 that has the complete 8 cortical clusters and 146 epochs for each type of perturbation. We fit our hierarchical model on each cortical cluster under the physical stand pull perturbation. For each case, we considered a maximum model order $P_{max} = 10$ and discount factor values on a range of $(0.9, 0.99)$ (with equal spacing of 0.01) for structure and system levels. The initial parameters $n_{f,0}$ and $n_{b,0}$ are set to be 1 and $h_{f,0}$ and $h_{b,0}$ are set to be 10 for all m. In addition, we set the initial prior parameters as $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = \boldsymbol{0}$ and $\boldsymbol{C}^*_{f,m,0} = \boldsymbol{C}^*_{b,m,0} = 10\boldsymbol{I}_{146}$ for all m. The model orders selected by DIC are 6 ± 1 (mean ± SD) for different cortical clusters and perturbation types. The discount factors are mostly selected at 0.99, which suggests the time-varying coefficients change slowly over time.

Figure 2.5 shows the estimated relative change of the log power of spectral density (PSD) over time during the standing pull condition with respect to the estimated log power spectral density at time -0.5s for the same condition. In other words, we fit the hierarchical model to all the epochs for each cluster for the entire time period and summarize the results obtained in terms of the baseline estimated effects for each cluster. Instead of presenting the summaries in terms of the estimated log PSD at each time, we compute estimated log PSD at time -0.5s before perturbation onset then subtracted its value from the estimated log PSD at each of the remaining times to obtain estimates of the relative changes. Positive values (red) in the figure indicate increased spectral power compared to the baseline, while negative values (blue) indicate decreased spectral power with respect to the baseline. The vertical dashed lines show that the pull perturbation begins at 0 second and ends at 1 second. We present results for only three clusters, namely, supplementary motor, left sensorimotor and right sensorimotor. For the supplementary motor cluster, we observe increased spectral power at frequencies in the range of $4 - 8$ Hz after the pull perturbation onset. The spectral power returns to the baseline level after the perturbation offset at 1 second. For the left and right sensorimotor clusters, there is decreased spectral power occurring between 10 and 15 Hz after the pull perturbation onset. Similar cortical spectral fluctuation patterns can be found in Peterson and Ferris (2019). Our model found no clear activity patterns at low gamma power $(30 - 50$ Hz), either. We provide estimated mean and uncertainty measures of relative change of the log PSD of cluster supplementary motor in Figure 2.6. The increased spectral power

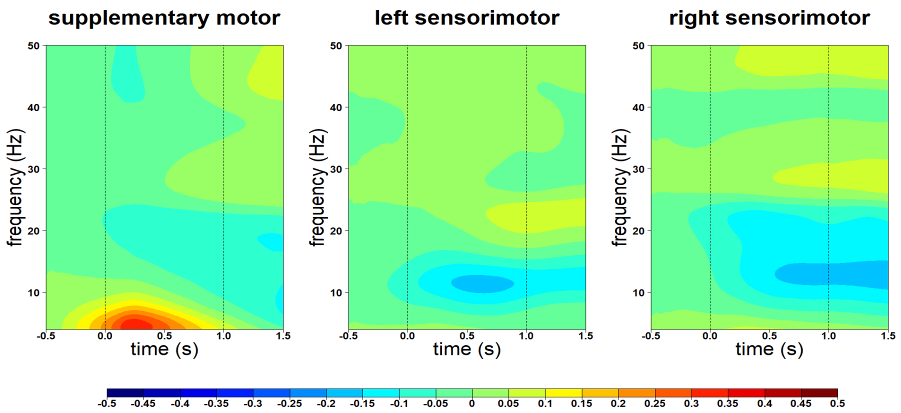

Figure 2.5: Cortical event-related spectral perturbations (ERSPs) to physical pulls.

at frequencies in the range of 4-8 Hz after the pull perturbation onset is also observed in the lower and upper uncertainty bands.

A key advantage of using the proposed hierarchical PARCOR model is that it jointly models all the 146 epochs corresponding to different trials for Subject 25, and is able to infer their latent time-frequency structure without having to average these signals over the different trials, and also without fitting different models individually to each of the series and then averaging the results from those models. We also note that posterior computations in the proposed PARCOR hierarchical approach are very fast, allowing us to jointly model 146 time series.

2.4.2 Analysis of Multi-Channel EEG Data

The hierarchical PARCOR model can also be used to capture common underlying features across different EEG channels recorded simultaneously on the same subject. 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 data set, code named `Ictal19`, analyzed in West et al. (1999) using univariate TVARs and in Prado et al. (2001) and using dynamic

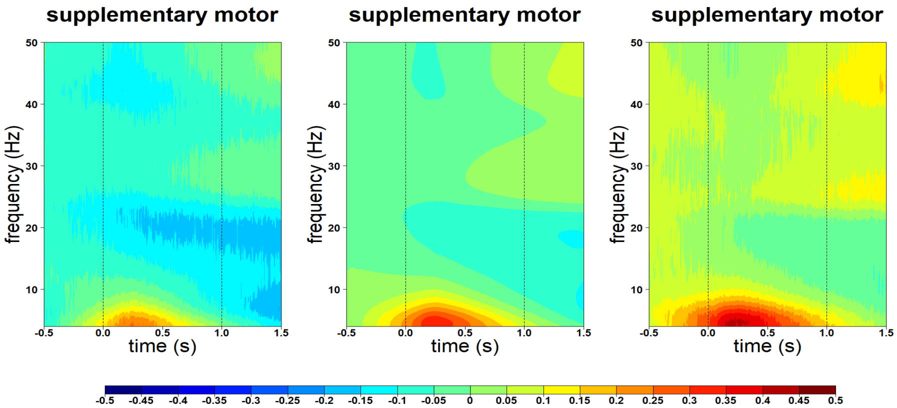

Figure 2.6: Left: Lower bound of a 95% posterior interval of the log-spectral density of supplementary motor cluster. Middle: Estimated mean of the log-spectral density of supplementary motor cluster. Right: Upper bound of a 95% posterior interval of the log-spectral density of supplementary motor cluster.

regression models. Nakajima and West (2017) presents an analysis of these data latent threshold TV-VAR models.

As an illustration, we use the hierarchical PARCOR model to analyze 9 channels, specifically channels F_3, F_z, F_4, C_3, C_z, C_4, P_3, P_z and P_4. shown in Figure 2.7. We chose these channels because they are closely located and because based on previous analyses we expect strong underlying similarities in their temporal structure over time. The original recordings of about $26,000$ observations per channel were sub-sampled every sixth observation from the highest amplitude portion of the seizure, leading to a set of time series of $3,600$ observations (corresponding to $84.375s$) per channel (Prado et al., 2001).

We analyzed the $K = 9$ series listed above jointly using a hierarchical PARCOR model. We considered a maximum model order $P_{max} = 15$ and discount factor values on a grid in the $(0.99, 0.999)$ range (with equal spacing of 0.001). We set $n_{f,0} = n_{b,0} = 1$ and $h_{f,0} = h_{b,0} = 10$ for all m. In addition, we set the same initial prior parameters $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = \mathbf{0}$ and $\boldsymbol{C}^*_{f,m,0} = \boldsymbol{C}^*_{b,m,0} = 10\boldsymbol{I}_9$. The optimal model order was found to be 11. The results shown in this section

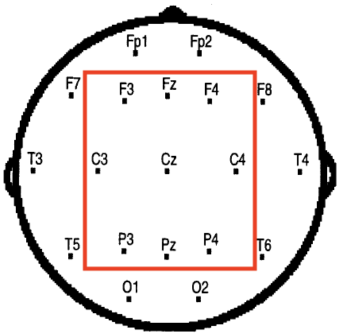

Figure 2.7: Representation of the `Ictal19` electrode placement. Here we focus on the nine channels in the region highlighted.

correspond to a hierarchical PARCOR model with this model order. Higher-order models were also fitted leading to similar but slightly smoother results in terms of the estimated spectral densities.

Figure 2.8 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 15 Hz). Each EEG channel shows a decrease in the dominant frequency over time, starting around 5 Hz and ending around approximate 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. In addition, the hierarchical PARCOR model allows us to obtain an estimate of the baseline spectral density underlying the 9 EEG channels. The estimated mean and 95% posterior interval of the baseline log-spectral density is shown in Figure 2.9 and provides a time-frequency summary of the features underlying the EEG time series.

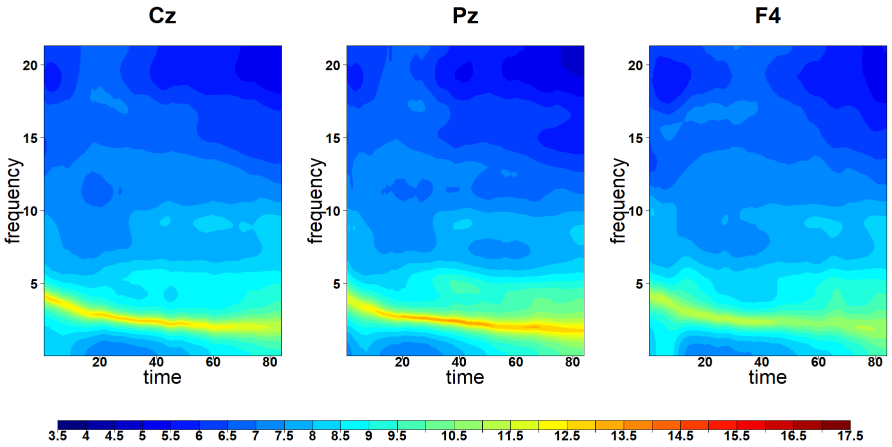

Figure 2.8: Estimated log-spectral densities for channels Cz, Pz and F4.

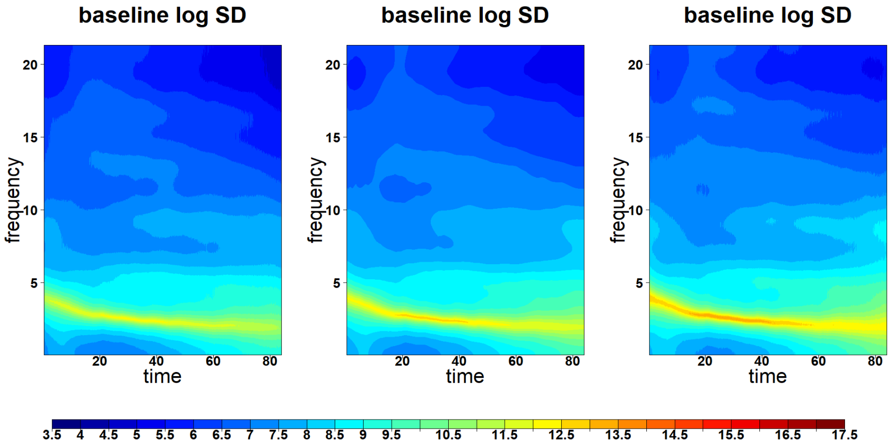

Figure 2.9: Left: Lower bound of a 95% posterior interval of the baseline log-spectral density. Middle: Estimated mean of baseline log-spectral densities. Right: Upper bound of a 95% posterior interval of the baseline log-spectral density.

2.5 Discussion

In this chapter, we present a dynamic hierarchical approach to model multiple times series in the PARCOR domain. This PARCOR model is more parsimonious than alternative time and frequency domain approaches often used in practice for the analysis of multiple time series, particularly for cases that involve modeling a relatively large number of time series with non-stationary time-dependency structure. We develop and implement algorithms for posterior inference in the hierarchical PARCOR setting that are computationally efficient and do not require the use of time consuming simulation-based approaches such as MCMC. We illustrate the performance of the proposed model and posterior inference algorithms in a simulation study and highlight its advantages in the analysis of two data sets consisting of multiple brain signals recorded under specific clinical/experimental conditions. The dynamic hierarchical structure in the PARCOR model allows us to infer the time-frequency characteristics of the individual time series, as well as those of their common underlying structure, which is of significant practical relevance in many practical settings, as illustrated in Section 2.4.

The proposed hierarchical PARCOR approach is computationally efficient and parismonious compared to state-space representations such as TV-VAR models. However, the proposed hierarchical PARCOR model cannot infer time-frequency relationships across multiple time series over time, such as time-varying coherence and partial coherence. In addition, a limited number of approaches are available for computationally efficient and scientifically interpretable analysis of multivariate non-stationary time series. In the next chapter, we consider extensions of the PARCOR framework for multivariate time series settings.

Chapter 3

Efficient Bayesian PARCOR Approaches for Dynamic Modeling of Multivariate Time Series

We develop a fast and accurate Bayesian lattice filtering and smoothing approach for multivariate non-stationary time series. The performance of the proposed time-varying vector PARCOR (or TV-VPARCOR) models is illustrated in simulation studies and in the analysis of multivariate non-stationary temporal data arising in neuroscience and environmental applications.

3.1 Introduction

In this chapter we extend the Bayesian lattice filter approach of Yang et al. (2016) explained in Section 1.2 of Chapter 1 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 (Prado et al., 2021, West and Harrison, 1997) 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 3.3 and 3.4. 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).

3.2 Models and Methods for Posterior Inference

3.2.1 TV-VAR Models and Lattice Structure

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

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

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

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,

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

$\boldsymbol{A}_{j,t}^{(P)}$ and $\boldsymbol{D}_{j,t}^{(P)}$ denote, respectively, the $K \times K$ time-varying matrices of forward and backward TV-VAR(P) coefficients for $j = 1, \dots, P$. Similarly, $\boldsymbol{A}_{j,t}^{(m)}$ and $\boldsymbol{D}_{j,t}^{(m)}$ denote the time-varying matrices of forward and backward TV-VAR(m) coefficients for $j = 1, \dots, m$.

Then, we write the m-stage of the lattice filter in terms of the pair of input-output relations between the forward and backward K-dimensional vector predictions, as follows,

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{\beta}_{f,m,t}^{(m)} \boldsymbol{b}_{t-m}^{(m-1)} + \boldsymbol{f}_t^{(m)}, \qquad \boldsymbol{f}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{f,m,t}), \tag{3.1}$$

$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{\beta}_{b,m,t}^{(m)} \boldsymbol{f}_{t+m}^{(m-1)} + \boldsymbol{b}_t^{(m)}, \qquad \boldsymbol{b}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{b,m,t}), \tag{3.2}$$

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

As mentioned in Section 1.2 of Chapter 1, the DLM formulation for univariate time series in the PARCOR domain fully avoids matrix computations, significantly improving computational efficiency, especially when the model order is relatively large. In the multivariate cases, the TV-VPARCOR approach requires matrix computations in the filtering and smoothing steps, however, it still offers significant computational savings when compared to using TV-VAR representations in the time domain. More specifically, the TV-VPARCOR approach results in DLM representations with state vectors of dimension K^2, with K the number of time series, leading to operations with matrices of dimension $K^2 \times K^2$. In contrast, a time-domain representation of a TV-VAR model requires state vectors of dimension K^2P, leading to operations with matrices of dimension $K^2P \times K^2P$, which results in much larger matrices when $P \geq 2$, discussed in Chapter 1 Section 1.1. For instance, for $K = 10$ and $P = 10$, the TV-VPARCOR approach requires operations with matrices of dimension 100×100 at each time step of the filtering and smoothing process, while the time-domain TV-VAR approach requires operations with matrices of dimension 1000×1000. Therefore, the TV-VPARCOR approach leads to significantly reduction computational time in many practical settings.

3.2.2 Whittle Algorithm

In the multivariate scenarios, the Whittle algorithm (Zhou, 1992) connects between TV-VAR coefficient matrices and TV-VPARCOR coefficient matrices similar as Durbin-Levinson algorithm (see Section 1.2.3). For each stage m of the lattice structure above, we obtain the forward and backward TV-VAR coefficient matrices, $\boldsymbol{A}_{m,t}^{(P)}$ and $\boldsymbol{D}_{m,t}^{(P)}$, from the time-varying forward and backward PARCOR coefficient matrices, $\boldsymbol{\beta}_{f,m,t}^{(m)}$ and $\boldsymbol{\beta}_{b,m,t}^{(m)}$ as follows,

$$\boldsymbol{A}_{j,t}^{(m)} = \boldsymbol{A}_{j,t}^{(m-1)} - \boldsymbol{A}_{m,t}^{(m)}\boldsymbol{D}_{m-j,t}^{(m-1)}, \tag{3.3}$$

$$\boldsymbol{D}_{j,t}^{(m)} = \boldsymbol{D}_{j,t}^{(m-1)} - \boldsymbol{D}_{m,t}^{(m)}\boldsymbol{A}_{m-j,t}^{(m-1)}, \quad j = 1, \dots, m-1, \tag{3.4}$$

with $\boldsymbol{A}_{m,t}^{(m)} = \boldsymbol{\beta}_{f,m,t}^{(m)}$ and $\boldsymbol{D}_{m,t}^{(m)} = \boldsymbol{\beta}_{b,m,t}^{(m)}$, for $m = 1, \dots, P$.

3.2.3 Model Specification and Inference

Our proposed model specification uses equations (3.1) and (3.2) as observational level equations of multivariate DLMs (Prado et al., 2021, West and Harrison, 1997) on the forward and backward PARCOR time-varying coefficients. These multivariate DLMs are specified as follows. For each t, let $\text{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)})$ and $\text{vec}(\boldsymbol{\beta}_{b,m,t}^{(m)})$ be the vectorized forward and backward PARCOR coefficients, i.e., these are K^2 vectors obtained by stacking the forward and backward PARCOR coefficient matrices at time t, $\boldsymbol{\beta}_{f,m,t}^{(m)}$ and $\boldsymbol{\beta}_{b,m,t}^{(m)}$, 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 (3.1) and (3.2) can be rewritten as

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{B}_{t-m}^{(m-1)} \text{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)}) + \boldsymbol{f}_t^{(m)}, \quad \boldsymbol{f}_t^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{f,m,t}), \tag{3.5}$$

$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{F}_{t+m}^{(m-1)} \text{vec}(\boldsymbol{\beta}_{b,m,t}^{(m)}) + \boldsymbol{b}_t^{(m)}, \quad \boldsymbol{b}_t^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{b,m,t}), \tag{3.6}$$

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 $\text{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)})$ and $\text{vec}(\boldsymbol{\beta}_{b,m,t}^{(m)})$, respectively. In order to complete the MDLM structure we specify random walk evolution equations for $\text{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)})$ and $\text{vec}(\boldsymbol{\beta}_{b,m,t}^{(m)})$ as follows,

$$\text{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)}) = \text{vec}(\boldsymbol{\beta}_{f,m,t-1}^{(m)}) + \boldsymbol{\epsilon}_{f,m,t}, \quad \boldsymbol{\epsilon}_{f,m,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W}_{f,m,t}), \tag{3.7}$$

$$\text{vec}(\boldsymbol{\beta}_{b,m,t}^{(m)}) = \text{vec}(\boldsymbol{\beta}_{b,m,t-1}^{(m)}) + \boldsymbol{\epsilon}_{b,m,t}, \quad \boldsymbol{\epsilon}_{b,m,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W}_{b,m,t}), \tag{3.8}$$

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

$$\mathrm{vec}(\boldsymbol{\beta}^{(m)}_{f,m,0})|\mathcal{D}_{f,m,0} \sim \mathcal{N}(\boldsymbol{m}_{f,m,0}, \boldsymbol{C}_{f,m,0}), \tag{3.9}$$

$$\mathrm{vec}(\boldsymbol{\beta}^{(m)}_{b,m,0})|\mathcal{D}_{b,m,0} \sim \mathcal{N}(\boldsymbol{m}_{b,m,0}, \boldsymbol{C}_{b,m,0}), \tag{3.10}$$

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

Given $\boldsymbol{\Omega}_{f,m,t}$, and $\boldsymbol{W}_{f,m,t}$ for all $t = 1, \ldots, T$, and all $m = 1, \ldots, P$, equations (3.5), (3.7) and (3.9) define a normal MDLM (see, e.g., Prado et al., 2021, Chapter 10) for the forward time-varying PARCOR. Similarly, given $\boldsymbol{\Omega}_{b,m,t}$ and $\boldsymbol{W}_{b,m,t}$ for all t and all m, equations (3.6), (3.8) and (3.10) 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 $\boldsymbol{W}_{f,m,t}$, and $\boldsymbol{W}_{b,m,t}$. We also assume $\boldsymbol{\Omega}_{f,m,t} = \boldsymbol{\Omega}_{f,m}$ and $\boldsymbol{\Omega}_{b,m,t} = \boldsymbol{\Omega}_{b,m}$ for all t, and use the approach of Triantafyllopoulos (2007) to obtain estimates of $\boldsymbol{\Omega}_{f,m}$ and $\boldsymbol{\Omega}_{b,m}$, 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

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

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 $k = 1, \ldots, K$, and then choose $\delta_{f,m}$ and $\delta_{b,m}$ optimally according to some criterion for each stage m (this is discussed in Section 3.2.4). This structure for $\boldsymbol{\Delta}_{f,m}$ and $\boldsymbol{\Delta}_{b,m}$ allows us to obtain closed form expressions for $\boldsymbol{W}_{f,m,t}$ and $\boldsymbol{W}_{b,m,t}$ 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 $\mathcal{D}_{f,m,t}$ denote all the information available up to time t at stage m for the forward model, with $\mathcal{D}_{f,m,t} = \{\mathcal{D}_{f,m,t-1}, \boldsymbol{f}_t^{(m-1)}\}$. Consider the posterior expectation of $\boldsymbol{\Omega}_{f,m}$ up to time t, i.e., $E(\boldsymbol{\Omega}_{f,m}|\mathcal{D}_{f,m,t})$, and assume that $\lim_{t\to\infty} E(\boldsymbol{\Omega}_{f,m}|\mathcal{D}_{f,m,t}) = \boldsymbol{\Omega}_{f,m}$. Let $n_{f,m,0}$ be a positive scalar and $\boldsymbol{S}_{f,m,0}$ be the prior expectation of $\boldsymbol{\Omega}_{f,m}$. Assume that at time $t-1$, we have that $\text{vec}(\boldsymbol{\beta}^{(m)}_{f,m,t-1})|\mathcal{D}_{f,m,t-1}$ is approximately distributed as $N(\boldsymbol{m}_{f,m,t-1}, \boldsymbol{C}_{f,m,t-1})$, and so, $E(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{f,m,t-1})$ is approximated by $\boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{m}_{f,m,t-1}$ and $V(\boldsymbol{f}_t^{(m-1)}|\mathcal{D}_{f,m,t-1})$ is approximated by $\boldsymbol{Q}_{f,m,t-1} = \boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{R}_{f,m,t}(\boldsymbol{B}_{t-m}^{(m-1)})' + \boldsymbol{S}_{f,m,t-1}$, with $\boldsymbol{R}_{f,m,t} = \boldsymbol{C}_{f,m,t-1} + \boldsymbol{W}_{f,m,t}$, for some $\boldsymbol{S}_{f,m,t-1}$. Then, following Theorem 1 of Triantafyllopoulos (2007) we have that, if $\boldsymbol{\Omega}_{f,m}$ is bounded, $\boldsymbol{S}_{f,m,t}$ will approximate $\boldsymbol{\Omega}_{f,m}$ for t large, with

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

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

In addition to showing that $\boldsymbol{S}_{f,m,t}$ converges to $\boldsymbol{\Omega}_{f,m}$ as t goes to infinity, Triantafyllopoulos (2007), presents empirical results that that aim to assess the quality of the approximation. In particular, a simulation study in Triantafyllopoulos (2007) generates 1000 time series with length 500 under three different scenarios and compares the approximation of $\boldsymbol{\Omega}_{f,m}$ given by $\boldsymbol{S}_{f,m,t}$ with estimates of $\boldsymbol{\Omega}_{f,m}$ obtained using a MCMC estimation procedure based on a blocked Gibbs sampler for state space models (Gamerman and Lopes, 2006) as a gold standard. The proposed covariance estimator has good performance, empirically converging to the true values of the unknown observation covariance matrix. Additional examples with TV-VAR processes are also considered. Our simulation studies also show that $\boldsymbol{S}_{f,m,t}$ closely approximates the true $\boldsymbol{\Omega}_{f,m}$.

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:

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

  and

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

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

- The one-step forecast error vector is given by $\boldsymbol{e}_{f,m,t} = \boldsymbol{f}_t^{(m-1)} - \boldsymbol{B}_{t-m}^{(m-1)} \boldsymbol{m}_{f,m,t-1}$.

- Using Bayes' theorem and the equations above we can obtain the approximate posterior distribution at time t as $\mathrm{vec}(\boldsymbol{\beta}_{f,m,t}^{(m)})|\mathcal{D}_{f,m,t} \approx \mathcal{N}(\boldsymbol{m}_{f,m,t}, \boldsymbol{C}_{f,m,t})$, where

$$\boldsymbol{m}_{f,m,t} = \boldsymbol{m}_{f,m,t-1} + \boldsymbol{U}_{f,m,t}\boldsymbol{e}_{f,m,t}, \tag{3.11}$$

$$\boldsymbol{C}_{f,m,t} = \boldsymbol{\Delta}_{f,m}\boldsymbol{C}_{f,m,t-1}\boldsymbol{\Delta}_{f,m} + \boldsymbol{U}_{f,m,t}\boldsymbol{Q}_{f,m,t}\boldsymbol{U}'_{f,m,t}, \tag{3.12}$$

$$\boldsymbol{U}_{f,m,t} = \boldsymbol{\Delta}_{f,m}\boldsymbol{C}_{f,m,t-1}\boldsymbol{\Delta}_{f,m}\boldsymbol{B}_{t-m}^{(m-1)}\boldsymbol{Q}_{f,m,t}^{-1}. \tag{3.13}$$

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

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

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

where the mean and covariance are computed recursively via

$$\begin{aligned} \boldsymbol{a}_{f,m,T}(t-T) &= \boldsymbol{m}_{f,m,t} - \boldsymbol{J}_{f,m,t}(\boldsymbol{a}_{f,m,t+1} - \boldsymbol{a}_{f,m,T}(t-T+1)), & (3.14)\\ \boldsymbol{R}_{f,m,T}(t-T) &= \boldsymbol{C}_{f,m,t} - \boldsymbol{J}_{f,m,t}(\boldsymbol{R}_{f,m,t+1} - \boldsymbol{R}_{f,m,T}(t-T+1)), & (3.15) \end{aligned}$$

for $t = (T-1), \dots, 1$, with $\boldsymbol{J}_{f,m,t} = \boldsymbol{C}_{f,m,t}\boldsymbol{R}_{f,m,t+1}^{-1}$, and starting values $\boldsymbol{a}_{f,m,T}(0) = \boldsymbol{m}_{f,m,T}$ and $\boldsymbol{R}_{f,m,T}(0) = \boldsymbol{C}_{f,m,T}$. 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 (3.1) and (3.2), respectively, which, combined with the random walk evolution equations (3.7) and (3.8), and the priors (3.9) and (3.10), define the multivariate PARCOR forward and backward models. Then, use the sequential filtering equations (3.11) to (3.13) to obtain the estimated $\{\boldsymbol{S}_{f,1,T}\}$ and $\{\boldsymbol{S}_{b,1,T}\}$. Use the sequential filtering equations (3.11) to (3.13) along with the smoothing equations (3.14) and (3.15) to obtain a series of estimated parameters $\{\text{vec}(\hat{\boldsymbol{\beta}}_{f,1,t}^{(1)})\}$, $\{\text{vec}(\hat{\boldsymbol{\beta}}_{b,1,t}^{(1)})\}$ for $t = 1 : T$. These estimated parameters are set at the posterior means of the smoothing distributions, i.e., the values in equation (3.14) for the forward case and a similar equation in the backward case.

3. Use the observational equations (3.1) and (3.2) 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 $2 - 3$ above until $\{\text{vec}(\hat{\boldsymbol{\beta}}_{f,m,t}^{(m)})\}$, $\{\text{vec}(\hat{\boldsymbol{\beta}}_{b,m,t}^{(m)})\}$, $\{\boldsymbol{S}_{f,m,T}\}$ and $\{\boldsymbol{S}_{b,m,T}\}$ have been obtained for all $m = 1, \ldots, P$.

5. Finally, use $\{\text{vec}(\hat{\boldsymbol{\beta}}_{f,m,t}^{(m)})\}$ and $\{\text{vec}(\hat{\boldsymbol{\beta}}_{b,m,t}^{(m)})\}$, for $m = 1, \ldots, P$, and equations (3.3) and (3.4) to obtain the forward and backward TV-VAR coefficient matrices via Whittle's algorithm (Section 3.2.2).

3.2.4 Model Selection

In order 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 equation (3.1) with $m = 1$. Using the selected optimal $\boldsymbol{\Delta}_{f,1}$ and $\boldsymbol{\Delta}_{b,1}$, 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, i.e., $m = 2$, 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 $\boldsymbol{\Delta}_{f,1}$ and $\boldsymbol{\Delta}_{b,1}$. We obtain optimal $\boldsymbol{\Delta}_{f,2}$, $\boldsymbol{\Delta}_{b,2}$ 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 $\{\boldsymbol{\Delta}_{f,m}, \boldsymbol{\Delta}_{b,m}, \mathcal{L}_{f,m}\}, m = 1, \ldots, P_{\max}$, 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 deviance information criterion (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 (3.5) for $t =$

$1, \ldots, T$, using $\mathrm{vec}(\hat{\boldsymbol{\beta}}^{(m)}_{f,m,t})$ and $\boldsymbol{S}_{f,m,T}$. In this way we obtain the first term in the calculation of the DIC for model order m.

- Obtain samples, $\mathrm{vec}(\boldsymbol{\beta}^{(m)}_{f,m,t,s})$, for $s = 1, \ldots, S$, from the approximate sequential filtering equations with distributions $\mathcal{N}(\boldsymbol{m}_{f,m,t}, \boldsymbol{C}_{f,m,t})$, and use these samples to compute the estimated number of parameters related only to stage m which we denote as $\hat{p}^{m}_{DIC,m}$. 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 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 $\widehat{DIC}_m$.

3.2.5 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. The estimated spectral density matrix $\hat{\boldsymbol{g}}(t, \omega)$ can be computed by plugging the estimated TV-VAR coefficients $\hat{A}^{(P)}_{m,t}$ and estimated covariance matrix $\boldsymbol{S}_{f,P,T}$ in equation (1.6) in Chapter 1 Section 1.1.1.

In addition, 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 3.2.3. 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 3.4.2.

3.2.6 Forecasting

In this section, we show how to obtain h-steps ahead forecasts. In order to have a non-explosive behavior in the forecasts, we assume the series is locally stationary in the future, i.e, $\boldsymbol{\beta}_{f,m,t}^{(m)} = \boldsymbol{\beta}_{b,m,t}^{(m)}$ at time $t = T+1, \ldots, T+h$. Then, the approximate h-steps ahead forecast posterior distribution of the PARCOR coefficients, with $h > 0$, is approximated as $(\boldsymbol{\beta}_{f,m,T+h}^{(m)} | \mathcal{D}_{f,m,T}) \approx \mathcal{N}(\boldsymbol{m}_{f,m,T}(h), \boldsymbol{C}_{f,m,T}(h))$, where

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

with $\boldsymbol{W}_{f,m,T+1} = \Delta_{f,m} \boldsymbol{C}_{f,m,T} \Delta_{f,m} - \boldsymbol{C}_{f,m,T}$, for $m = 1, \ldots, P$. Then, we apply Whittle's algorithm (Section 3.2.2) to transform the PARCOR coefficients, $\boldsymbol{\beta}_{f,P,T+h}^{(P)}$, into TV-VAR coefficients $\boldsymbol{A}_{j,T+h}^{(P)}$ and $\boldsymbol{D}_{j,T+h}^{(P)}$, for $j = 1, \ldots, P$. Finally, we obtain the h-steps ahead forecasts using

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

3.3 Simulation Studies

In this section 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.

3.3.1 Bivariate TV-VAR(2) Processes

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

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

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 $\phi_{2,1,2,t} = r_{4,t}$.

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 the true squared coherence between the first and second component $\rho_{12}^2(t, \omega) = 0$ for all t and ω. Figure 3.1 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 3.4 and 3.5.

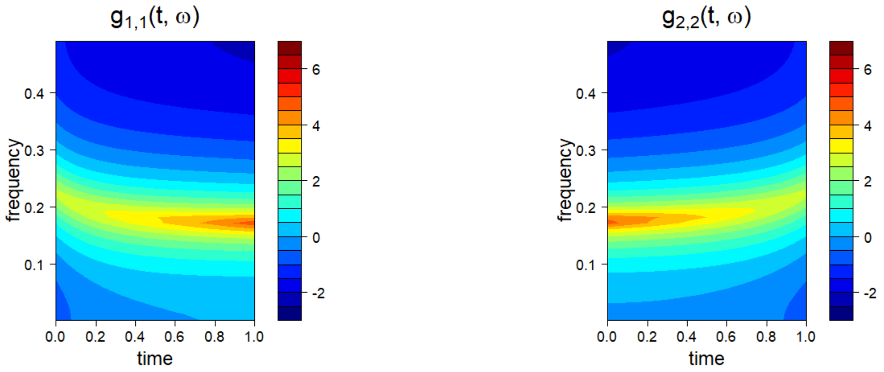

Figure 3.1: Case $\phi_{1,1,2} = 0$. Left: True log spectral density $g_{11}(t, \omega)$. Right: True log spectral density $g_{22}(t, \omega)$.

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, i.e., $t = 1025 : 1034$. 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 $n_{f,m,0} = n_{b,m,0} = 1$, $\boldsymbol{S}_{f,m,0} = \boldsymbol{S}_{b,m,0} = \boldsymbol{I}_2$, $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = (0, 0, 0, 0)'$ and $\boldsymbol{C}_{f,m,0} = \boldsymbol{C}_{b,m,0} = \boldsymbol{I}_4$. 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 $m = 1, \ldots, 5$. 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, in order 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).

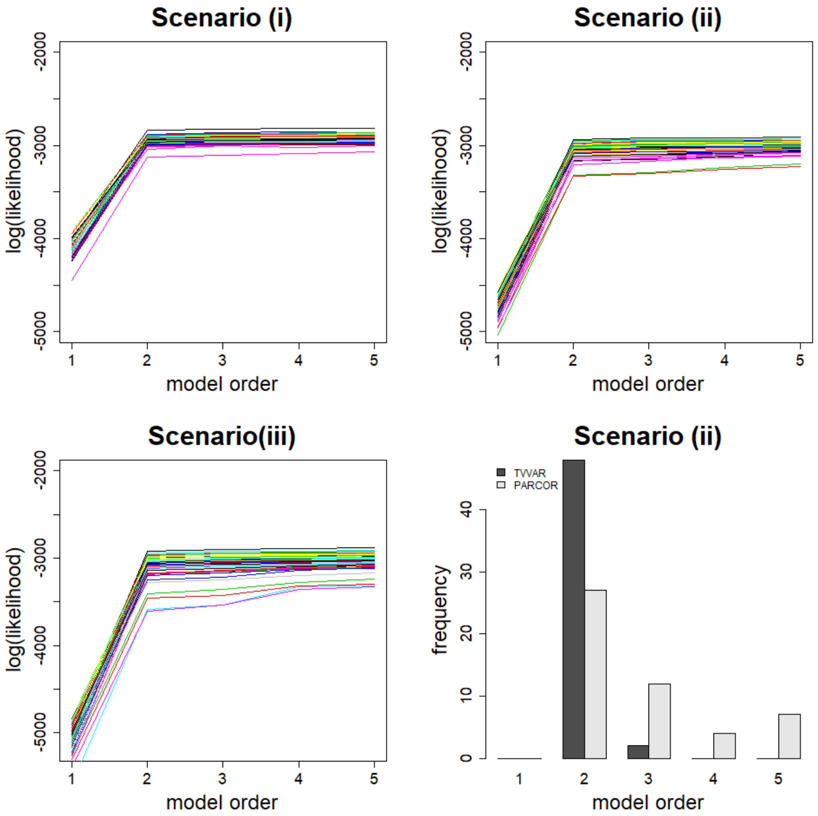

Figure 3.2: Top: BLF-scree plots of the 50 realizations of the for scenarios (i) and (ii). Bottom: BLF-scree plot for scenario (iii) and optimal model orders for scenario (ii).

Figure 3.2 shows the BLF-scree plots obtained from the PARCOR approach for each of the 50 datasets under the three scenarios for model orders $m = 1, \ldots, 5$. 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 $m = 1, \ldots, 5$ and each dataset under the three scenarios. The bottom right plot in Figure 3.2 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 3.3, 3.4 and 3.5, 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 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)).

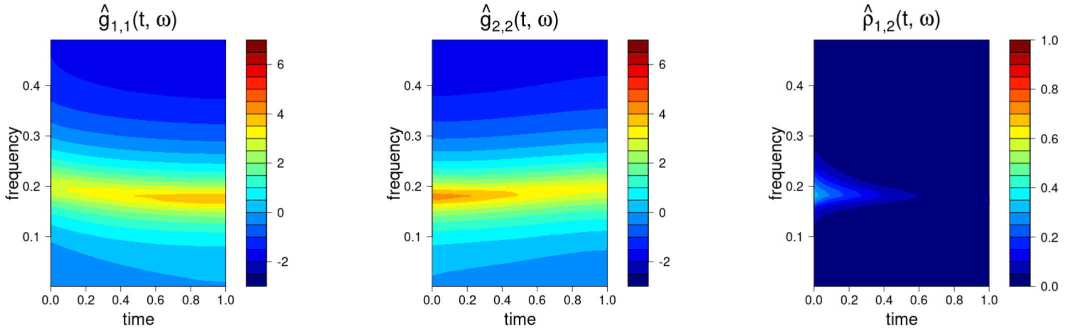

Figure 3.3: Case with $\phi_{1,1,2,t} = \phi_{2,1,2,t} = 0$. Left: Estimated average log spectral density of the first component. Middle: Estimated average log spectral density of the second component. Right: Estimated average squared coherence.

In order 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 average squared error (ASE) for each of

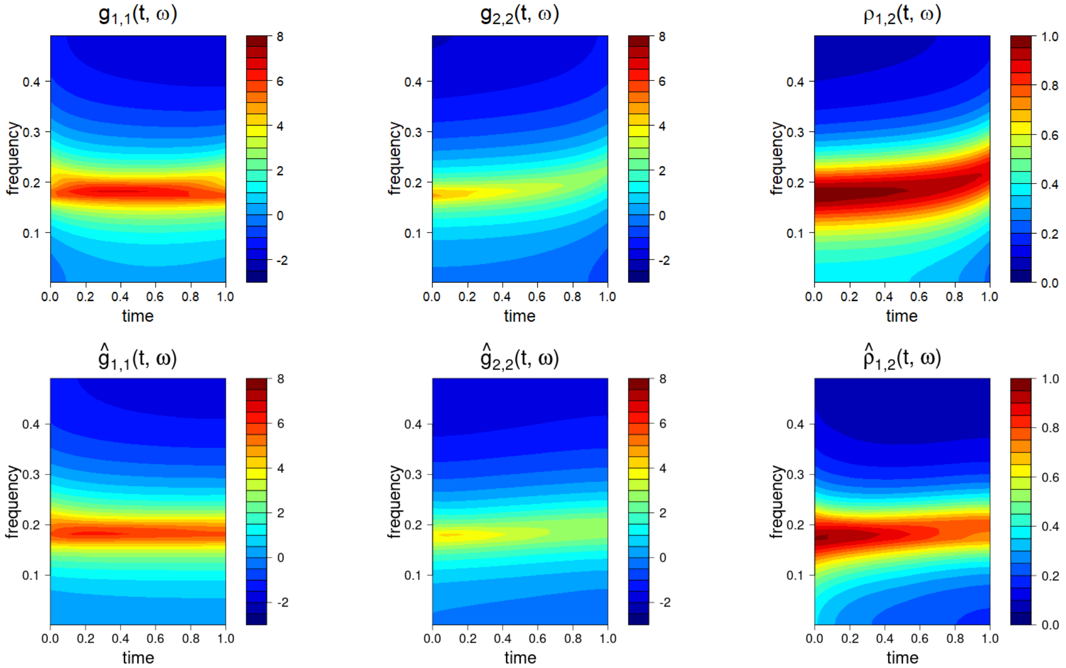

Figure 3.4: Case with $\phi_{1,1,2,t} = -0.8$ and $\phi_{2,1,2,t} = 0$ for all t. Top: True log spectral density $g_{1,1}(t, \omega)$ (left), true log spectral density $g_{2,2}(t, \omega)$ (middle), true squared coherence $\rho^2_{1,2}(t, \omega)$ (right). Bottom: Estimated $\hat{g}_{1,1}(t, \omega)$ (left), estimated $\hat{g}_{2,2}(t, \omega)$ (middle), estimated $\hat{\rho}^2_{1,2}(t, \omega)$ (right).

the models in each of the three simulation scenarios. Similar to equation (2.19) in Chapter 2 Section 2.3, the ASE of spectral density matrix is defined as follows (Ombao et al., 2001)

$$\mathrm{ASE}_n = (TL)^{-1} \sum_{t=1}^{T} \sum_{l=1}^{L} \left(\log \hat{\boldsymbol{g}}(t, \omega_l) - \log \boldsymbol{g}(t, \omega_l)\right)^2, \tag{3.16}$$

where $\omega_l = 0, 0.001, 0.011, \ldots, 0.5$. Note that we have $n = 50$ simulated data sets for each of the three scenarios. Table 3.1 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

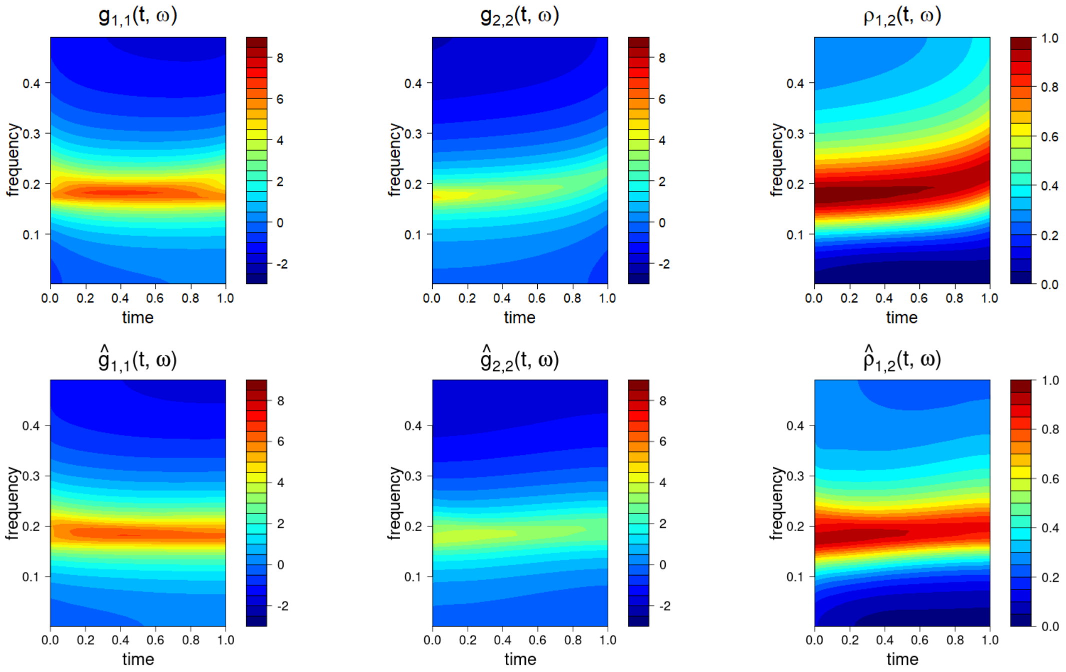

Figure 3.5: Case with $\phi_{1,1,2,t} = r_{3,t}$ and $\phi_{2,1,2,t} = r_{4,t}$. Top: True log spectral density $g_{1,1}(t, \omega)$ (left), true log spectral density $g_{2,2}(t, \omega)$ (middle), true squared coherence $\rho^2_{1,2}(t, \omega)$ (right). Bottom: Estimated $\hat{g}_{1,1}(t, \omega)$ (left), estimated $\hat{g}_{2,2}(t, \omega)$ (middle), estimated $\hat{\rho}^2_{1,2}(t, \omega)$ (right).

comparable performance in terms of estimating the time-frequency characteristics of the original process while being computationally more efficient. In fact, Table 3.2 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.

	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 3.1: 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 $t = 1 : 1024$.

Finally, Table 3.3 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

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

Table 3.2: Computation times (in seconds) for TV-VPARCOR and TV-VAR models.

parameters are non-zero and varying over time.

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

Table 3.3: 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.

3.3.2 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:

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

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

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 585s of computation time for $P_{\max} = 3$, while the TV-VAR model required 3379s 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 3.6 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

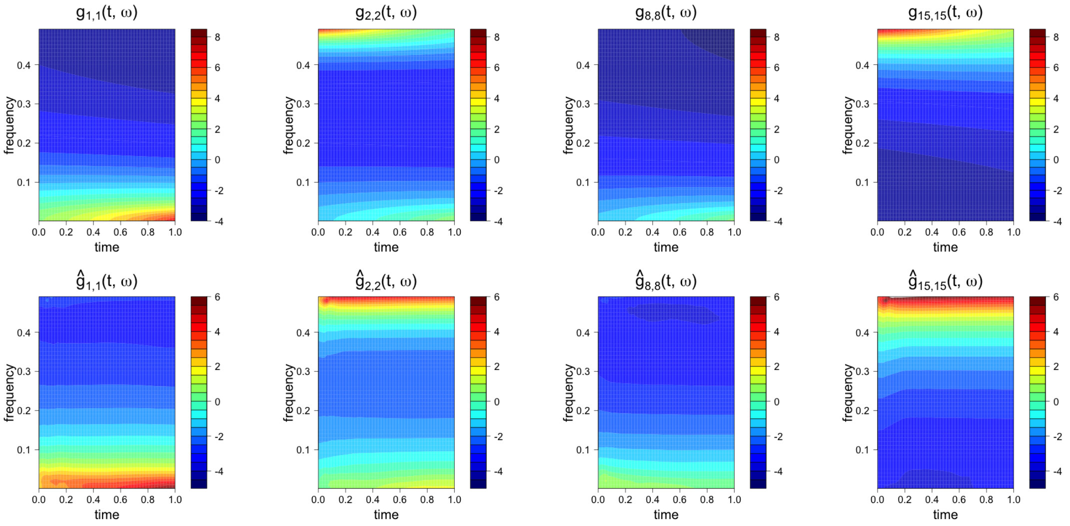

Figure 3.6: Top: True log spectral densities of time series components 1, 2, 8 and 15. Bottom: estimated log spectral densities of the same components obtained from the PARCOR approach with model order 1.

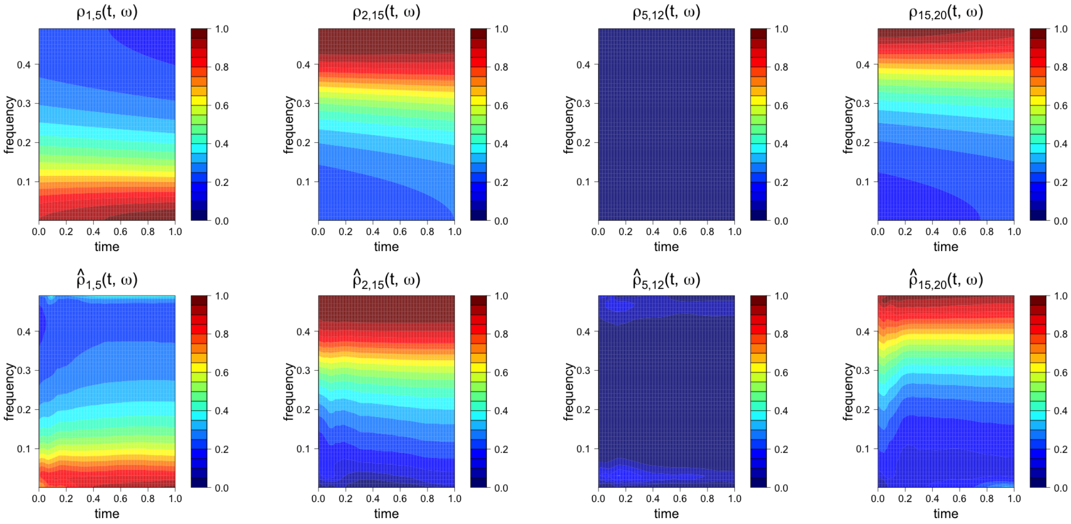

Figure 3.7: Top: True coherence between components 1 and 5, 2 and 15, 5 and 12, and 15 and 20. Bottom: Corresponding estimated coherences obtained from the PARCOR model.

and 15. Figure 3.7 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 A.1 in the Supplementary Material).

3.3.3 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 3.4.1 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) 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 3.8 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 3.9 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.

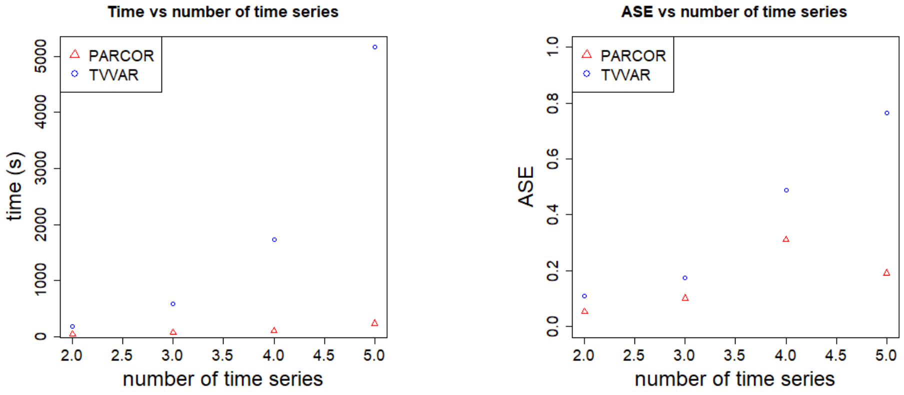

Figure 3.8: Left plot: Running time against number of time series. Right plot: estimated ASE against number of time series.

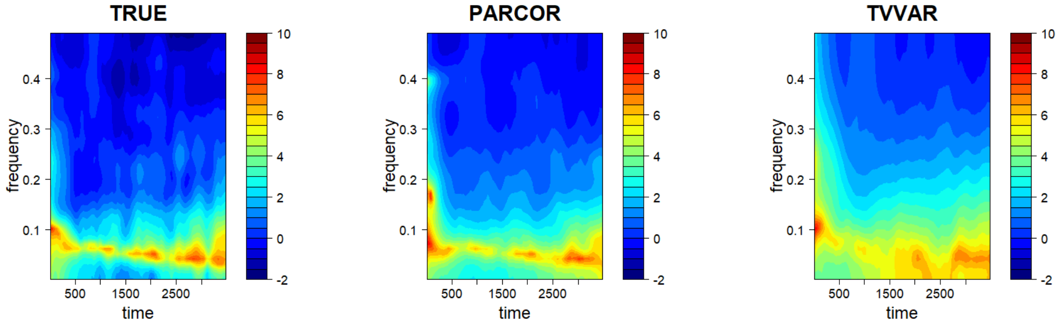

Figure 3.9: Left: True spectral density of the first time series in the first simulation. Middle: PARCOR estimated spectral density of the first time series. Right: TV-VAR estimated spectral density of the first time series.

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 to 10 to the EEG data and then using the estimated parameters from these models to simulate the data. Figure 3.10 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 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.

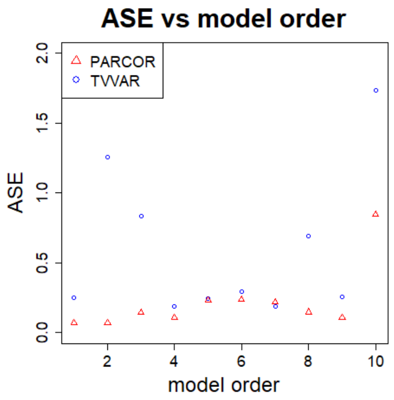

Figure 3.10: ASE values against model order.

Finally, we also considered a study in which the data was 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 Section A.2 of Supplementary Material.

3.4 Case Studies

3.4.1 Analysis of multi-channel EEG data

In this section, we jointly analyzed the same $K = 9$ of the EEG channels as Section 2.4.2 of Chapter 2 by 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 $n_{f,m,0} = n_{b,m,0} = 1$, and $S_{f,m,0} = S_{b,m,0} = 2000\boldsymbol{I}_9$ for all m. In addition, we set the same initial prior parameters $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = \boldsymbol{0}$ and $\boldsymbol{C}_{f,m,0} = \boldsymbol{C}_{b,m,0} = 1000\boldsymbol{I}_{81}$. The computation time to run the search for the optimal model with $P_{\max} = 20$ in this data set was 1,142 seconds in an Inter(R) Xeon(R) server with CPU E5-4650 with 2 cores and 2.70GHz. The optimal model order was found to be 5 (see Figure A.4 in the Supplementary Material) and so, the results presented here correspond to a TV-VPARCOR 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 3.11 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 18Hz). Furthermore, each EEG channel shows a decrease in the dominant frequency over time, starting around 5Hz and ending around approximately 3Hz. 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 chan-

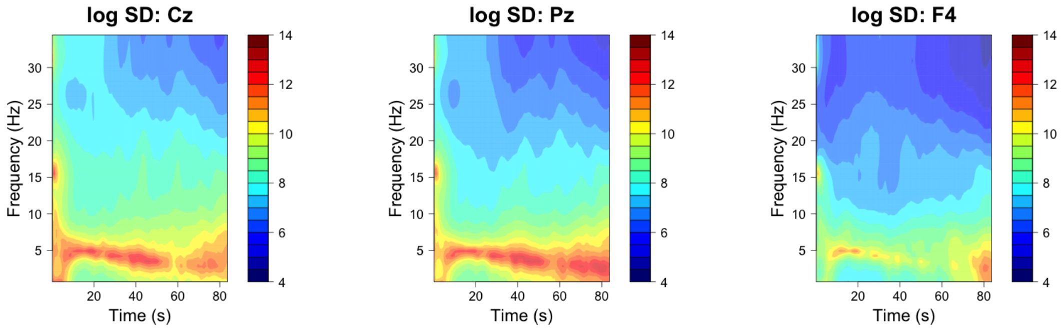

Figure 3.11: Estimated log-spectral densities for channels Cz, Pz and F4.

nels 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 3.12 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 35Hz. 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 10-15s, and approximately until about 50s, there is a strong coherence between F4 and Pz and Cz only at the dominant frequency of 3-5Hz that dissipates towards the end of the seizure. The partial coherence across pairs of 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 3.12 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 5Hz 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).

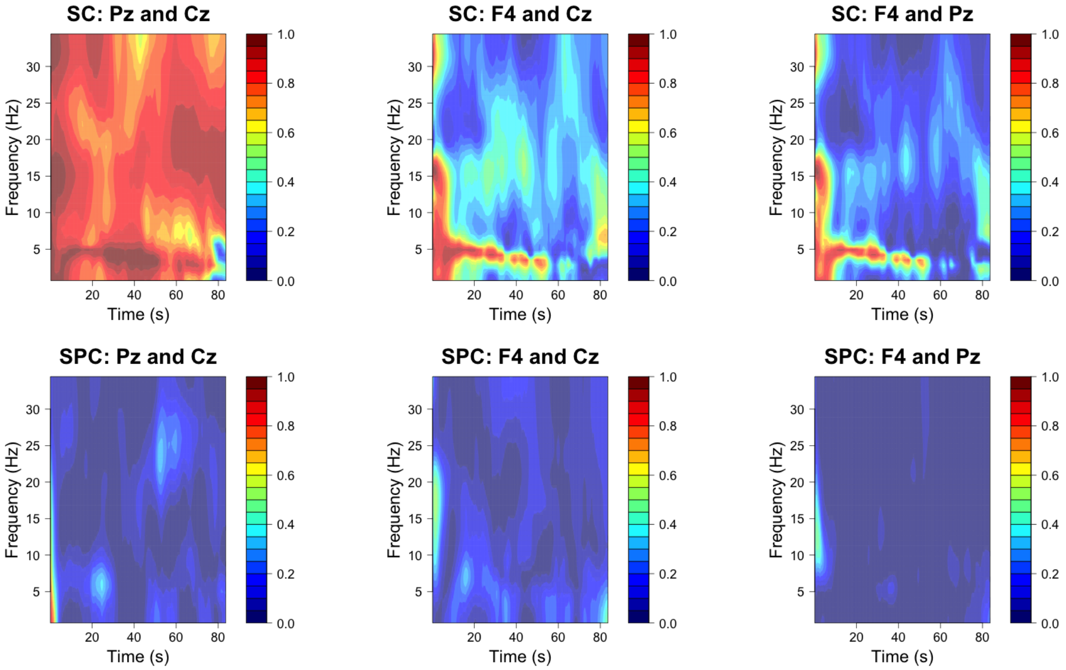

Figure 3.12: Top plots: Squared coherence between Pz and Cz, F4 and Cz, and F4 and Pz, respectively. Bottom plots: Squared partial coherence between Pz and Cz, F4 and Cz, and F4 and Pz.

Finally, we also estimated direct transfer functions and partial directed coherence between channels as explained in Chapter 1 Section 1.1.1. Figure 3.13 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.

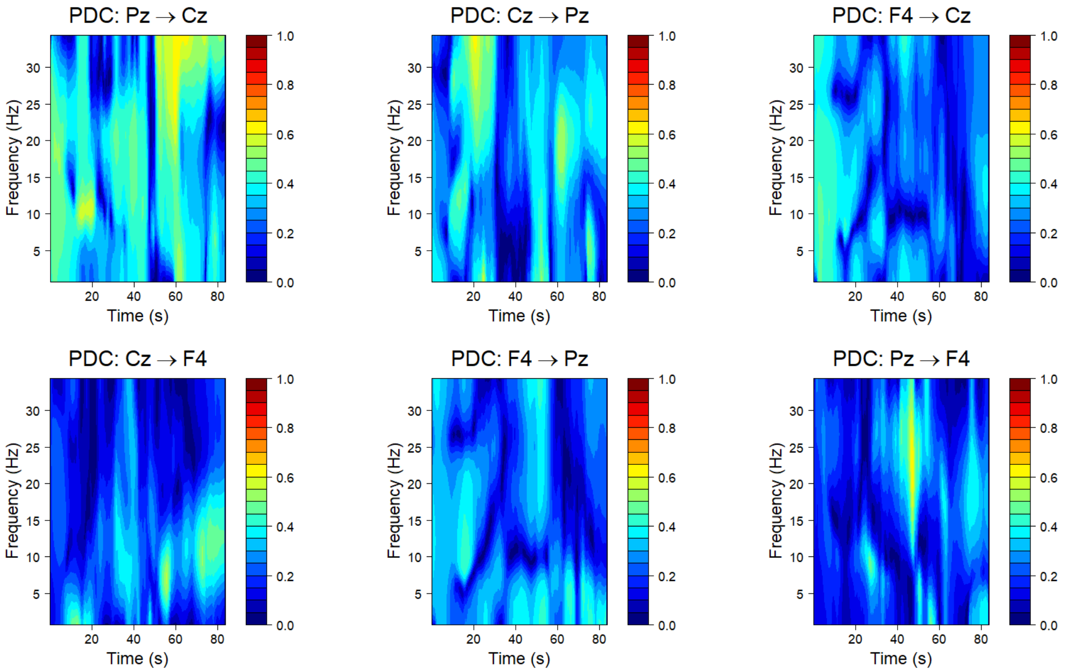

Figure 3.13: Estimated partial directed coherence among the three channels Pz, Cz and F4.

3.4.2 Analysis of Multi-Location Wind Data

We analyze wind component data derived from median wind speed and direction measurements taken every 4 hours from June 1st 2010 to August 15th in 3 stations in Northern California. These data were obtained from the Iowa Environmental Mesonet (IEM) Automated Surface Observing System (ASOS) Network, a publicly available database (see `http://mesonet.agron.iastate.edu/ASOS/`). ASOS stations are located at airports and take observations and basic reports from the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). For additional information about the ASOS measurements see NOA (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: $n_{f,m,0} = n_{b,m,0} = 1$, and $S_{f,m,0} = S_{b,m,0} = 5\boldsymbol{I}_6$, $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = \boldsymbol{0}$ and $\boldsymbol{C}_{f,m,0} = \boldsymbol{C}_{b,m,0} = 10\boldsymbol{I}_{36}$ for all m. The computation time to run the search for the optimal model with $P_{\max} = 10$ in this data set was 35.72 seconds in an Inter(R) Xeon(R) server with CPU E5-4650 with 2 cores and 2.70GHz. The optimal model order chosen by the approximate DIC calculation is $P = 3$ (see Figure A.5 in the Supplementary Material). 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 3.14 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 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.

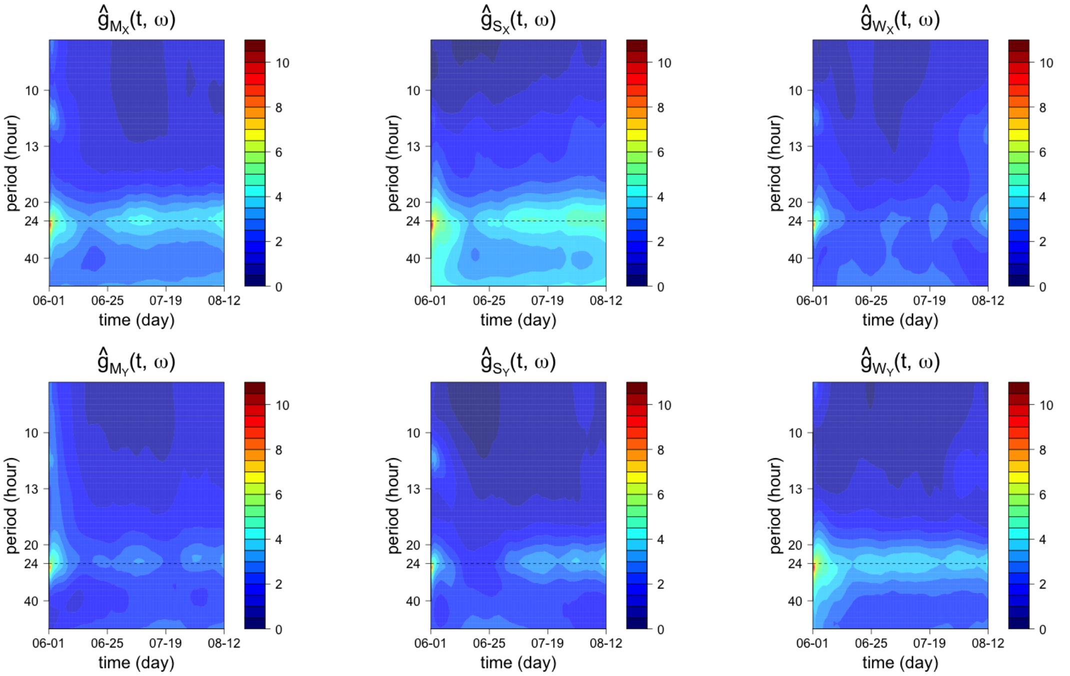

Figure 3.14: Top row: Estimated log-spectral densities of the East-West (X) components for Monterey, Salinas and Watsonville. Middle row: Estimated log-spectral densities of the North-South (Y) components for Monterey, Salinas and Watsonville.

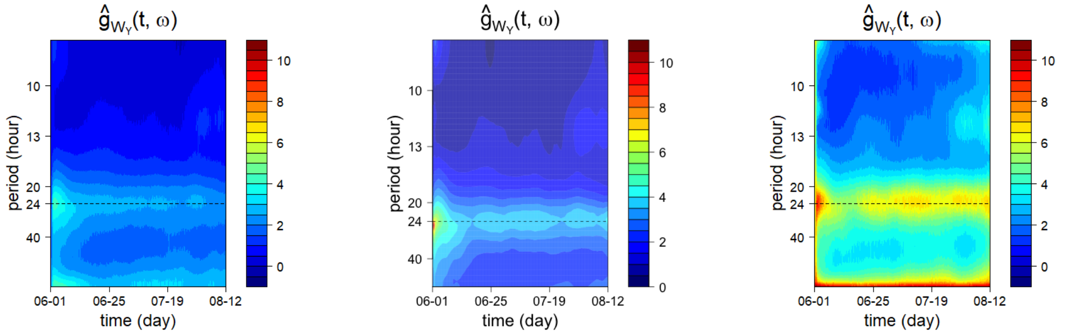

Figure 3.15: Left: Lower bound of a 95% posterior interval of the log-spectral density of the North-South (Y) component in Watsonville. Middle: Estimated mean log-spectral density of the North-South (Y) component in Watsonville. Right: Upper bound of 95% posterior interval of the log-spectral density of the North-South (Y) component in Watsonville.

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 3.15. 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 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 3.16 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 of Monterey and Salinas). Furthermore, the estimated squared partial coherence (see Figure A.6 in the Supplementary Material) 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 3.2.6. Figure 3.17 shows 48 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.

3.5 Discussion

In this chapter, 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 average squared error between the true and estimated

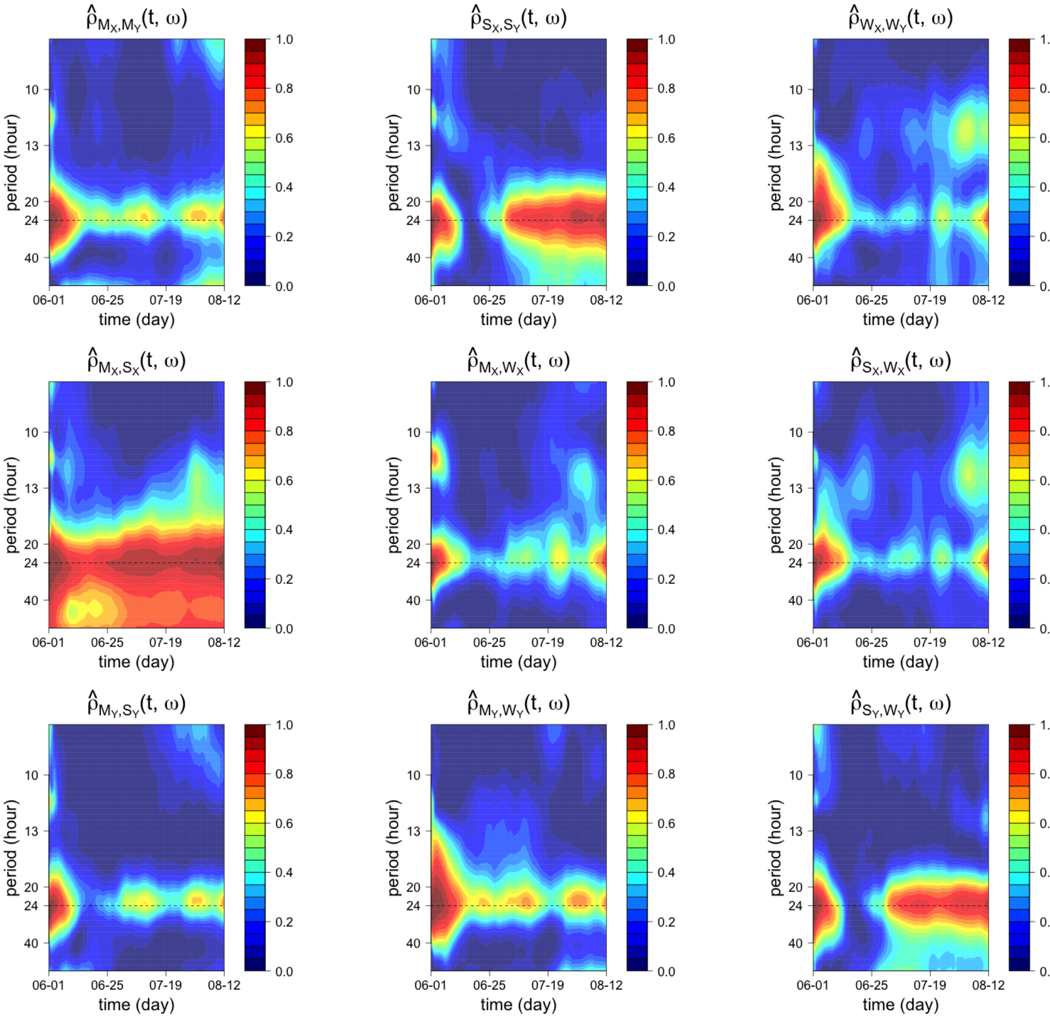

Figure 3.16: Top row: Estimated squared coherences between the East-West (X) component and North-South (Y) component in Monterey, Salinas and Watsonville. Middle row: Estimated squared coherences between the East-West (X) components of Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville. Bottom row: Estimated squared coherences between the North-South components in Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville.

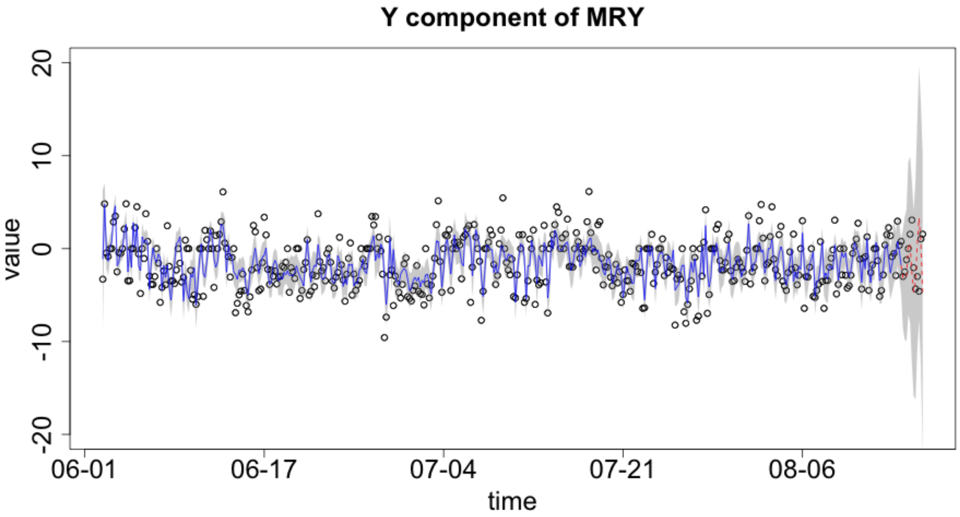

Figure 3.17: Observed Monterey North-South (Y component) Wind Component (dots), smoothed estimates obtained from the posterior mean values (solid blue line) and corresponding 90% bands (gray shade), 48 hours forecast (dotted red line) and corresponding 90% bands (gray shade).

time-varying spectral densities as shown in extensive simulation studies and in the analysis of two multivariate time series data sets. 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.

We note that the proposed hierarchical time-varying PARCOR and the TV-VPARCOR approaches developed in Chapters 2 and 3, respectively, are computationally efficient compared to standard state-space representations of TV-VAR models. However, they still present issues of overfitting, as in many practical settings we may expect sparsity in the model parameters or have situations in which some of parameters should remain constant over the entire observation period. In addition, in some large-dimensional settings, the proposed DIC under PARCOR framework (Section 3.2.4) may tend to select lower model order because of the over-penalization on the number of parameters. In order to solve these issues, we consider model extensions that place shrinkage priors on the time-varying PARCOR coefficients allowing for automatical selection of the model order, inducing shrinkage on the time-varying paramters towards zero or some other constant value over time. In Chapter 4 we develop and evaluate a new class of shrinkage TV-VPARCOR models.

Chapter 4

Multivariate Time Series Analysis and Forecasting via Time-Varying PARCOR Shrinkage Models

In this chapter, we induce a shrinkage structure in the TV-VPARCOR framework to address the overfitting issues mentioned in Section 3.5. After fully specifying the model structure we present the details of MCMC and ISVB algorithms for exact and approximate posterior inference, respectively. The performance of the proposed shrinkage TV-VPARCOR model is illustrated in simulation studies and in the analysis of multi-location wind component data.

4.1 Introduction

We propose a flexible and computationally efficient approach for analysis and forecasting of multi-dimensional time series that has the following features. First, following Chapter 3, we use a modeling framework in the partial autocorrelation (PARCOR) domain rather than a time-domain approach. This results in more parsimonious and flexible representations that greatly reduce computational time while simultaneously achieving accuracy in model fitting and prediction. Second, following Bitto and Frühwirth-Schnatter (2019), we use a prior structure that induces shrinkage and allows the model parameters to be time-varying or static as needed, providing greater flexibility and significantly reducing overfitting. Rather than using this prior structure in time-domain models as done in Bitto and Frühwirth-Schnatter (2019), we place the prior on the PARCOR model parameters. Third, we provide two algorithms for achieving posterior inference. One algorithm uses Markov chain Monte Carlo (MCMC) methods to obtain full posterior inference. An alternative algorithm develops a computationally efficient and accurate importance sampling variational Bayes (ISVB) approach for approximate posterior inference. The ISVB algorithm is one of the main contributions of this chapter as it allows users to obtain reliable approximate inference in large-dimensional settings in this flexible multivariate model that also incorporates shrinkage priors. Once full or approximate posterior inference is obtained for the multivariate time-varying PARCOR coefficients, posterior estimates for the implied TV-VAR coefficient matrices and innovations covariance matrices can be obtained via the Whittle's algorithm discussed in Section 3.2.2. Posterior estimates (Section 3.2.5 of Chapter 3) for any function of such matrices, such as the multivariate spectra and functions of the spectra, can also be obtained, via the equation (1.6) in Chapter 1 Section 1.1.1. The performance of shrinkage TV-VPARCOR model is illustrated in simulation studies (Section 4.4) and in the analysis of wind component data (Section 4.5).

4.2 The TV-VPARCOR shrinkage model

4.2.1 Time-varying variance-covariance matrices

We follow the PARCOR lattice representation of the TV-VAR model developed in Section 3.2.1 of Chapter 3. In order to model the time-varying forward and backward variances $\boldsymbol{\Omega}_{f,m,t}$ and $\boldsymbol{\Omega}_{b,m,t}$ in equations (3.1) and (3.2), we consider triangular reductions as proposed in Lopes et al. (2016). Here we detail this formulation only for $\boldsymbol{\Omega}_{f,m,t}$. A similar approach is used for $\boldsymbol{\Omega}_{b,m,t}$. More specifically, we take

$$\boldsymbol{\Phi}_{f,m,t}\boldsymbol{\Omega}_{f,m,t}\boldsymbol{\Phi}'_{f,m,t} = \boldsymbol{\Sigma}_{f,m,t}, \tag{4.1}$$

where $\boldsymbol{\Sigma}_{f,m,t}$ is a diagonal matrix with diagonal elements $\sigma^2_{i,f,m,t}$ for $i = 1, \dots, K$, and $\boldsymbol{\Phi}_{f,m,t}$ is a lower triangular matrix with ones on the main diagonal. We rewrite $\boldsymbol{\Phi}_{f,m,t} = \boldsymbol{I}_K + \boldsymbol{\Phi}^*_{f,m,t}$, where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix and $\boldsymbol{\Phi}^*_{f,m,t}$ is a lower triangular matrix with zeros on the main diagonal. We can rewrite the forward lattice structure (3.1) by considering the transformation $\tilde{\boldsymbol{f}}_t^{(m)} = \boldsymbol{\Phi}_{f,m,t}\boldsymbol{f}_t^{(m)}$, such that

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{\beta}^{(m)}_{f,m,t}\boldsymbol{b}^{(m-1)}_{t-m} + \boldsymbol{\Phi}^{-1}_{f,m,t}\tilde{\boldsymbol{f}}_t^{(m)}, \qquad \tilde{\boldsymbol{f}}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{f,m,t}). \tag{4.2}$$

Further arrangements lead to

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{\Gamma}^{(m)}_{f,m,t}\boldsymbol{b}^{(m-1)}_{t-m} + \boldsymbol{\Phi}^*_{f,m,t}\left(-\boldsymbol{f}_t^{(m-1)}\right) + \tilde{\boldsymbol{f}}_t^{(m)}, \quad \tilde{\boldsymbol{f}}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{f,m,t}), \tag{4.3}$$

where $\boldsymbol{\Gamma}^{(m)}_{f,m,t} = \boldsymbol{\Phi}_{f,m,t}\boldsymbol{\beta}^{(m)}_{f,m,t}$. Now, the transformed forward prediction errors $\tilde{f}_{it}^{(m)}$ for $i = 1, \dots K$, are all independent and follow normal distributions $\mathcal{N}(0, \sigma^2_{i,f,m,t})$.

Finally, the model can be recast as a multivariate dynamic regression, i.e.,

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{\Lambda}_{f,m,t}\boldsymbol{x}_{f,t}^{(m-1)} + \tilde{\boldsymbol{f}}_t^{(m)}, \quad \tilde{\boldsymbol{f}}_t^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{f,m,t}), \tag{4.4}$$

where $\boldsymbol{\Lambda}_{f,m,t} = (\boldsymbol{\Gamma}_{f,m,t}^{(m)}, \boldsymbol{\Phi}_{f,m,t}^{*})$ is the $K \times 2K$ matrix with time-varying coefficients $\lambda_{ij,f,m,t}$ in row i and column j. Note that $\lambda_{ij,f,m,t} = 0$ for $j > K+i-1$. In addition, $\boldsymbol{x}_{f,t}^{(m-1)} = ((\boldsymbol{b}_{t-m}^{(m-1)})', (-\boldsymbol{f}_t^{(m-1)})')'$ is a $2K$-dimensional column vector.

4.2.2 Additional model structure and prior specifications

We begin by specificying the structure of the time-varying parameters $\lambda_{ij,f,m,t}$ in (4.4). These parameters are assumed to follow independent random walks:

$$\lambda_{ij,f,m,t} = \lambda_{ij,f,m,t-1} + w_{ij,f,m,t}, \qquad w_{ij,f,m,t} \sim \mathcal{N}(0, \theta_{ij,f,m}), \tag{4.5}$$

with initial distribution $\lambda_{ij,f,m,0} \sim \mathcal{N}(\lambda_{ij,f,m}, \theta_{ij,f,m} p_{ij,f,m,0})$. We implement a conjugate prior on $p_{ij,f,m,0}$ as proposed in Bitto and Frühwirth-Schnatter (2019), that is, we assume $p_{ij,f,m,0} \sim \mathcal{G}^{-1}(\nu^p, (\nu^p - 1)c^p)$ with hyperparameters $c^p = 1$ and $\nu^p = 20$. Priors for $\lambda_{ij,f,m}$ and $\theta_{ij,f,m}$ are detailed below. Furthermore, and again following an approach similar to that in (Bitto and Frühwirth-Schnatter, 2019), we implement a non-centered parametrization of the DLM model (Frühwirth-Schnatter and Wagner, 2010) by applying a standardized transformation on $\lambda_{ij,f,m,t}$ as follows:

$$\lambda_{ij,f,m,t} = \lambda_{ij,f,m} + \sqrt{\theta}_{ij,f,m}\tilde{\lambda}_{ij,f,m,t}, \qquad t = 0, \ldots, T. \tag{4.6}$$

Here, $\sqrt{\theta}_{ij,f,m}$ denotes the standard deviation of the error in the random walk state equation for the jth coefficient in the ith equation (4.5). Thus, using (4.6), the ith $(i = 1, \dots, K)$ component of the forward PARCOR lattice equation (4.4) is given by

$$f_{it}^{(m-1)} = \boldsymbol{\lambda}_{i,f,m}\boldsymbol{x}_{f,t}^{(m-1)} + \tilde{\boldsymbol{\lambda}}_{i,f,m,t}\sqrt{\boldsymbol{\Theta}}_{i,f,m}\boldsymbol{x}_{f,t}^{(m-1)} + \tilde{f}_{it}^{(m)},\ \tilde{f}_{it}^{(m)} \sim \mathcal{N}(0, \sigma_{i,f,m,t}^2), \tag{4.7}$$

where $\boldsymbol{\lambda}_{i,f,m} = (\lambda_{i,1,f,m}, \dots, \lambda_{i,K+i-1,f,m}, 0, \dots, 0)$, is a $2K$-dimensional row vector, $\tilde{\boldsymbol{\lambda}}_{i,f,m,t} = (\tilde{\lambda}_{i,1,f,m,t}, \dots, \tilde{\lambda}_{i,K+i-1,f,m,t}, 0, \dots, 0)$, is also a $2K$-dimensional row vector, and finally, $\sqrt{\boldsymbol{\Theta}}_{i,f,m} = \mathrm{diag}(\sqrt{\theta}_{i,1,f,m}, \dots, \sqrt{\theta}_{i,K+i-1,f,m}, 0, \dots, 0)$ is a $2K \times 2K$ diagonal matrix.

Then the evolution process in (4.5) now becomes a random walk with standard normal independent increments, i.e.,

$$\tilde{\lambda}_{ij,f,m,t} = \tilde{\lambda}_{ij,f,m,t-1} + \tilde{w}_{ij,f,m,t}, \qquad \tilde{w}_{ij,f,m,t} \sim \mathcal{N}(0,1), \tag{4.8}$$

and initial value $\tilde{\lambda}_{ij,f,m,0} \sim \mathcal{N}(0, p_{ij,f,m,0})$.

We impose a normal-gamma shrinkage prior (Griffin and Brown, 2010) on the static components $\lambda_{ij,f,m}$ and $\sqrt{\theta}_{ij,f,m}$. Other shrinkage priors can be considered, but normal-gamma shrinkage priors facilitate the implementation of the importance sampling variational Bayes algorithm for approximate inference while showing comparable or improved performance than other shrinkage priors in our simulation studies.

Due to the flexibility of the lattice structure, we can induce the shrinkage priors starting at stage $m = 1$ or only consider shrinkage after some certain stage m_0. Once again, our extensive simulation studies have shown that imposing shrinkage priors for stages $m = 2, \dots, P$ leads to more stable results than using shrinkage priors for all stages $m = 1, \dots, P$, when there is evidence that the model order is above 1. For stage $m = 1$ we propose using the PARCOR prior specification defined in Section 3.2.3, and for $m > 1$ we use a hierarchical normal-gamma prior for $\sqrt{\theta}_{ij,f,m}$ for $i = 1 : K$ and $j = 1 : K + i - 1$, i.e.,

$$\sqrt{\theta}_{ij,f,m}|\xi^2_{ij,f,m} \sim \mathcal{N}\left(0, \xi^2_{ij,f,m}\right), \quad \xi^2_{ij,f,m}|a^{\xi}_{i,f}, \kappa^2_{i,f} \sim \mathcal{G}\left(a^{\xi}_{i,f}, a^{\xi}_{i,f}\kappa^2_{i,f}/2\right) \tag{4.9}$$

$$\kappa^2_{i,f} \sim \mathcal{G}(d_1, d_2), \quad a^{\xi}_{i,f} \sim \mathcal{E}(b^{\xi}), \tag{4.10}$$

The prior variance $\xi^2_{ij,f,m}$ controls the local shrinkage. The global shrinkage is controlled by $\kappa^2_{i,f}$ which is specific for each time series, but shared across j and m. We assume $\kappa^2_{i,f}$ follows a gamma distribution with fixed hyperparameters d_1 and d_2. The parameter $a^{\xi}_{i,f}$ controls the tail behavior of the prior. A smaller value of $a^{\xi}_{i,f}$ will lead to a distribution $p(\sqrt{\theta}_{ij,f,m}|\xi^2_{ij,f,m})$ with heavier tails and will place more mass at zero. We assume that $a^{\xi}_{i,f}$ follows an exponential distribution with a fixed hyperparameter $b^{\xi} \geq 1$ as in Griffin and Brown (2010) and Bitto and Frühwirth-Schnatter (2019). Setting $a^{\xi}_{i,f} = 1$ leads to the Bayesian Lasso prior (Belmonte et al., 2014).

Similarly, we consider the following prior on $\lambda_{ij,f,m}$ for $i = 1 : K$ and $j = 1 : K + i - 1$:

$$\lambda_{ij,f,m}|\tau^2_{ij,f,m} \sim \mathcal{N}(0, \tau^2_{ij,f,m}), \quad \tau^2_{ij,f,m}|a^{\tau}_{i,f}, \nu^2_{i,f} \sim \mathcal{G}(a^{\tau}_{i,f}, a^{\tau}_{i,f}\nu^2_{i,f}/2), \tag{4.11}$$

$$\nu^2_{i,f} \sim \mathcal{G}(e_1, e_2), \quad a^{\tau}_{i,f} \sim \mathcal{E}(b^{\tau}). \tag{4.12}$$

For the prior specification of $\sigma^2_{i,f,m,t}$, for $i = 1, \ldots, K$ and $m = 1, \ldots, P$, we consider two cases: the homoscedastic case with $\sigma^2_{i,f,m,t} = \sigma^2_{i,f,m}$ for all t, and the heteroscedastic case with $\sigma^2_{i,f,m,t}$ time-dependent. In the homoscedastic case, we follow Bitto and Frühwirth-Schnatter (2019) and use a hierarchical conjugate prior distribution as follows,

$$\sigma^2_{i,f,m}|n_{i,f,m,0}, s_{i,f,m,0} \sim \mathcal{G}^{-1}\left(\frac{n_{i,f,m,0}}{2}, \frac{n_{i,f,m,0}s_{i,f,m,0}}{2}\right),$$
$$s_{i,f,m,0} \sim \mathcal{G}(g_{i,f,m,0}, n_{i,f,m,0}G_{i,f,m,0}),$$

with hyperparameters $n_{i,f,m,0}, g_{i,f,m,0}$ such that $G_{i,f,m,0} = g_{i,f,m,0}/[E(\sigma^2_{i,f,m})(n^2_{i,f,m,0} - n_{i,f,m,0})]$, with $n_{i,f,m,0} = 2.5$ and $g_{i,f,m,0} = 5$, where $E(\sigma^2_{i,f,m})$ is a prior estimate of $\sigma^2_{i,f,m}$.

For heteroscedasticity, we implement a stochastic volatility (SV) specification as in Jacquier et al. (2002), where $\sigma^2_{i,f,m,t} = \exp\{h_{i,f,m,t}\}$ and the log volatility $h_{i,f,m,t}$ follows an AR(1) process:

$$h_{i,f,m,t}|h_{i,f,m,t-1}, \mu^h_{i,f,m}, \phi^h_{i,f,m}, (\sigma^h_{h,i,f,m})^2 \sim \mathcal{N}\left(M, (\sigma^h_{i,f,m})^2\right),$$

where $M = \mu^h_{i,f,m} + \phi^h_{i,f,m}(h_{i,f,m,t-1} - \mu^h_{i,f,m})$. The initial state $h_{i,f,m,0}$ is assumed to follow the stationary distribution of the AR(1) process, i.e.

$$h_{i,f,m,0}|\mu^h_{i,f,m}, \phi^h_{i,f,m}, (\sigma^h_{i,f,m})^2 \sim \mathcal{N}\left(\mu^h_{i,f,m}, (\sigma^h_{i,f,m})^2/(1 - (\phi^h_{i,f,m})^2\right).$$

As in Kastner and Frühwirth-Schnatter (2014) we consider independent priors, i.e.,

$$p\left(\mu^h_{i,f,m}, \phi^h_{i,f,m}, (\sigma^h_{i,f,m})^2\right) = p(\mu^h_{i,f,m})p(\phi^h_{i,f,m})p\left((\sigma^h_{i,f,m})^2\right),$$

where $\mu^h_{i,f,m} \sim \mathcal{N}(b^\mu, B^\mu)$, $(\phi^h_{i,f,m} + 1)/2 \sim \mathcal{B}(a^\phi, b^\phi)$, and $(\sigma^h_{i,f,m})^2 \sim \mathcal{G}(1/2, 1/\left(2B^\sigma\right))$, with hyperparameters $b^\mu = 0, B^\mu = 100, a^\phi = 20, b^\phi = 1.5$, and $B^\sigma = 1$.

The equations above complete the model structure for the forward lattice equations. The backward lattice structure has a similar form with different regressors $\boldsymbol{x}_{b,t}^{(m-1)} = ((\boldsymbol{f}_{t+m}^{(m-1)})', -(\boldsymbol{b}_t^{(m-1)})')'$ and their corresponding time-varying coefficients, that is,

$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{\Lambda}_{b,m,t}\boldsymbol{x}_{b,t}^{(m-1)} + \tilde{\boldsymbol{b}}_t^{(m)}, \quad \tilde{\boldsymbol{b}}_t^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{b,m,t}), \tag{4.13}$$

which, if triangular reduction and the non-centered parameterization are applied, results on the following equation for the ith component:

$$b_{it}^{(m-1)} = \boldsymbol{\lambda}_{i,b,m}\boldsymbol{x}_{b,t}^{(m-1)} + \tilde{\boldsymbol{\lambda}}_{i,b,m,t}\sqrt{\boldsymbol{\Theta}}_{i,b,m}\boldsymbol{x}_{b,t}^{(m-1)} + \tilde{b}_{it}^{(m)},\ \tilde{b}_{it}^{(m)} \sim \mathcal{N}(0, \sigma^2_{i,b,m,t}), \tag{4.14}$$

where $\boldsymbol{\lambda}_{i,b,m} = (\lambda_{i,1,b,m}, \ldots, \lambda_{i,K+i-1,b,m}, 0, \ldots, 0)$, is a 2K-dimensional row vector, $\tilde{\boldsymbol{\lambda}}_{i,b,m,t} = (\tilde{\lambda}_{i,1,b,m,t}, \ldots, \tilde{\lambda}_{i,K+i-1,b,m,t}, 0, \ldots, 0)$, is also a 2K-dimensional row vector, and, $\sqrt{\boldsymbol{\Theta}}_{i,b,m} = \text{diag}(\sqrt{\theta}_{i,1,b,m}, \ldots, \sqrt{\theta}_{i,K+i-1,b,m}, 0, \ldots, 0)$ is a $2K \times 2K$ diagonal matrix. The shrinkage prior structure on the static model parameters and the prior structure on $\sigma^2_{i,b,m,t}$ have the same form as those specified above for the forward model. A random walk evolution is assumed on the $\tilde{\lambda}_{i,\cdot,b,m,t}$. We omit the details due to space limitations.

4.3 Full and approximate posterior inference

In this section we discuss two algorithms for posterior inference. The first algorithm uses a simulation-based Markov Chain Monte Carlo approach to obtain full posterior inference. The second algorithm uses an importance sampling variational Bayes (ISVB) approach to obtain fast approximate posterior estimation. This second algorithm is one of the key contributions of this paper as it allows user to obtain fast and reliable approximate posterior estimation of the proposed models for relatively large-dimensional multivariate settings.

4.3.1 General MCMC algorithm for the forward PARCOR shrinkage model

Our proposed MCMC algorithm is detailed below. Before starting the algorithm, we fix the potential model order P and set all the prior hyperparameters, i.e., we set $\nu^p, c^p, d_1, d_2, b^\xi$, e_1, e_2, b^τ, and the variance hyperparameters $n_{i,\cdot,m,0}$, $g_{i,\cdot,m,0}$ for the homoscedastic case or $b^\mu, B^\mu, a^\phi, b^\phi$ and B^σ in the time-varying case. Therefore, at each stage m, $m = 1 : P$, the list of forward parameters that are sampled is given by $\{\tilde{\boldsymbol{\lambda}}_{f,m,0:T}, \boldsymbol{\lambda}_{f,m}, \boldsymbol{\theta}_{f,m}, \boldsymbol{\sigma}^2_{f,m}, \boldsymbol{\xi}^2_{f,m}, \boldsymbol{\tau}^2_{f,m}, \boldsymbol{a}^\xi_f, \boldsymbol{a}^\tau_f, \boldsymbol{p}_{f,m,0}, \boldsymbol{s}_{f,m,0}\}$, for the homoscedastic case. The list will also include $\mathbf{h}_{f,m,0:T}$, $\boldsymbol{\mu}^h_{f,m}$, $\boldsymbol{\phi}^h_{f,m}$ and $\boldsymbol{\sigma}^2_{f,m}$ for the heteroscedastic case. There is a similar set of parameters for backward PARCOR model.

MCMC algorithm

1. Set the initial values for the forward and backward prediction errors as $\boldsymbol{f}_t^{(0)} = \boldsymbol{b}_t^{(0)} = \boldsymbol{y}_t$, for $t = 1 : T$. For all $m = 1 : P$, set initial values for $p_{ij,f,m,0}$ as well as initial values for the set of forward and backward fixed parameters $\boldsymbol{\Psi}_{i,f,m} = (\boldsymbol{\lambda}_{i,f,m}, \sqrt{\boldsymbol{\theta}}_{i,f,m})$ and $\boldsymbol{\Psi}_{i,b,m} = (\boldsymbol{\lambda}_{i,b,m}, \sqrt{\boldsymbol{\theta}}_{i,b,m})$, with $\sqrt{\boldsymbol{\theta}}_{i,\cdot,m} = (\sqrt{\theta}_{i,1,\cdot,m}, \dots, \sqrt{\theta}_{i,K+i-1,\cdot,m})'$ and also set initial values for $\sigma^2_{i,\cdot,m}$ (or $\sigma^2_{i,\cdot,m,t}$ in the heterocedastic case), for $i = 1, \dots, K$.

2. Use $\{\boldsymbol{f}_t^{(0)}\}$ and $\{\boldsymbol{b}_t^{(0)}\}$ as vectors of responses in the observational level equations (4.7) and (4.14), respectively. Combine these equations with the corresponding evolution equations and use the forward filtering Backward Sampling (FFBS) algorithm (Carter and R.Kohn, 1994, Frühwirth-Schnatter, 1994) to sample the forward and backward latent states $\tilde{\boldsymbol{\lambda}}_{i,f,m,0:T}$ and $\tilde{\boldsymbol{\lambda}}_{i,b,m,0:T}$, for $m = 1$ conditional on all rest of parameters, for $i = 1, \dots, K$.

3. For $m = 1$, sample the scale parameters of the initial distribution $p_{ij,f,m,0}$ for each $i = 1, \ldots, K$, and $j = 1 : K + i - 1$, from

$$p_{ij,f,m,0} | \tilde{\boldsymbol{\lambda}}_{j,f,m,0} \sim \mathcal{G}^{-1} \left(\nu^p + \frac{1}{2}, (\nu^p - 1)c^p + \frac{1}{2} \tilde{\boldsymbol{\lambda}}^2_{j,f,m,0} \right).$$

4. For $i = 1 : K$, jointly sample the forward static components $\boldsymbol{\Psi}_{i,f,m}$ for $m = 1$ from a multivariate Gaussian distribution denoted as $p(\boldsymbol{\Psi}_{i,f,1} | \tilde{\boldsymbol{\lambda}}_{i,f,1,t}, \boldsymbol{\tau}^2_{i,f,1}, \boldsymbol{\xi}^2_{i,f,1}, f^{(0)}_{it})$. Here $\boldsymbol{\tau}^2_{i,f,m}$ collects the $\tau^2_{ij,f,m}$ and $\boldsymbol{\xi}^2_{i,f,m}$ collects the $\xi^2_{ij,f,m}$ for all j for a given i and m. Similarly, jointly sample the backward static component $\boldsymbol{\Psi}_{i,b,1}$ from a multivariate Gaussian distribution. The computation details can be found in Section B.1.2 of the Supplementary Material. We also implement the ancillarity-sufficiency interweaving strategy (ASIS) introduced by Yu and Meng (2011) that for this particular case redraws the forward and backward static components $\boldsymbol{\Psi}_{i,f,m}$ and $\boldsymbol{\Psi}_{i,b,m}$ in order to improve the mixing performance. Again, details can be found in Section B.1.3 of the Supplementary Material.

5. Sample $\xi^2_{ij,f,1}$ and $\tau^2_{ij,f,1}$, for $i = 1 : K$ and $j = 1 : K + i - 1$, from conditionally independent generalized inverse Gaussian distributions given in Section B.1.4 of the Supplementary Material. Sample $\xi^2_{i,b,1}$ and $\tau^2_{ij,f,1}$ in a similar way.

6a. For the homoscedastic case, sample $\sigma^2_{i,f,m}$ for $m = 1$ given the rest of parameters from conditionally independent inverted gamma distributions, i.e.,

$$\sigma^2_{i,f,m} | \ldots \sim \mathcal{G}^{-1} \left(\frac{n_{i,f,m,0} + T}{2}, \frac{1}{2} \left(n_{i,f,m,0} s_{i,f,m,0} + \sum_{t=1}^{T} \left(f^{(m-1)}_{it} - \boldsymbol{Z}_{i,f,m,t} \boldsymbol{\Psi}_{i,f,m} \right)^2 \right) \right),$$

where $\boldsymbol{Z}_{i,f,m,t}$ is defined in Section B.1.2 of the Supplementary Material.

Then sample $s_{i,f,m,0}$ from a Gamma distribution as follows

$$s_{i,f,m,0}|\sigma^2_{i,f,m,t} \sim \mathcal{G}\left(g_{i,f,m,0} + \frac{n_{i,f,m,0}}{2}, n_{i,f,m,0}\left(G_{i,f,m,0} + \frac{1}{2\sigma^2_{i,f,m}}\right)\right).$$

6b. For the heteroscedastic case, we implement an interweaving strategy for boosting MCMC estimation of SV models (Kastner and Frühwirth-Schnatter, 2014) to sample $\mathbf{h}_{i,f,m,0:T}$, as well as $\mu^h_{i,f,m}, \phi^h_{i,f,m}$, and $(\sigma^h_{i,f,m})^2$ for $m = 1$.

7. Use the observational equations (4.7) and (4.14) to obtain the new series of forward and backward prediction errors $\{\tilde{\boldsymbol{f}}^{(1)}_t\}$ and $\{\tilde{\boldsymbol{b}}^{(1)}_t\}$, for $t = 1 : T$, at stage $m = 1$. Using $\boldsymbol{f}^{(m)}_t = \boldsymbol{\Phi}^{-1}_{f,m,t}\tilde{\boldsymbol{f}}^{(m)}_t$ and $\boldsymbol{b}^{(m)}_t = \boldsymbol{\Phi}^{-1}_{b,m,t}\tilde{\boldsymbol{b}}^{(m)}_t$, we obtain $\boldsymbol{f}^{(1)}_t$ and $\boldsymbol{b}^{(1)}_t$.

8. Repeat steps $2 - 7$ above for $m = 2 : P$ until all the forward and backward PARCOR coefficients $\boldsymbol{\Gamma}^{(m)}_{f,m,t}$, $\boldsymbol{\Gamma}^{(m)}_{b,m,t}$, $\boldsymbol{\Phi}^*_{f,m,t}$ and $\boldsymbol{\Phi}^*_{b,m,t}$ have been sampled for all $m = 1, \ldots, P$. Then, samples of $\boldsymbol{\Omega}_{f,m,t}$ can be obtained using (4.1) with the sampled $\boldsymbol{\Phi}^*_{f,m,t}$ and the sampled $\boldsymbol{\Sigma}_{f,m,t}$ obtained from steps above.

9. Sample $(\kappa^2_{i,f}|a^\xi_{i,f}, \boldsymbol{\xi}^2_{i,f})$ and $(\kappa^2_{i,b}|a^\xi_{i,b}, \boldsymbol{\xi}^2_{i,b})$, as well as $(\nu^2_{i,f}|a^\tau_{i,f}, \boldsymbol{\tau}^2_{i,f})$ and $(\nu^2_{i,b}|a^\tau_{i,b}, \boldsymbol{\tau}^2_{i,b})$ for $i = 1 : K$ from the gamma distributions given in Section B.1.5 of the Supplementary Material.

10. Sample $a^\xi_{i,f}$ and $a^\tau_{i,f}$ as well as $a^\xi_{i,b}$ and $a^\tau_{i,b}$, for $i = 1 : K$, by using a random walk Metropolis-Hasting (MH) step with proposal distribution $\log a^{\xi,\text{new}}_{i,\cdot} \sim \mathcal{N}(\log a^\xi_{i,\cdot}, c^2_\xi)$ and $\log a^{\tau,\text{new}}_{i,\cdot} \sim \mathcal{N}(\log a^\tau_{i,\cdot}, c^2_\tau)$. Details can be found in Section B.1.6 and B.1.7 of the Supplementary Material.

11. Transform the sampled $\boldsymbol{\Gamma}^{(m)}_{f,m,t}$, $\boldsymbol{\Gamma}^{(m)}_{b,m,t}$ to obtain samples of the forward and backward PARCOR coefficients $\boldsymbol{\beta}^{(m)}_{f,m,t}$ and $\boldsymbol{\beta}^{(m)}_{b,m,t}$, for $m = 1 : P$.

12. Finally, use the sampled $\boldsymbol{\beta}_{f,m,t}^{(m)}$ and $\boldsymbol{\beta}_{b,m,t}^{(m)}$, for $m = 1 : P$, and Whittle's algorithm to obtain the sampled forward and backward TV-VAR coefficients $\boldsymbol{A}_{m,t}^{(P)}$ and $\boldsymbol{D}_{m,t}^{(P)}$ for $m = 1 : P$ and all t.

4.3.2 Importance Sampling Variational Bayes algorithm for approximate inference

In this section we develop an Importance Sampling Variational Bayes (ISBV) algorithm for fast approximate posterior inference in the multivariate TV-VPARCOR shrinkage model. Variational Bayes (VB) is an optimization method for fast, approximate posterior inference that can also be implemented for settings that deal with state-space models (Ostwald et al., 2014).

In order to improve the computational efficiency, we consider only the homoscedastic case and assume that $\sigma^2_{i,f,m,t} = \sigma^2_{i,f,m}$ for all t. Users only need to specify positive scalars $n_{i,f,m,0}$ and $n_{i,b,m,0}$ that correspond to prior degrees of freedom as well as $s_{i,f,m,0}$ and $s_{i,b,m,0}$ which are prior estimates of $\sigma^2_{i,f,m}$ and $\sigma^2_{i,b,m}$, respectively. In addition, we also fix $p_{ij,f,m,0} = 1$ for all i, j and m.

Let $\boldsymbol{\Pi}_{f,m} = \left\{ \tilde{\boldsymbol{\lambda}}_{f,m,1:T}, \boldsymbol{\sigma}^{\mathbf{2}}_{f,m}, \boldsymbol{\lambda}_{f,m}, \boldsymbol{\Theta}_{f,m}, \boldsymbol{\xi}^2_{f,m}, \boldsymbol{\tau}^2_{f,m}, \boldsymbol{a}^{\xi}_{f}, \boldsymbol{\kappa}^2_{f}, \boldsymbol{a}^{\tau}_{f}, \boldsymbol{\nu}^2_{f} \right\}$ for $m = 1 : P$, denote the set of all the parameters in the forward multivariate PARCOR model. Variational Bayes (VB) is an optimization method for fast, approximate posterior inference that can also be implemented for settings that deal with state-space models (Ostwald et al., 2014). Within the VB framework, we approximate the posterior distribution $p(\boldsymbol{\Pi}_{f,m} | \boldsymbol{f}^{(m)}_{1:T})$ with an arbitrary variational distribution $r(\boldsymbol{\Pi}_{f,m})$ which minimizes the Kullback-Leibler (KL) divergence between these distributions (Kullback and Leibler, 1951). Such optimal variational distribution maximizes their corresponding evidence lower bound (ELBO), therefore, in order to find this distribution we optimize the ELBO.

A common choice for the family of variational distributions over which we optimize the ELBO assumes a factorization over different sets of variables known as a mean-field approximation (Beal, 2003). Here we use the following factorization

$$r(\boldsymbol{\Pi}_{f,m}) = r(\tilde{\boldsymbol{\lambda}}_{f,m,1:T}, \boldsymbol{\sigma}^{\mathbf{2}}_{f,m}) r(\boldsymbol{\lambda}_{f,m}) r(\boldsymbol{\Theta}_{f,m}) r(\boldsymbol{\tau}^2_{f,m}) r(\boldsymbol{\xi}^2_{f,m}) r(\boldsymbol{\kappa}^2_f) r(\boldsymbol{\nu}^2_f) r(\boldsymbol{a}^{\xi}_f) r(\boldsymbol{a}^{\tau}_f). \tag{4.15}$$

This factorization makes an assumption of stochastic independence between some sets of variables. It has been shown that for each component of the factorization, the ELBO is maximized by $r(\boldsymbol{\Pi}_{f,m,c})$ such that

$$r(\boldsymbol{\Pi}_{f,m,c}) \propto \exp\left\{ \int \log p(\boldsymbol{f}^{(m-1)}_{1:T}, \boldsymbol{\Pi}_{f,m}) r(\boldsymbol{\Pi}_{f,m,-c}) \mathrm{d}\boldsymbol{\Pi}_{f,m,-c} \right\}, \tag{4.16}$$

where $\boldsymbol{\Pi}_{f,m,c}$ denotes the set of variables in the component being maximized and $\boldsymbol{\Pi}_{f,m,-c}$ denotes the variables not in that component of the partition. We will also use the notation $< g(\boldsymbol{\Pi}_{f,m,c}) >$ defined as:

$$< g(\boldsymbol{\Pi}_{f,m,c}) >= \int \log g(\boldsymbol{\Pi}_{f,m}) r(\boldsymbol{\Pi}_{f,m,-c}) d\boldsymbol{\Pi}_{f,m,-c}. \tag{4.17}$$

The general variational algorithm is described below.

1. Fix the potential model order P. Set all the hyperparameters, i.e., $e_{\cdot,1}, e_{\cdot,2}, b^{\tau}$, $d_{\cdot,1}, d_{\cdot,2}, b^{\xi}$, $p_{ij,f,m,0}$, $n_{i,\cdot,m,0}$ and $s_{i,\cdot,m,0}$ for $i = 1 : K$, and $m = 1 : P$. Set the initial values for the forward and backward prediction errors as $\boldsymbol{f}^{(0)}_t = \boldsymbol{b}^{(0)}_t = \boldsymbol{y}_t$, for $t = 1, \ldots, T$. Set $l = 0$ and initialize $< \tilde{\boldsymbol{\lambda}}_{i,\cdot,m,t} >^{\{l\}}$, $< \boldsymbol{\lambda}_{i,\cdot,m} >^{\{l\}}$, $< \sqrt{\boldsymbol{\Theta}}_{i,\cdot,m} >^{\{l\}}$ for all i, t and m.

2. For each $i = 1 : K$ use the components of $\{\boldsymbol{f}^{(0)}_t\}$ and $\{\boldsymbol{b}^{(0)}_t\}$ as the responses

in the observational level equations (4.7) and (4.14), respectively, and also substitute $\boldsymbol{\lambda}_{i,f,m}$ by $< \boldsymbol{\lambda}_{i,f,m} >^{\{l\}}$ and $\boldsymbol{\lambda}_{i,b,m}$ by $< \boldsymbol{\lambda}_{i,b,m} >^{\{l\}}$ in these equations. Then, use the forward filtering backward smoothing algorithm (West and Harrison, 1997) to update the forward and backward variational approximations $< \tilde{\boldsymbol{\lambda}}_{i,f,m,0:T} >^{\{l+1\}}$, $< \tilde{\boldsymbol{\lambda}}_{i,b,m,0:T} >^{\{l+1\}}$, as well as $< \sigma^2_{i,f,1} >^{\{l+1\}}$ and $< \sigma^2_{i,b,1} >^{\{l+1\}}$ which can be obtained by the sequential filtering equations in Supplementary Material Section B.2.1.

3. For $i = 1, \ldots, K$, and $j = 1, \ldots, K + i - 1$, update $< \lambda_{ij,f,1} >^{\{l+1\}}$ with the approximate distributions for the forward static components as $r^{\{l+1\}}(\lambda_{ij,f,1}) = \mathcal{N}\left(\mu^{\{l+1\}}_{\lambda_{ij,f,1}}, \Sigma^{\{l+1\}}_{\lambda_{ij,f,1}}\right)$. Details appear in Supplementary Material Section B.2.2. A similar update is used to obtain $< \lambda_{ij,b,1} >^{\{l+1\}}$ and the approximate distribution of the backward static component $r^{\{l+1\}}(\lambda_{ij,b,1})$ with the corresponding Gaussian distribution.

4. For $i = 1, \ldots, K$, and $j = 1, \ldots, K + i - 1$, update $< \sqrt{\theta}_{ij,f,1} >^{\{l+1\}}$ with the approximate distribution of the forward standard deviation of the latent process via $r^{\{l+1\}}(\sqrt{\theta}_{ij,f,1}) = \mathcal{N}\left(\mu^{\{l+1\}}_{\sqrt{\theta}_{ij,f,1}}, \Sigma^{\{l+1\}}_{\sqrt{\theta}_{ij,f,1}}\right)$. The details are in Section B.2.3 of the Supplementary Material. Similarly, update $< \sqrt{\theta}_{ij,b,1} >^{\{l+1\}}>$ and the approximate distribution $r^{\{l+1\}}(\sqrt{\theta}_{ij,b,1})$ using the corresponding Gaussian distribution.

5. For $i = 1, \ldots, K$, and $j = 1, \ldots, K + i - 1$, update $< \tau^2_{ij,f,1} >^{\{l+1\}}$ with variational distribution

$$r^{\{l+1\}}(\tau^2_{ij,f,1}) = \mathcal{GIG}\left(\frac{1}{2}, \frac{1}{2} < \nu^2_{i,f} >^{\{l\}}, < \lambda^2_{ij,f,1} >^{\{l+1\}}\right).$$

6. For $i = 1, \ldots, K$, and $j = 1, \ldots, K + i - 1$, update $< \xi^2_{ij,f,1} >^{\{l+1\}}$ with

variational distribution

$$r^{\{l+1\}}(\xi^2_{ij,f,1}) = \mathcal{GIG}\left(\frac{1}{2}, \frac{1}{2} < \kappa^2_{i,f} >^{\{l\}}, < \theta_{ij,f,1} >^{\{l+1\}}\right)$$

7. Use observational equations (4.7) and (4.14) to obtain the new series of forward and backward prediction errors, $\{< \tilde{\boldsymbol{f}}_t^{(m)} >^{\{l+1\}}\}$ and $\{< \tilde{\boldsymbol{b}}_t^{(m)} >^{\{l+1\}}\}$, for $t = 1, \dots, T$, at stage $m = 1$. Using

$$\begin{aligned} < \boldsymbol{f}_t^{(m)} >^{\{l+1\}} &=< \boldsymbol{\Phi}^{-1}_{f,m,t} >^{\{l+1\}}< \tilde{\boldsymbol{f}}_t^{(m)} >^{\{l+1\}}, \\ < \boldsymbol{b}_t^{(m)} >^{\{l+1\}} &=< \boldsymbol{\Phi}^{-1}_{b,m,t} >^{\{l+1\}}< \tilde{\boldsymbol{b}}_t^{(m)} >^{\{l+1\}}, \end{aligned}$$

   we obtain $\{< \boldsymbol{f}_t^{(1)} >^{\{l+1\}}\}$ and $\{< \boldsymbol{b}_t^{(1)} >^{\{l+1\}}\}$.

8. Repeat steps $2 - 7$ above until all the forward and backward estimates $< \tilde{\boldsymbol{\lambda}}_{\cdot,m,1:T} >^{\{l+1\}}$, $< \boldsymbol{\sigma}^2_{\cdot,m} >^{\{l+1\}}$, $< \boldsymbol{\lambda}_{\cdot,m} >^{\{l+1\}}$, $< \boldsymbol{\Theta}_{\cdot,m} >^{\{l+1\}}$, $< \boldsymbol{\tau}^2_{\cdot,m} >^{\{l+1\}}$, $< \boldsymbol{\xi}^2_{\cdot,m} >^{\{l+1\}}$ have been updated for $m = 1, \dots, P$.

9. Transform all the estimated forward and backward regression coefficients back to obtain estimated forward and backward PARCOR coefficients $< \boldsymbol{\beta}^{(m)}_{f,m,0:T} >^{\{l+1\}}$ and $< \boldsymbol{\beta}^{(m)}_{b,m,0:T} >^{\{l+1\}}$, for $m = 1, \dots, P$. The estimated $< \boldsymbol{\Omega}_{f,m} >^{\{l+1\}}$ can be obtained using decomposition (4.1).

10. For $i = 1, \dots, K$, update $< \nu^2_{i,f} >^{\{l+1\}}$ with variational distribution

$$r^{\{l+1\}}(\nu^2_{i,f}) = \mathcal{G}\left(e_1 + A, e_2 + B\right),$$

where

$$\begin{aligned} A &= (K+i-1)P < a^{\tau}_{i,f} >^{\{l\}}, \\ B &= \frac{< \bar{\tau}^2_{i,f} >^{\{l+1\}}}{2} < a^{\tau}_{i,f} >^{\{l\}} (K+i-1)P, \\ < \bar{\tau}^2_{i,f} >^{\{l+1\}} &= \frac{1}{(K+i-1)P} \left(\sum_{j=1}^{K+i-1} \sum_{m=1}^{P} < \tau^2_{ij,f,m} >^{\{l+1\}} \right). \end{aligned}$$

11. For $i = 1, \ldots, K$, update $< \kappa^2_{i,f} >^{\{l+1\}}$ with variational distribution

$$r^{\{l+1\}}(\kappa^2_{i,f}) = \mathcal{G}\left(d_1 + C, d_2 + D\right),$$

where

$$\begin{aligned} C &= (K+i-1)P < a^{\xi}_{i,f} >^{\{l\}}, \\ D &= \frac{< \bar{\xi}^2_{i,f} >^{\{l+1\}}}{2} < a^{\xi}_{i,f} >^{\{l\}} (K+i-1)P, \\ < \bar{\xi}^2_{i,f} >^{\{l+1\}} &= \frac{1}{(K+i-1)P} \left(\sum_{j=1}^{K+i-1} \sum_{m=1}^{P} < \xi^2_{ij,f,m} >^{\{l+1\}} \right). \end{aligned}$$

12. For $i = 1, \ldots, K$, update $r^{\{l+1\}}(a^{\xi}_{i,f})$ and $r^{\{l+1\}}(a^{\tau}_{i,f})$ using importance sampling (IS) based on log-normal proposal distributions for each case. Further details of this IS step can be found in the Supplemental Material Section B.2.5 and B.2.4.

13. Calculate the change of L_2-norm of $< \mathbf{\Pi}_{\cdot,1:P} >^{\{l+1\}}$. If the change is smaller than some ϵ, $\epsilon = 10^{-5}$ for example, then all parameters $< \mathbf{\Pi}_{\cdot,1:P} >^{\{l+1\}}$ have converged. If not, set $l = l + 1$, and repeat steps $2 - 12$.

14. Once all parameters $r(\mathbf{\Pi}_{\cdot,1:P})$ have converged, use estimated $\boldsymbol{\beta}^{(m)}_{f,m,t}$ and $\boldsymbol{\beta}^{(m)}_{b,m,t}$, for $m = 1, \ldots, P$, and the Whittle algorithm (Section 3.2.2 of Chapter

3) to obtain the forward and backward TV-VAR coefficient $\boldsymbol{A}_{m,t}^{(P)}$ and $\boldsymbol{D}_{m,t}^{(P)}$ for $m = 1, \dots, P$.

4.4 Simulation Studies

We illustrate our proposed approach in the analysis of simulated data. We use the ASE (see equation (3.16) in Section 3.3.1) between the estimated spectral density matrix and the true spectral density matrix as a performance criterion. We consider the TV-VPARCOR proposed in Chapter 3 as a benchmark model for comparison.

4.4.1 Bivariate TV-VAR(2) processes

We simulated 100 bivariate time series of length $T = 1024$ from the following TV-VAR(2) model defined in Section 3.3.1 of Chapter 3. In scenario (i) the true square coherence $\rho_{1,2}^2(t, \omega)$ is 0 for all t and ω, since the off-diagonal elements $a_{1,1,2,t} = a_{2,1,2,t} = 0$ for all t. Figure 4.1 shows the true spectral densities $g_{11}(t, \omega)$ and $g_{22}(t, \omega)$ in this scenario. The true log spectral densities and square coherences for scenario (ii) are shown in Figure 4.2. Scenario (iii) is not displayed due to space limitations. We fit the proposed TV-VPARCOR models with Bayesian Lasso and Normal-Gamma priors to each of the 100 simulated bivariate time series for $t = 1 : 1024$ under cases (i), (ii) and (iii). Full and approximate posterior inference of the models under these two shrinkage priors are obtained via MCMC and the ISVB algorithm, respectively. We set a maximum of model order $P_{\max} = 5$, and set the hyperparameters of the priors as $d_{f,i} = d_{b,i} = e_{f,i} = e_{b,i} = 0.001$, for $i = 1, 2$. For the Normal-Gamma shrinkage prior, we set the additional hyperparameters $b^{\xi} = b^{\tau} = 10$. We also set the variances hyperparameters to $b^{\mu} = 0, B^{\mu} = 100, a^{\phi} =$

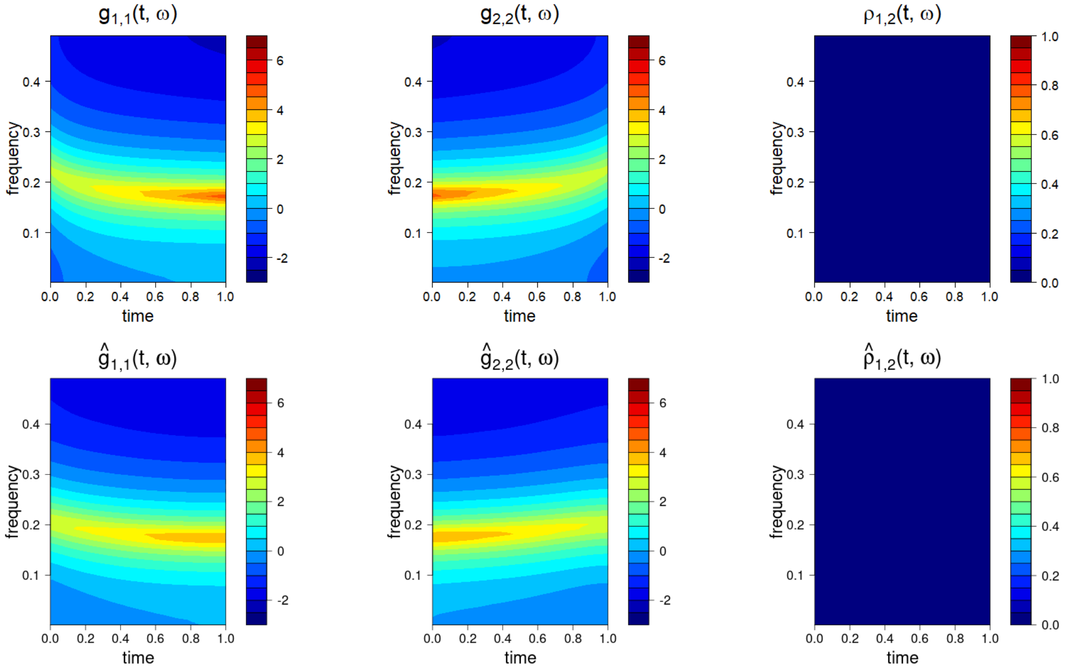

Figure 4.1: Simulation scenario (i) with $a_{1,1,2,t} = a_{2,1,2,t} = 0$. Top row: True log spectral density $g_{11}(t, \omega)$ (left); true log spectral density $g_{22}(t, \omega)$ (center); true squared coherence (right). Bottom row: Corresponding estimated log spectral densities (left and center) and estimated squared coherence.

$20, b^{\phi} = 1.5$, and $B^{\sigma} = 1$. In MCMC algorithm, and we set $n_{i,m,0} = 1$ and $s_{i,\cdot,m,0} = 2$ for $m = 1 : P$. In order to obtain more numerical stable results, the shrinkage prior is only considered for $m > 1$. For stage $m = 1$ we use the conjugate prior structure defined in Section 3.2.3. For each of the 100 simulated time series in all three scenarios, we draw $2,000$ MCMC samples after a burn-in of length $13,000$. We implemented the variational Bayes algorithm with a maximum number of $15,000$ iterations. The algorithm stops when the change of L_2-norm of parameters is smaller than 10^{-5}. For comparison, we also fit the TV-VPARCOR with non-shrinkage conjugate priors developed in Chapter 3 to the simulated bivariate data with fixed model order 5.

Figures 4.1 and 4.2 summarize the posterior inference obtained from the shrinkage TV-VPARCOR approach with Normal-Gamma priors for scenarios (i)

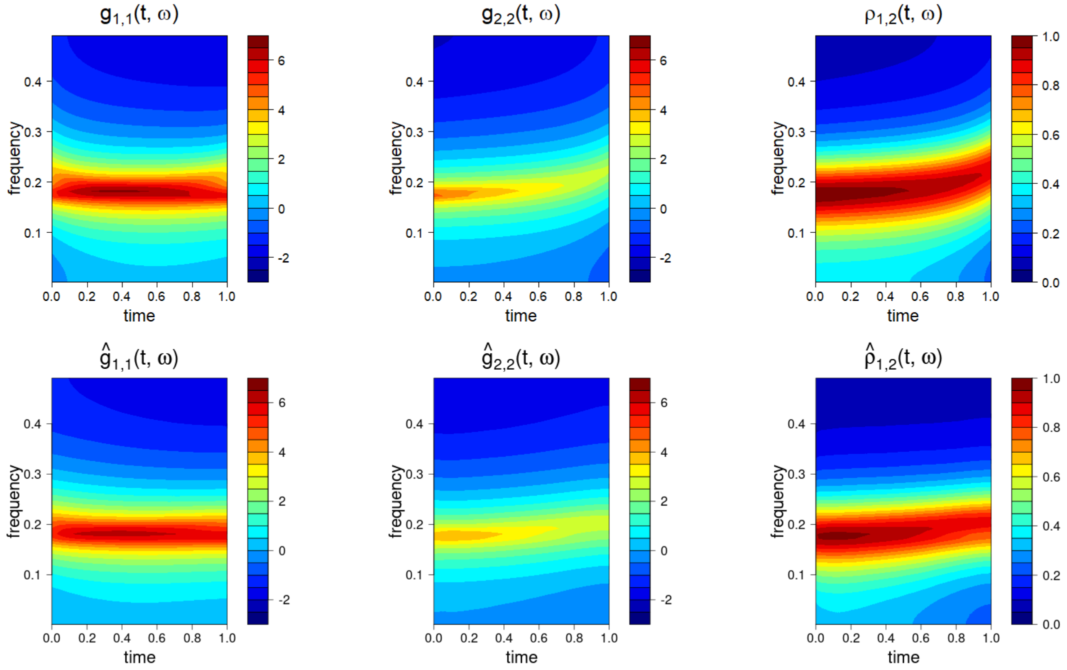

Figure 4.2: Case with $a_{1,1,2,t} = -0.8$ and $a_{2,1,2,t} = 0$ for all t. Top: True log spectral density $g_{11}(t, \omega)$ (left), true log spectral density $g_{22}(t, \omega)$ (middle), true squared coherence $\rho^2_{1,2,}(t, \omega)$ (right). Bottom: Posterior mean of $\hat{g}_{11}(t, \omega)$ (left), posterior mean of $\hat{g}_{22}(t, \omega)$ (middle), posterior mean of $\rho^2_{1,2}(t, \omega)$ (right)

and (ii). Estimates shown were obtained with the MCMC algorithm. The posterior means of the log spectral densities displayed in the figures are obtained by averaging over the 100 simulated data sets. The bivariate shrinkage TV-VPARCOR model not only is able to adequately capture the structure of the individual spectral densities and that of the squared coherences, but also shrinks those autoregressive coefficients to zero after lag 2 (Figure B.1, B.3 and B.5 in the Supplementary Material). We also show the results obtained with the variational Bayes algorithm in the Supplementary Material Section B.3.

To compare the performance of the shrinkage TV-VPARCOR and TV-VPARCOR 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. Table 4.1 summarizes the mean and standard deviations of the ASE based on ASE_n for the three scenarios. The shrinkage TV-VPARCOR(5) model with the MCMC and variational Bayes algorithms outperform TV-VPARCOR(5) in all scenarios. Especially, for case (i), shrinkage priors shrink all the cross-elements in the PARCOR coefficients to zero so that the estimated average squared coherence shown in Figure 4.1 is shrinked to zero. Table 4.2 presents the computation times for the TV-VPARCOR with non-shrinkage priors, as well as the TV-VPARCOR with Bayessian Lasso and Double Gamma TV-VPARCOR(5) priors obtained with the MCMC and variational Bayes algorithms averaging over the 100 realizations in each case. Note that the ISVB algorithm converges quickly before the maximum iterations. The models were run on a system with CPU Inter(R) i7-9750H with 6 cores and 2.60GHz.

4.4.2 20-Dimensional TV-VAR(1)

We show the performance of our shrinkage TV-VPARCOR models with sparse synthetic data. We simulate synthetic data from the 20-dimensional non-stationary TV-VAR(1) process with $T = 300$ used in Section 3.3.2 of Chapter 3. The (i, j) elements of the matrix of VAR coefficients at time t, are given as follows:

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

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

Note that in Section 3.3.2 of Chapter 3, the optimal order for the TV-

Table 4.1: Mean ASE values and corresponding standard deviations (in parentheses) for the log-spectral densities and squared coherences obtained from shrinkage TV-VPARCOR models with the MCMC and variational Bayes algorithms, and for TV-VPARCOR models with no shrinkage (top row for each case). All models assumed $P = 5$.

	Case (i): $a_{1,1,2,t} = a_{2,1,2,t} = 0$		
Model	g_{11}	g_{22}	ρ_{12}^2
Non-shrinkage conjugate	0.0363(0.0178)	0.0468(0.0246)	0.0031(0.0042)
Bayesian Lasso (MCMC)	0.0317(0.0118)	0.0390(0.0144)	0.00078(0.00077)
Bayesian Lasso (VI)	0.0307(0.0128)	0.0390(0.0141)	0.00057(0.0007)
Double Gamma (MCMC)	0.0308(0.0143)	0.0387(0.0155)	0.00035(0.00066)
Double Gamma (VI)	0.0324(0.0138)	0.0402(0.0149)	0.00049(0.0007)
	Case (ii): $a_{1,1,2,t} = -0.8, a_{2,1,2,t} = 0$		
Model	g_{11}	g_{22}	ρ_{12}^2
Non-shrinkage conjugate	0.0645(0.0331)	0.0506(0.0175)	0.0097(0.0127)
Bayesian Lasso (MCMC)	0.0448(0.0138)	0.0436(0.0129)	0.0048(0.0021)
Bayesian Lasso (VI)	0.0399(0.0118)	0.0421(0.0119)	0.0042(0.0020)
Double Gamma (MCMC)	0.0389(0.0148)	0.0406(0.0158)	0.0039(0.0024)
Double Gamma (VI)	0.0388(0.0111)	0.0426(0.0117)	0.0038(0.0019)
	Case (iii): $a_{1,1,2,t} = r_{3,t}, a_{2,1,2,t} = r_{4,t}$		
Model	g_{11}	g_{22}	ρ_{12}^2
Non-shrinkage conjugate	0.1006(0.2338)	0.0809(0.1493)	0.0099(0.0110)
Bayesian Lasso (MCMC)	0.0794(0.0409)	0.0694(0.0261)	0.0054(0.0033)
Bayesian Lasso (VI)	0.0439(0.0210)	0.0457(0.0138)	0.0041(0.0028)
Double Gamma (MCMC)	0.0527(0.0368)	0.0504(0.0231)	0.0041(0.0040)
Double Gamma (VI)	0.0430(0.0219)	0.0479(0.0132)	0.0040(0.0027)

VPARCOR analysis was $P = 1$ according to their DIC criterion. Therefore, we fit shrinkage TV-VPARCOR models considering $P = 1$, to make a fair comparison across the different models. We implement the Bayesian Lasso and the Normal-Gamma priors. Here we present approximate inference results obtained with the variational Bayes algorithm. For both prior distributions, we assumed prior hyperparameters $d_{f,i} = d_{b,i} = e_{f,i} = e_{f,i} = 0.001$, for $i = 1, 2$. In addition, $n_{f,m,0} = n_{b,m,0} = 1$ and $s_{i,f,m,0} = s_{i,b,m,0} = 2$. We run the variational Bayes algorithm with a maximum number of $30,000$ iterations. The algorithm stops when

Table 4.2: Average computation times (in seconds) for TV-VPARCOR, TV-VPARCOR with Bayesian Lasso and Double Gamma priors

Model	Case (i)	Case (ii)	Case (iii)
TV-VPARCOR(5)	2.56	5.62	6.38
Bayesian Lasso (MCMC)	560.95	555.95	564.4
Bayesian Lasso (VI)	49.64	46.17	46.56
Double Gamma (MCMC)	568.09	563.67	564.44
Double Gamma (VI)	47.47	46.41	47.2

the change in the L_2-norm of the model parameters is smaller than 10^{-3}.

Figure 4.3 shows the true and estimated log spectral densities from the shrinkage TV-VPARCOR model for 4 components of the 20-dimensional time series, namely, components $1, 2, 8$, and 15. Figure 4.4 shows the true and estimated squared coherences between components 1 and 5, components 2 and 15, components 5 and 12, and components 15 and 20. The time-frequency characteristics of the original multivariate non-stationary time series process can be captured by the shrinkage TV-VPARCOR models. The estimated squared coherence is shrunk to zero by TV-VPARCOR models. Moreover, the shrinkage TV-VPARCOR can detect the slowly time-varying of spectral density, like the second component of time series.

To evaluate the shrinkage effects of shrinkage PARCOR model, we compute the mean squared error on the TV-VAR coefficients obtained from the TV-VPARCOR models with the Bayesian Lasso and the Normal Gamma priors, as well as for the non-shrinkage conjugate priors. Table 4.3 shows that both shrinkage PARCOR models outperform the original TV-VPARCOR model with conjugate non-shrinkage priors because of the shrinkage effects.

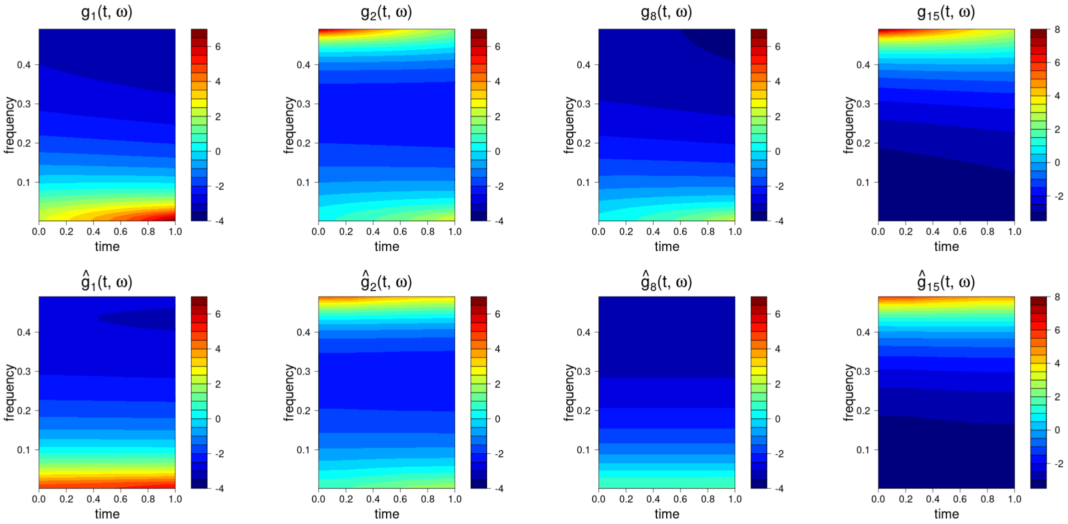

Figure 4.3: Top: True log spectral densities of time series components 1, 2, 8 and 15. Bottom: estimated log spectral densities of the same components obtained from the shrinkage TV-VPARCOR approach with model order 1.

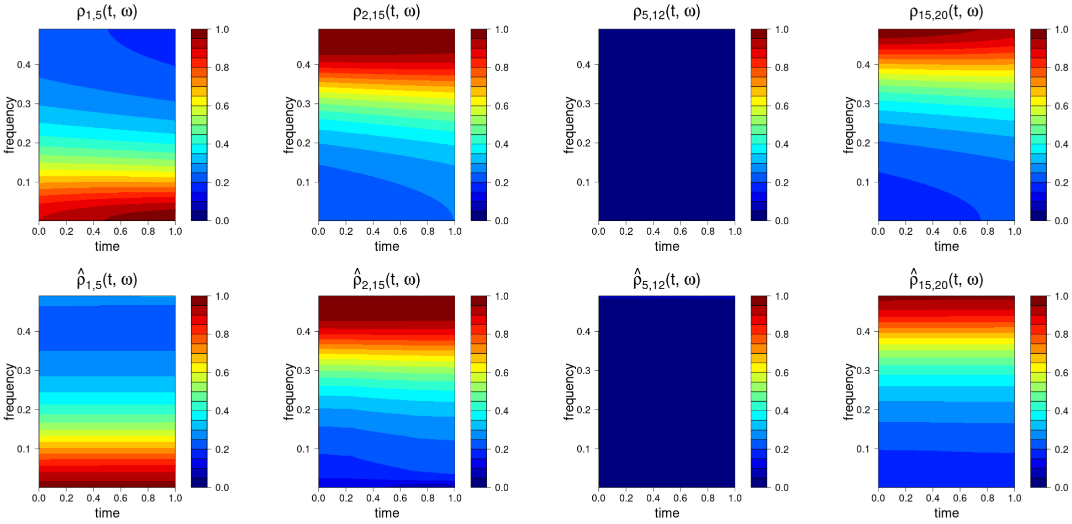

Figure 4.4: Top: True squared coherence between components 1 and 5, 2 and 15, 5 and 12, and 15 and 20. Bottom: Corresponding estimated squared coherences obtained from the shrinkage TV-VPARCOR model.

Table 4.3: The MSE on TVAR coefficients

Model	MSE
Bayesian Lasso	0.002734
Normal Gamma	0.0009651
TV-VPARCOR	0.014

4.5 Analysis of Multi-Location Wind Data

In Section 3.4.2 of Chapter 3, we analyze median wind speed and direction measurements taken every 4 hours from 1 June 2010 to 12 August in 3 stations in the Northern California, namely: Monterey (MRY), Salinas (SNS) and Watsonville (WVI). Therefore, the total number of series in this data set was 6, since each location has measurements for two wind components. Here we illustrate the performance of the shrinkage TV-VPARCOR models and the efficient variational inference by considering a larger data set with 10 locations resulting in a 20-dimensional time series data set. In particular, we consider the 3 locations above, i.e., Monterey (MRY), Salinas (SNS) and Watsonville (WVI) and add 7 more locations: Fairfield (SUU), Vacaville (VCB) and Napa County (APC) near the San Francisco Bay area, and Sacramento (SAC), West Sacramento (SMF), Fresno (FCH) and Merced (MCE), near the Central Valley area. The station locations are shown in Figure 4.5. For each station we have measurements of the wind direction and wind speed which we transform into the East-West component (X component) and the North-South component (Y component).

We apply the shrinkage TV-VPARCOR model with the Normal-Gamma prior for joint analysis of the $20-$dimensional time series corresponding to the wind time series components for the 10 stations. As suggested in Section 3.4.2 of Chapter 3, the optimal model order chosen by the approximate DIC calculation is $P = 3$. The posterior inference is obtained by the variational Bayes algorithm. We set

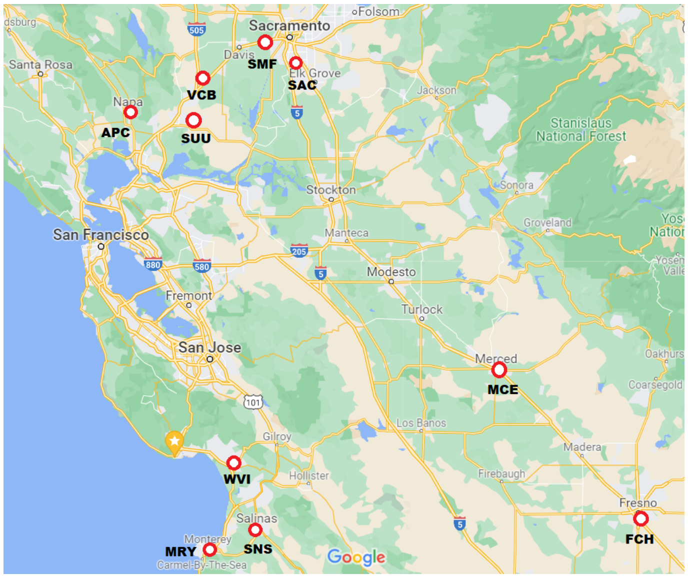

Figure 4.5: The 10 locations of stations in northern and central California.

the prior hyperparameters as follows: $n_{f,m,0} = n_{b,m,0} = 1$, $s_{i,f,m,0} = s_{i,b,m,0} = 5$ for all m, $d_{f,i} = d_{b,i} = e_{f,i} = e_{b,i} = 10^{-4}$ for $i = 1, 2$. We set up the maximum number of iterations to $10,000$. The algorithm stops when the change of L_2-norm of all parameters is smaller than 10^{-5}.

Figure 4.6 shows the estimated log spectral densities of the East-West component (X component) and the North-South component (Y component) for Monterey, Salinas, Watsonville, Fairfield, Vacaville and Napa County. There is a strong 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, which is consistent with results obtained in Chapter 3 Section 3.4.2. This dominant quasi-periodic behavior around 24 hours also appears after the beginning of July in the north-south component in Napa County. A less consistent quasi-periodic daily behavior can be also found in the East-West component of Vacaville and the North-South component of Fairfield and Vacaville after mid of July. Due to the shrinkage priors the estimated power of the log spectral den-

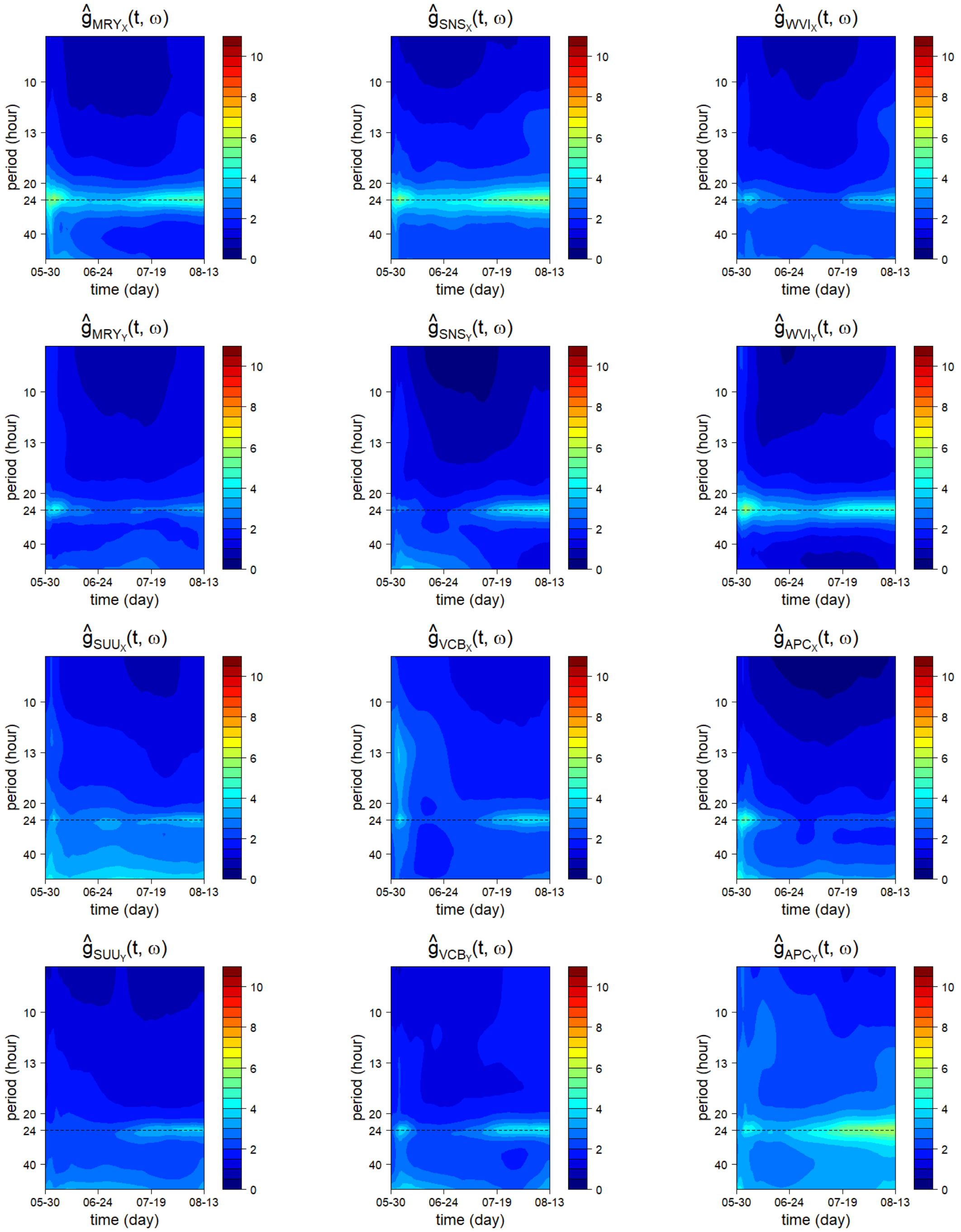

Figure 4.6: Estimated log spectral densities of the East-West (X) components (rows 1 and 3) and North-South (Y) components (rows 2 and 4) for 6 stations, such as Monterey, Salinas, Watsonville, Fairfield, Vacaville and Napa County.

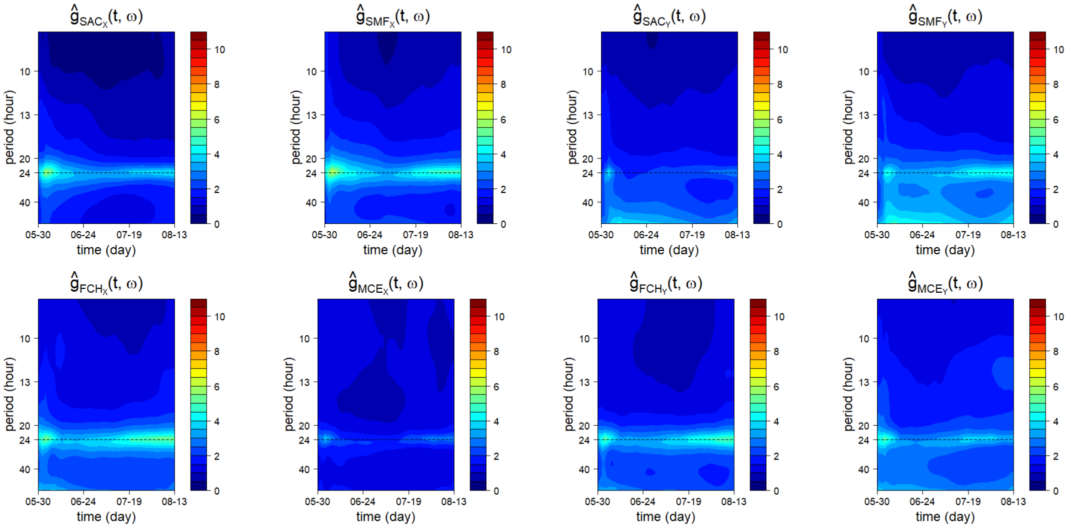

Figure 4.7: Top row: Estimated log-spectral densities of the East-West (X) and North-South (Y) components for Sacramento and West Sacramento. Bottom row: Estimated log-spectral densities of the East-West (X) and North-South (Y) components for Fresno and Merced.

sities is smaller than the estimated power from TV-VPARCOR models without shrinkage priors.

Figure 4.7 shows the log spectral densities of the East-West component and North-South component for Sacramento, West Sacramento, Fresno and Merced. The quasi-periodic daily pattern can be detected in East-West component for Sacramento, West Sacramento and Fresno, as well as North-South component for West Sacramento, Fresno and Merced.

We are particularly interested in the period between 22 hours and 26 hours. The proposed shrinkage TV-VPARCOR model provides a method to conduct inference on period-collapsed functionals. For example, the time-varying period-collapsed spectral matrices are given by

$$\boldsymbol{g}^{[22,26]}(t) = \int_{\frac{4}{26}}^{\frac{4}{22}} \boldsymbol{g}(t,\omega)d\omega,$$

where $\boldsymbol{g}(t,\omega)$ is defined in Chapter 1 Section 1.1.1. Similarly, the time-varying period-collapsed squared coherences between time series j and k is defined as

$$\tilde{\rho}^2_{jk}(\omega) = \left| \int_{\frac{4}{26}}^{\frac{4}{22}} g_{jk}(t,\omega) d\omega \right|^2 / \left\{ g_{jj}^{[22,26]}(t) g_{kk}^{[22,26]}(t) \right\},$$

where $g_{jj}^{[22,26]}(t) = \int_{\frac{4}{26}}^{\frac{4}{22}} g_{jj}(t,\omega) d\omega$ is the time-varying period-collapsed power of time series j.

We can explore the behavior of the time-varying period collapsed squared coherences across wind components for different locations. Figure 4.8 shows the time-varying 22-26 hours period-collapsed squared coherence of the East-West (X) components (left plot) between Salinas and the rest of the 9 locations. Similarly, the right plot in Figure 4.8 shows the $22-26$ hours period-collapsed squared coherence of the North-South (Y) components between Salinas and rest of the 9 locations. Figure 4.9 shows similar plots of the time-varying 22-26 hours period-collapsed squared coherence between Sacramento and rest of locations. For the East-West (X) component, West Sacramento, Sacramento and Monterey, are highly correlated with Salinas at period band between 22 hours and 26 hours for the entire time period. Sacramento and West Sacramento are highly correlated throughout the entire period and they also happen be very close locations. The squared coherence between Salinas and Napa is strong at the beginning of June, and then is decreasing over time. The period-collapsed squared coherences of the North-South components between Salinas and rest of locations are increasing after the mid of June. In addition, there are also increasing period-collapsed squared coherences of the North-South component between Sacramento and rest of locations after mid of June.

In order to evaluate the goodness of fit, we compute the mean square error (MSE) on the fitted residuals over the 20 time series components. The MSE

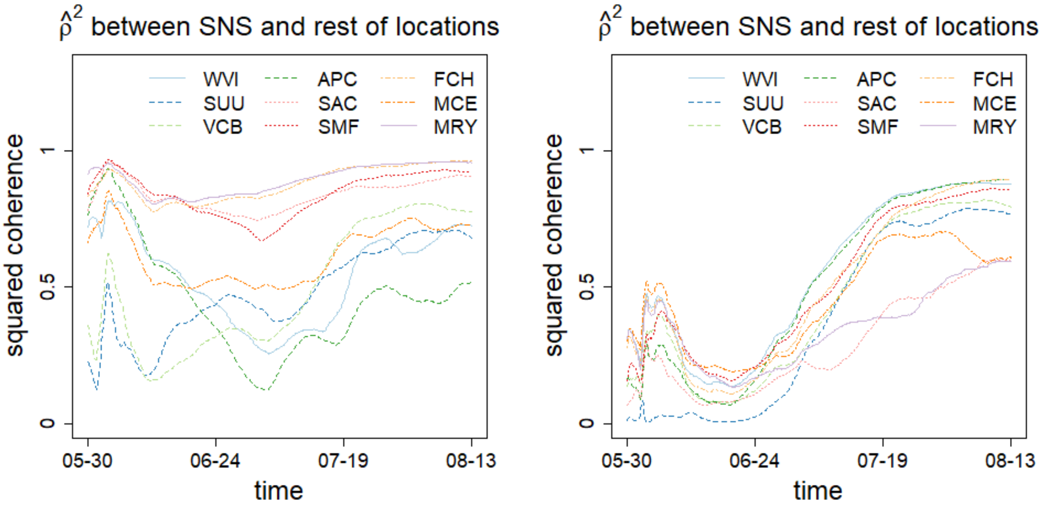

Figure 4.8: The time-varying period-collapsed (period between 22 hours and 26 hours) squared coherence of East-West (X) component (left plot) and North-South (Y) component (right plot) between Salinas and rest of 9 locations. The solid lines represent Monterey Bay Area. The dashed lines represent Bay Area. The dotted line represents Sacramento Valley. The dotdash line represents Central Valley.

was computed using the posterior mean of the TV-VPARCOR model parameters under the three different prior settings, i.e., under the Bayesian lasso prior, the Normal Gamma prior, and the non-shrinkage conjugate prior (Chapter 3). Once again we assume the same model order $P = 3$ for the three different prior settings. Table 4.4 shows that both shrinkage priors have the smaller MSE values than those obtained under the non-shrinkage conjugate prior, which suggests that the regularizing effect of the shrinkage priors on the TV-VPARCOR parameters improves the fit.

Table 4.4: MSE of the smoothed time series

Model	MSE
Bayesian Lasso	3.96
Normal Gamma	3.98
TV-VPARCOR	4.14

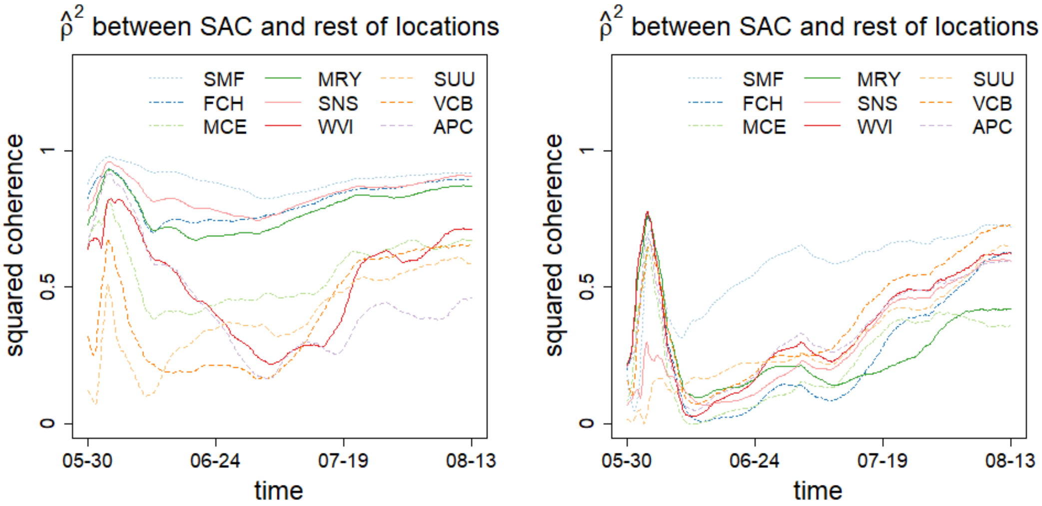

Figure 4.9: The time-varying period-collapsed (period between 22 hours and 26 hours) squared coherence of East-West (X) component (left plot) and North-South (Y) component (right plot) between Sacramento and rest of 9 locations. The solid lines represent Monterey Bay Area. The dashed lines represent Bay Area. The dotted line represents Sacramento Valley. The dotdash line represents Central Valley.

4.6 Discussions

In this chapter, we propose a Bayesian shrinkage time-varying PARCOR framework for multivariate time series. The shrinkage prior setting allows us to automatically reduce time-varying PARCOR coefficients to static ones. We consider the normal-gamma prior (Griffin and Brown, 2010) and Bayesian Lasso prior (Belmonte et al., 2014) and use the non-centered parameterization (Frühwirth-Schnatter and Wagner, 2010) on the dynamic multivariate PARCOR domain. An MCMC algorithm for full posterior inference is developed. In addition, an ISVB algorithm is also developed for fast approximate posterior inference. We illustrate the performance of the proposed model by comparing the accuracy of the proposed TV-VPARCOR shrinkage model with that of TV-VPARCOR model (Chapter 3) in two simulation studies. In addition, we show that the TV-VPARCOR model with shrinkage priors can be successfully used to analyze multivariate non-stationary wind component time series data, as illustrated in Section 4.5.

Chapter 5

Conclusion

In this thesis, we develop a series of new flexible and computational efficient methods for analysis and forecasting of multivariate and multiple non-stationary time series in the PARCOR domain. First, we extend the univariate PARCOR model (Yang et al., 2016) to the multivariate case. This approach showcases the advantages of Bayesian lattice filter structure in terms of computational efficiency for high-dimensional time series analysis. We develop a fast and accurate approximate posterior inference scheme on the TV-VPARCOR coefficients by adapting the approach of Triantafyllopoulos (2007) for estimating the covariance-variance matrices at the observational level for the DLM PARCOR representations at each stage of the lattice filter. We also specify the innovations at system levels using discount factors. A couple of quantitative and qualitative model selection criteria (e.g., DIC) are developed under the PARCOR framework. We use our model for analysis and prediction of multi-location wind components time series data over a period of time. These models allow us to detect quasi-periodic patterns through the estimated spectral densities of each of the wind time series components, and are useful for short-term forecasting. We also use our model to infer the time-frequency characteristic of multi-channel EEG data from patients who received

ECT. Second, we extend univariate PARCOR model to the case in which the data analysis involves multiple, rather than multivariate, time series. A hierarchical PARCOR model is developed, which is more parsimonious and computational efficient than the standard hierarchical state-space TVAR representations. We analyze multiple EEG signals recorded from repeated trials to detect the common underlying features without losing information. We illustrate these models in the analysis of multi-trial brain signals. Finally, we propose a TV-VPARCOR framework with Bayesian Lasso priors Park and Casella (2008) and Normal Gamma priors Griffin and Brown (2010). We consider the non-centered parameterization (Frühwirth-Schnatter and Wagner, 2010) on the TV-VPARCOR state-space representation. These shrinkage priors allow us to shrink time-varying parameters to static ones when the model is overfitting. We illustrate the performance of the shrinkage TV-VPARCOR models in the analysis of multi-location wind component temporal data from Northern California. The code used in this thesis is available at `https://github.com/Jayzhaowj/PARCOR`.

Future research on PARCOR domain is promising. First, the multivariate PARCOR model can be extended by adding latent structural level. For example, in the analysis of group-level EEG data presented in Chapter 2, the hierarchical PARCOR model allows us to extract the common underlying time-frequency features from the multiple recorded brain signals for each cortical cluster. However, it is impossible for this model to also capture the relationships between cortical clusters. A possible approach to address this problem can be to consider matrix-variate hierarchical dynamic models (MHDM) proposed by Landim and Gamerman (2000). The observation level of MHDM can be modeled by the multivariate Bayesian lattice filters. The TV-VPARCOR coefficients in the structural level can be decomposed into group-level effects and series-specific effects easily thanks to the parsimonious PARCOR representations. At the evolution level, the random walk process can be implemented on the corresponding parameters. We will also consider regularized TV-VPARCOR models by extending the approach of Proietti and Giovannelli (2018), which proposes a regularized Durbin-Levinson algorithm that imposes penalty weights controlled by trapezoidal kernels on the PARCOR coefficients, to the multivariate case. Finally, we will consider spatio-temporal extensions of the lattice filter approach to more adequately deal with spatio-temporal data such as the multi-location wind component data.

Appendix A

Supplementary Materials

A.1 Supplementary figures for the 20-dimensional TV-VAR(1) example

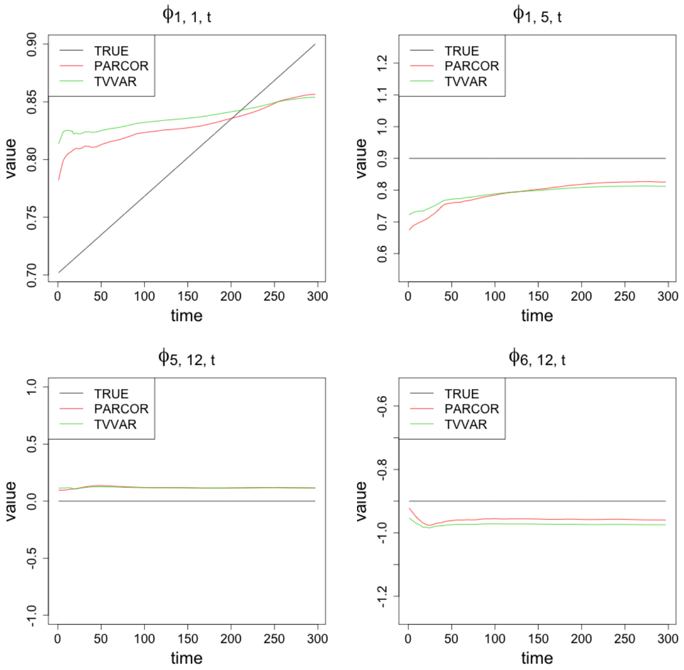

Figure A.1: True and estimate traces of the TV-VAR coefficients $\phi_{1,1,t}$, $\phi_{1,5,t}$, $\phi_{6,12,t}$ and $\phi_{5,12,t}$ obtained from the TV-VPARCOR and TV-VAR approaches.

A.2 Simulation study: Piecewise time series process

We simulated $T = 1024$ observations from the following piecewise TV-VAR process:

$$\boldsymbol{x}_t = \begin{cases} \boldsymbol{\Phi}_{1,t}\boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t, & 1 \leq t \leq 512, \\ \boldsymbol{\Phi}_{2,t}\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_{3,t}\boldsymbol{x}_{t-2} + \boldsymbol{\epsilon}_t, & 513 \leq t \leq 1024, \end{cases}$$

with

$$\boldsymbol{\Phi}_{1,t} = \begin{pmatrix} 0.35 & -0.85 \\ 0 & 0.5 \end{pmatrix}, \quad \boldsymbol{\Phi}_{2,t} = \begin{pmatrix} -0.45 & -0.4 \\ 0 & -0.32 \end{pmatrix}, \quad \boldsymbol{\Phi}_{3,t} = \begin{pmatrix} -0.35 & 0 \\ 0.85 & -0.21 \end{pmatrix},$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$.

Table A.1 shows a comparison between the mean ASE values for a TV-VPARCOR model of order 2 and a TV-VAR model of order 2 for these simulated data. We see that the TV-VPARCOR model has a much better performance than the TV-VAR and is able to better adapt to the abrupt changes in the simulated data.

Table A.1: Mean ASE values and corresponding standard deviations (in parentheses) obtained from TV-VPARCOR and TV-VAR models of order 2 for the piecewise TV-VAR simulated data.

	Piece-wise AR process		
Model	g_{11}	g_{22}	ρ^2_{12}
TV-VPARCOR	0.180(0.057)	0.055(0.026)	0.005(0.003)
TV-VAR	0.985(0.154)	0.482(0.241)	0.097(0.054)

Figure A.2 shows the true and estimated TV-VPARCOR log-spectral density and squared coherence. Uncertainty measures can also be obtained for any of these quantities. For instance, Figure A.3 shows 95% posterior intervals for the squared coherence. Once again, these plots show that the TV-VPARCOR model is able to detect regime changes at unknown time points.

A.3 Supplementary figures for the multi-channel EEG analysis

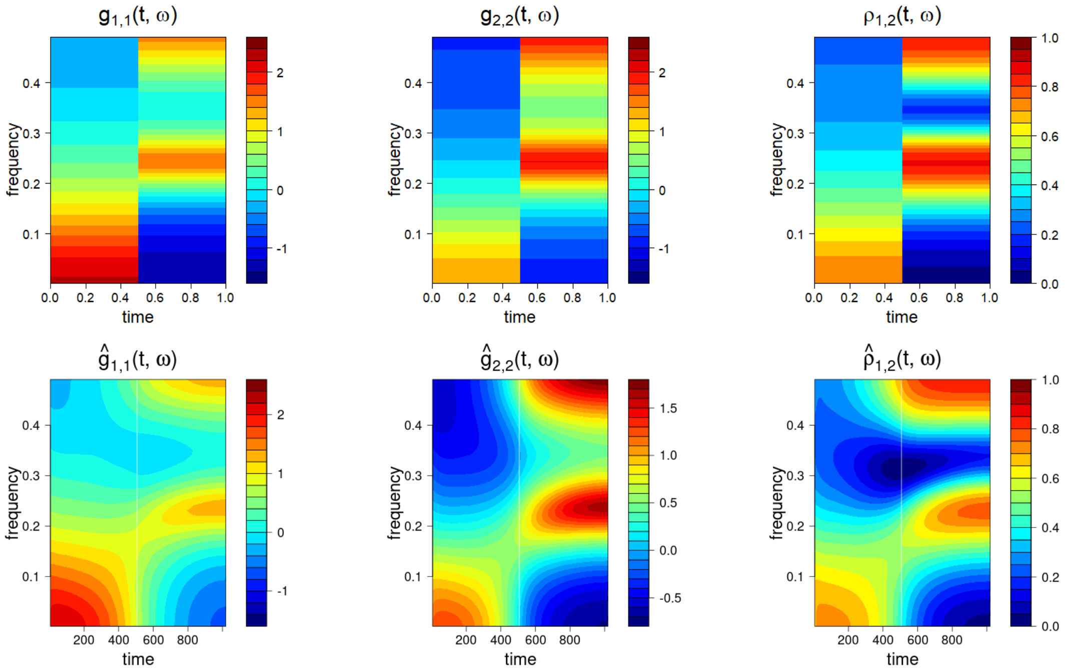

Figure A.2: Case: Piecewise Stationary VAR Process. Top: True log spectral density $g_{1,1}(t,\omega)$ (left), true log spectral density $g_{2,2}(t,\omega)$ (middle), true squared coherence $\rho^2_{1,2}(t,\omega)$ (right). Bottom: Estimated $\hat{g}_{1,1}(t,\omega)$ (left), estimated $\hat{g}_{2,2}(t,\omega)$ (middle), estimated $\hat{\rho}^2_{1,2}(t,\omega)$ (right).

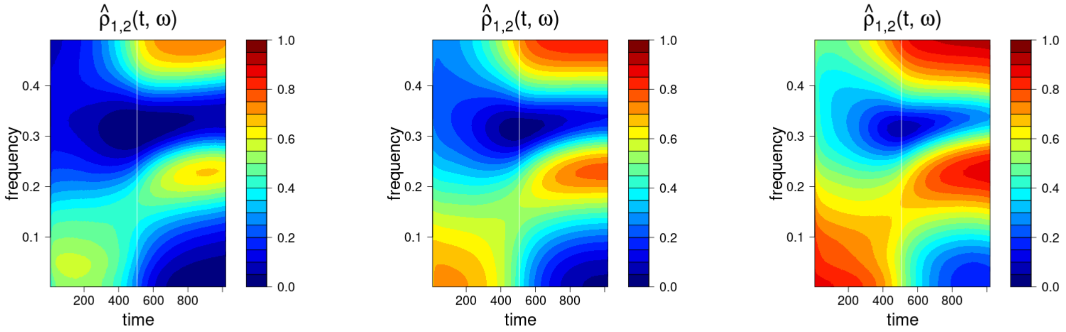

Figure A.3: Case: Piecewise VAR Process. Left: Lower 95% posterior band for the squared coherence between two time series. Middle: Posterior mean of the squared coherence. Right: Upper 95% posterior band for the squared coherence between two time series.

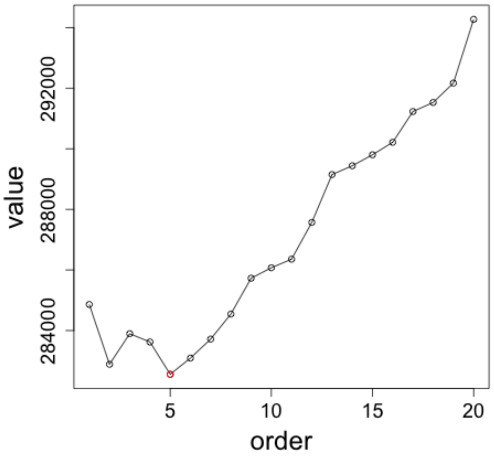

Figure A.4: Multi-channel EEG data: DIC values.

A.4 Supplementary figures for the multi-location wind component data

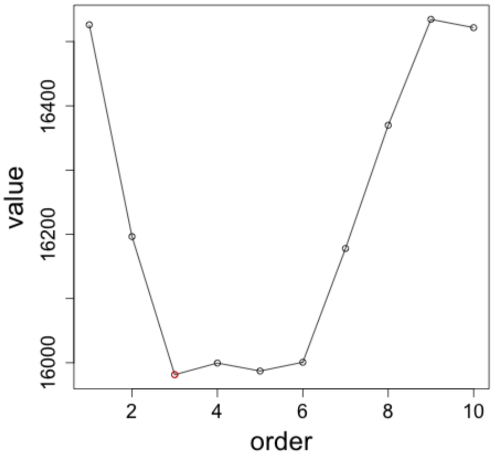

Figure A.5: Multi-location wind components: DIC values.

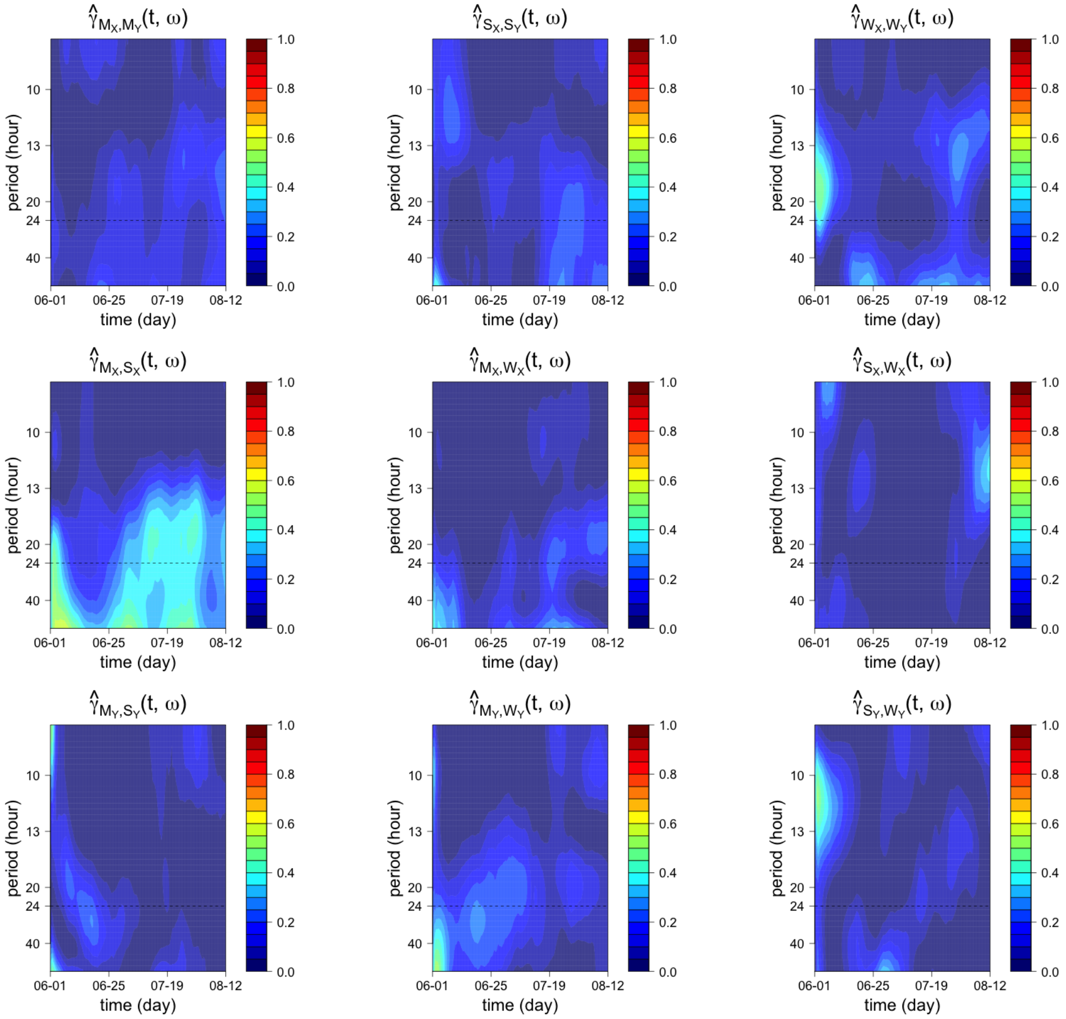

Figure A.6: Top row: The partial coherences between East-West (X) Components and North-South (Y) Components in Monterey, Salinas and Watsonville. Middle row: The partial coherences between Monterey and Salinas, Monterey and Watsonville, as well as Salinas and Watsonville in terms of East-West (X) components. Bottom row: The partial coherences between Monterey and Salinas, Monterey and Watsonville, as well as Salinas and Watsonville in terms of North-South (Y) components.

Appendix B

Algorithm details

B.1 MCMC Algorithm Details

B.1.1 MCMC Forward Filtering Backwards Sampling

Let illustrate the detail of sampling the forward latent states $\tilde{\boldsymbol{\lambda}}_{i,f,m,t}$ for $i = 1, \ldots, K$ by Forward Filtering Backward Sampling (FFBS) (see e.g. Carter and R.Kohn, 1994, Frühwirth-Schnatter, 1994). The procedure of sampling the backward latent states $\tilde{\boldsymbol{\lambda}}_{i,b,m,t}$. The algorithm is implemented for a slight modification of the non-centered TVP model (4.7) and (4.5), given by:

$$\bar{f}_{it}^{(m-1)} = \tilde{\boldsymbol{\lambda}}_{i,f,m,t}\boldsymbol{F}_{i,f,m,t} + \tilde{f}_{it}^{(m)}, \qquad \tilde{f}_{it}^{(m)} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{i,f,m,t}), \tag{B.1}$$

$$\tilde{\boldsymbol{\lambda}}'_{i,f,m,t} = \tilde{\boldsymbol{\lambda}}'_{i,f,m,t-1} + \tilde{\boldsymbol{w}}_{i,f,m,t}, \qquad \tilde{\boldsymbol{w}}_{i,f,m,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{K+i-1}), \tag{B.2}$$

with transformed response $\bar{f}_{it}^{(m-1)} = f_{it}^{(m-1)} - \boldsymbol{\lambda}_{i,f,m}\boldsymbol{x}_{f,t}^{(m-1)}$, and the regressor $\boldsymbol{F}_{i,f,m,t} = \sqrt{\boldsymbol{\Theta}}_{i,f,m}\boldsymbol{x}_{f,t}^{(m-1)}$. Let $D_{i,f,m,t}$ denote all the information available up to time t at stage m for the ith equation of the forward model, with $D_{i,f,m,t} = \{D_{i,f,m,t-1}, f_{it}^{(m-1)}\}$.

- For $t = 1, \ldots, T$, and $i = 1, \ldots, K$, forward filtering:
  - Posterior at $t-1 : p(\tilde{\boldsymbol{\lambda}}_{i,f,m,t-1} | D_{i,f,m,t-1})$ is approximately distributed as
  $$\mathcal{N}(\boldsymbol{m}_{i,f,m,t-1}, \boldsymbol{C}_{i,f,m,t-1}),$$
  - The one-step ahead forecast mean and variance at time t are given by:
  $$E(\bar{f}_{it}^{(m-1)} | D_{i,f,m,t-1}) \approx \boldsymbol{F}_{i,f,m,t}^T \boldsymbol{m}_{i,f,m,t-1},$$
  and
  $$V(\bar{f}_{it}^{(m-1)} | D_{i,f,m,t-1}) \approx Q_{i,f,m,t} = \boldsymbol{F}_{i,f,m,t}^T \boldsymbol{R}_{i,f,m,t} \boldsymbol{F}_{i,f,m,t} + \sigma_{i,f,m,t-1}^2,$$
  where $\boldsymbol{R}_{i,f,m,t} = \boldsymbol{C}_{i,f,m,t-1} + \boldsymbol{I}_K$.
  - Apply Bayes' theorem and the above equations, we can obtain the approximate posterior filtered distribution at time t as $\tilde{\boldsymbol{\lambda}}_{i,f,m,t} | D_{i,f,m,t} \approx \mathcal{N}(\boldsymbol{m}_{i,f,m,t}, \boldsymbol{C}_{i,f,m,t})$,
  $$\begin{aligned} \boldsymbol{m}_{i,f,m,t} &= \boldsymbol{m}_{i,f,m,t-1} + \boldsymbol{U}_{i,f,m,t} e_{i,f,m,t}, \\ \boldsymbol{C}_{i,f,m,t} &= \boldsymbol{R}_{i,f,m,t} - \boldsymbol{U}_{i,f,m,t} \boldsymbol{U}'_{i,f,m,t} Q_{i,f,m,t}, \\ \boldsymbol{U}_{i,f,m,t} &= \boldsymbol{R}_{i,f,m,t} \boldsymbol{F}_{i,f,m,t} Q_{i,f,m,t}^{-1}. \end{aligned}$$
- For $i = 1, \ldots, K$, backward sampling $p(\tilde{\boldsymbol{\lambda}}_{i,f,m,t} | D_{i,f,m,t|T}) :$
  - For $t = T$, sample $\tilde{\boldsymbol{\lambda}}_{i,f,m,t} | D_{i,f,m,T} \sim \mathcal{N}(\boldsymbol{m}_{i,f,m,t}, \boldsymbol{C}_{i,f,m,t})$.
  - For $t = T-1, \ldots, 1$, sample $\tilde{\boldsymbol{\lambda}}_{i,f,m,t} | D_{i,f,m,T} \sim \mathcal{N}(\boldsymbol{m}_{i,f,m,t|T}, \boldsymbol{C}_{i,f,m,t|T})$,

where

$$\boldsymbol{m}_{i,f,m,t|T} = \boldsymbol{m}_{i,f,m,t} + \boldsymbol{C}_{i,f,m,t}\boldsymbol{R}^{-1}_{i,f,m,t+1}(\tilde{\boldsymbol{\lambda}}_{i,f,m,t+1} - \boldsymbol{m}_{i,f,m,t}),$$
$$\boldsymbol{C}_{i,f,m,t|T} = \boldsymbol{C}_{i,f,m,t} - \boldsymbol{C}_{i,f,m,t}\boldsymbol{R}^{-1}_{i,f,m,t+1}\boldsymbol{C}_{i,f,m,t}.$$

B.1.2 Sample the static parameters

Conditional on the state process $\tilde{\boldsymbol{\lambda}}_{i,f,m,t}$, the observation equation (4.7) of the non-centered state space model defines an expanded regression model:

$$f_{it}^{(m-1)} = \boldsymbol{Z}_{i,f,m,t}\boldsymbol{\Psi}_{i,f,m} + \boldsymbol{\epsilon}_{i,f,m,t}, \boldsymbol{\epsilon}_{i,f,m,t} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{f,m}\right), \tag{B.3}$$

with regression coefficient $\boldsymbol{\Psi}_{i,f,m} = (\boldsymbol{\lambda}_{i,f,m}, \sqrt{\boldsymbol{\Theta}}_{i,f,m})'$ and $\boldsymbol{Z}_{i,f,m,t} = \left((\boldsymbol{x}^{(m-1)}_{f,t})', \tilde{\boldsymbol{\lambda}}_{i,f,m,t} \circ (\boldsymbol{x}^{(m-1)}_{f,t})'\right)$ is a $1 \times 2K$ regressor ith row vector. Where notation $\circ$ represents the element wise multiplication. Under the conjugate prior $\boldsymbol{\Psi}_{i,f,m} \sim \mathcal{N}(\boldsymbol{\Psi}_{i,0}, \boldsymbol{\Sigma}^{\psi}_{i,f,m,0})$, where $\boldsymbol{\Psi}_{i,0} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}^{\psi}_{i,f,m,0} = \text{Diag}\left(\tau^2_{i,f,m}, \xi^2_{i,f,m}\right)$, it follows that the conditional posterior distribution $p(\boldsymbol{\Psi}_{i,f,m}|\cdot)$ is a multivariate Gaussian distribution,

$$\boldsymbol{\Psi}_{i,f,m}|\tilde{\boldsymbol{\Gamma}}_{i,f,m,1:T}, \boldsymbol{\tau}^2_{i,f,m}, \sigma^2_{i,f,m}, \quad f^{(m-1)}_{i,1:T} \sim \mathcal{N}_{2(K+i-1)}\left(\boldsymbol{\Psi}_{i,T,f,m}, \boldsymbol{\Sigma}^{\psi}_{i,T,f,m}\right),$$

with

$$\boldsymbol{\Psi}_{i,T,f,m} = \boldsymbol{\Sigma}^{\psi}_{i,T,f,m}\left(\sum_{t=1}^{T}\sigma^{-2}_{i,f,m} f_{it}^{(m-1)}\boldsymbol{Z}'_{i,f,m,t}\right),$$
$$\boldsymbol{\Sigma}^{\phi}_{i,T,f,m} = \left(\sum_{t=1}^{T}\sigma^{-2}_{i,f,m}\boldsymbol{Z}'_{i,f,m,t}\boldsymbol{Z}_{i,f,m,t} + \boldsymbol{\Sigma}^{-1,\psi}_{i,f,0}\right)^{-1}.$$

B.1.3 Ancillarity-Sufficiency Interweaving Strategy

For each $i = 1, \dots, K,\ j = 1, \dots, K$, redraw the constant $a_{ij,f,m}$ and the square root of the process variance $\sqrt{\theta}_{ij,f,m}$ through interweaving into the state equation of the centered parameterization:

3.a Use the transformation $\lambda^{(m)}_{ij,f,m,t} = \lambda_{ij,f,m} + \tilde{\lambda}_{ij,f,m,t}\sqrt{\theta}_{ij,f,m}$ to match the draws of the latent process $\tilde{\lambda}_{ij,f,m,0}, \dots, \tilde{\lambda}_{ij,f,m,T}$ in the non-centered to the latent process $\lambda^{(m)}_{ij,f,m,0}, \dots, \lambda^{(m)}_{ij,f,m,T}$ in the centered parameterization and store the sign of $\sqrt{\theta}_{ij,f,m}$.

3.b Update $\lambda_{ij,f,m}$ and $\theta_{ij,f,m}$ in the centered paramterization by sampling $\theta^{new}_{ij,f,m}$ from generalized inverse Gaussian posterior

$$\theta_{ij,f,m}| \dots \sim \mathcal{GIG}\left(\frac{T}{2}, \frac{1}{\xi^2_{ij,f,m}}, \sum_{t=1}^{T}(\lambda^{(m)}_{ij,f,m,t} - \lambda^{(m)}_{ij,f,m,t-1})^2 + (\lambda^{(m)}_{ij,f,m,0} - \lambda_{ij,f,m})^2\right),$$

and $\lambda^{new}_{ij,f,m}$ from the Gaussian posterior

$$\lambda_{ij,f,m}| \dots \sim \mathcal{N}\left(\frac{\lambda_{ij,f,0}\tau^2_{ij,f,m}}{\tau^2_{ij,f,m} + \theta_{ij,f,m}}, \frac{\tau^2_{ij,f,m}\theta_{ij,f,m}}{\tau^2_{ij,f,m} + \theta_{ij,f,m}}\right).$$

B.1.4 Sample hyperparameters $\xi^2_{ij,f,m}$ and $\tau^2_{ij,f,m}$

Sample the prior variance $\xi^2_{ij,f,m}$ and $\tau^2_{ij,f,m}$ for $i = 1, \dots, K, j = 1, f, K + i - 1.$ from conditionally independent generalized inverse Gaussian distributions as follows:

$$\xi^2_{ij,f,m}|\sqrt{\theta}_{ij,f,m}, a^{\xi}_{i,f}, \kappa^2_{i,f} \sim \mathcal{GIG}\left(a^{\xi}_{i,f} - 1/2, a^{\xi}_{i,f}\kappa^2_{i,f}, \theta_{ij,f,m}\right), \tag{B.4}$$

$$\tau^2_{ij,f,m}|\lambda^2_{ij,f,m}, a^{\tau}_{i,f}, \nu^2_{i,f} \sim \mathcal{GIG}\left(a^{\tau}_{i,f} - 1/2, a^{\tau}_{i,f}\lambda^2_{i,f}, a^2_{ij,f,m}\right). \tag{B.5}$$

B.1.5 Sample the hyperparameters $\nu^2_{i,f}$ and $\kappa^2_{i,f}$

Sample the hyperparameters $\nu^2_{i,f}|a^\tau_{i,f}, e_1, e_2, \boldsymbol{\tau}^2_{i,f,m}$ and $\kappa^2_{i,f}|a^\xi_{i,f}, d_1, d_2, \boldsymbol{\xi}^2_{i,f,m}$ from the corresponding conditional posteriors:

$$\nu^2_{i,f}|a^\tau_{i,f}, e_{f,1}, e_{f,2}, \boldsymbol{\tau}^2_{i,f,m} \sim \mathcal{G}\left(e_{f,1} + a^\tau_{i,f}K, e_{f,2} + \frac{\bar{\tau}^2_{i,f}}{2} a^\tau_{i,f} K\right), \tag{B.6}$$

$$\kappa^2_{i,f}|a^\xi_{i,f}, d_{f,1}, d_{f,2}, \boldsymbol{\xi}^2_{i,f,m} \sim \mathcal{G}\left(d_{f,1} + a^\xi_{i,f}K, d_{f,2} + \frac{\bar{\xi}^2_{i,f}}{2} a^\xi_{i,f} K\right), \tag{B.7}$$

where $\bar{\tau}^2_{i,f}$ and $\bar{\xi}^2_{i,f}$ are the averages of the variances in the shrinkage priors:

$$\bar{\tau}^2_{i,f} = \frac{1}{(K+i-1)P} \sum_{j=1}^{K+i-1} \sum_{m=1}^{P} \tau^2_{ij,f,m}, \quad \bar{\xi}^2_{i,f} = \frac{1}{(K+i-1)P} \sum_{j=1}^{K+1-i} \sum_{m=1}^{P} \xi^2_{ij,f,m}.$$

B.1.6 Sample hyperparameter $a^\tau_{i,f}$

Sample from $a^\tau_{i,f}|\boldsymbol{\lambda}_{i,f,1}, \ldots, \boldsymbol{\lambda}_{i,f,P}, \nu^2_{i,f}$ using a random walk Metropolis Hastings (MH) step based on proposing $\log a^{\tau,\text{new}}_{i,f} \sim \mathcal{N}(\log a^\tau_{i,f}, c^2_\tau)$. The accept probability for $a^{\tau,\text{new}}_{i,f}$ is:

$$\min\left\{1, \frac{p(a^{\tau,\text{new}}_{i,f})a^{\tau,\text{new}}_{i,f}}{p(a^\tau_{i,f})a^\tau_{i,f}} \prod_{j=1}^{(K+i-1)} \prod_{m=1}^{P} \frac{p(\lambda_{ij,f,m}|a^{\tau,\text{new}}_{i,f}, \nu^2_{i,f})}{p(\lambda_{ij,f,m}|a^\tau_{i,f}, \nu^2_{i,f})}\right\}.$$

B.1.7 Sample hyperparameter $a^\xi_{i,f}$

Sample from $a^\xi_{i,f}|\sqrt{\boldsymbol{\Theta}}_{i,f,1}, \ldots, \sqrt{\boldsymbol{\Theta}}_{i,f,P}, \kappa^2_{i,f}$ using a random walk Metropolis Hastings (MH) step based on proposing $\log a^{\xi,\text{new}}_{i,f} \sim \mathcal{N}(\log a^\xi_{i,f}, c^2_\xi)$. The accept probability for $a^{\xi,\text{new}}_{i,f}$ is:

$$\min\left\{1, \frac{p(a^{\xi,\text{new}}_{i,f})a^{\xi,\text{new}}_{i,f}}{p(a^\xi_{i,f})a^\xi_{i,f}} \prod_{j=1}^{(K+i-1)} \prod_{m=1}^{P} \frac{p(\sqrt{\theta}_{ij,f,m}|a^{\xi,\text{new}}_{i,f}, \kappa^2_{i,f})}{p(\sqrt{\theta}_{ij,f,m}|a^\xi_{i,f}, \kappa^2_{i,f})}\right\}.$$

B.2 VB Algorithm Details

B.2.1 VB Forward Filtering Backward Smoothing

The non-centered TVP model is similar as (B.1) and (B.2). However, the transformed response becomes $\bar{f}_{it}^{(m-1)} = f_{it}^{(m-1)} - \langle \boldsymbol{\lambda}_{i,f,m} \rangle^{\{l\}} \boldsymbol{x}_{f,t}^{(m-1)}$, and the regressor is $\boldsymbol{F}_{i,f,m,t} = \left\langle \sqrt{\boldsymbol{\Theta}}_{i,f,m} \right\rangle^{\{l\}} \boldsymbol{x}_{f,t}^{(m-1)}$. For $t = 1, \ldots, T$, we can update $r^{\{l+1\}}(\tilde{\boldsymbol{\lambda}}_{f,m,t}) \approx \mathcal{T}_{n_{i,f,m,t}}(\boldsymbol{m}_{i,f,m,t}, \boldsymbol{C}_{i,f,m,t})$.

- Forward filter computation is similar as forward filter illustrated in Section B.1.1.
    - Posterior at $t-1$: $p(\tilde{\boldsymbol{\lambda}}_{i,f,m,t-1} | D_{i,f,m,t-1})$ is approximately distributed as

      $$\mathcal{T}_{n_{i,f,m,t-1}}(\boldsymbol{m}_{i,f,m,t-1}, \boldsymbol{C}_{i,f,m,t-1}),$$

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

      $$E(\bar{f}_{it}^{(m-1)} | D_{i,f,m,t-1}) \approx \boldsymbol{F}_{i,f,m,t}^T \boldsymbol{m}_{i,f,m,t-1},$$

      and

      $$V(\bar{f}_{it}^{(m-1)} | D_{i,f,m,t-1}) \approx Q_{i,f,m,t} = \boldsymbol{F}_{i,f,m,t}^T \boldsymbol{R}_{i,f,m,t} \boldsymbol{F}_{i,f,m,t} + s_{i,f,m,t-1},$$

      where $\boldsymbol{R}_{i,f,m,t} = \boldsymbol{C}_{i,f,m,t-1} + \boldsymbol{I}_K$, and let $e_{i,f,m,t} = \bar{f}_{it}^{(m-1)} - \boldsymbol{F}_{i,f,m,t}^T \boldsymbol{m}_{i,f,m,t-1}$.

    - Apply Bayes' theorem and the above equations, we can obtain the approximate posterior filtered distribution at time t as

      $$\tilde{\boldsymbol{\lambda}}_{i,f,m,t} | D_{i,f,m,t} \approx \mathcal{T}_{n_{i,f,m,t}}(\boldsymbol{m}_{i,f,m,t}, \boldsymbol{C}_{i,f,m,t}),$$

where

$$
\begin{aligned}
n_{i,f,m,t} &= n_{i,f,m,t-1} + 1, \\
s_{i,f,m,t} &= s_{i,f,m,t-1} + \frac{s_{i,f,m,t-1}}{n_{i,f,m,t}} \left(\frac{e^2_{i,f,m,t}}{Q_{i,f,m,t}} - 1 \right), \\
\boldsymbol{U}_{i,f,m,t} &= \boldsymbol{R}_{i,f,m,t} \boldsymbol{F}_{i,f,m,t} Q^{-1}_{i,f,m,t}, \\
\boldsymbol{m}_{i,f,m,t} &= \boldsymbol{m}_{i,f,m,t-1} + \boldsymbol{U}_{i,f,m,t} e_{i,f,m,t}, \\
\boldsymbol{C}_{i,f,m,t} &= \frac{s_{i,f,m,t}}{s_{i,f,m,t-1}} \left(\boldsymbol{R}_{i,f,m,t} - \boldsymbol{U}_{i,f,m,t} \boldsymbol{U}'_{i,f,m,t} Q_{i,f,m,t} \right).
\end{aligned}
$$

- Backward Smoother for $i = 1, \ldots, K$:
  - For $t = T$, $r^{\{l+1\}}(\tilde{\boldsymbol{\lambda}}_{i,f,m,t}) = \mathcal{T}_{n_{i,f,m,T}}(\boldsymbol{m}_{i,f,m,T}, \boldsymbol{C}_{i,f,m,T})$.
  - For $t = T-1, \ldots, 1$,

$$
r^{\{l+1\}}(\tilde{\boldsymbol{\lambda}}_{i,f,m,t}) = \mathcal{T}_{n_{i,f,m,T}}(\boldsymbol{m}_{i,f,m,t|T}, \frac{s_{i,f,m,T}}{s_{i,f,m,t}} \boldsymbol{C}_{i,f,m,t|T}),
$$

where

$$
\begin{aligned}
\boldsymbol{m}_{i,f,m,t|T} &= \boldsymbol{m}_{i,f,m,t} + \boldsymbol{C}_{i,f,m,t} \boldsymbol{R}^{-1}_{i,f,m,t+1} (\boldsymbol{m}_{i,f,m,t+1|T} - \boldsymbol{m}_{i,f,m,t}), \\
\boldsymbol{C}_{i,f,m,t|T} &= \boldsymbol{C}_{i,f,m,t} + \boldsymbol{C}_{i,f,m,t} \boldsymbol{R}^{-1}_{i,f,m,t+1} \left(\boldsymbol{C}_{i,f,m,t+1|T} - \boldsymbol{R}_{i,f,m,t+1} \right) \boldsymbol{R}^{-1}_{i,f,m,t+1} \boldsymbol{C}_{i,f,m,t}.
\end{aligned}
$$

B.2.2 Update the fixed regressor $r^{\{l+1\}}(\lambda_{ij,f,m})$

For $i = 1, \ldots, K$ and $j = 1, \ldots, K + i - 1$, update the fixed latent regressor $r^{\{l+1\}}(\lambda_{ij,f,m}) = \mathcal{N}(\mu^{\{l+1\}}_{\lambda_{ij,f,m}}, \Sigma^{\{l+1\}}_{\lambda_{ij,f,m}})$, where

$$
\begin{aligned}
\Sigma^{\{l+1\}}_{\lambda_{ij,f,m}} &= \left(\left\langle \frac{1}{\sigma^2_{i,f,m}} \right\rangle^{\{l+1\}} \sum_{t=1}^{T} x^{(m-1)}_{j,f,t} + \left\langle \frac{1}{\tau^2_{ij,f,m}} \right\rangle^{\{l\}} \right)^{-1}, \\
\mu^{\{l+1\}}_{\lambda_{ij,f,m}} &= \Sigma^{\{l+1\}}_{\lambda_{ij,f,m}} \cdot \left(\left\langle \frac{1}{\sigma^2_{i,f,m}} \right\rangle^{\{l+1\}} \sum_{t=1}^{T} \left(< B_t > x^{(m-1)}_{j,f,t} \right) \right), \\
< B_t > &= f^{(m)}_{j,t} - \left\langle \tilde{\boldsymbol{\lambda}}_{i,f,m,t} \right\rangle^{\{l+1\}} \mathrm{diag}\left(\left\langle \sqrt{\boldsymbol{\Theta}}_{i,f,m} \right\rangle^{\{l\}} \right) \boldsymbol{x}^{(m-1)}_t - \sum_{k \neq j} x^{(m-1)}_{k,f,t} < \lambda_{ik,f,m} >^{\{l\}},
\end{aligned}
$$

B.2.3 Update the variance of process $r^{\{l+1\}}(\sqrt{\theta}_{ij,f,m})$

For $i = 1, \ldots, K$, and $j = 1, \ldots, K + i - 1$, update the variance of process $r^{\{l+1\}}(\sqrt{\theta}_{ij,f,m}) = \mathcal{N}(\mu^{\{l+1\}}_{\sqrt{\theta}_{ij,f,m}}, \Sigma^{\{l+1\}}_{\sqrt{\theta}_{ij,f,m}})$, where

$$
\begin{aligned}
\Sigma^{\{l+1\}}_{\sqrt{\theta}_{ij,f,m}} &= \left(\left\langle \frac{1}{\sigma^2_{i,f,m}} \right\rangle^{\{l+1\}} \sum_{t=1}^{T} x^{(m-1)}_{j,f,t} \left\langle \tilde{\lambda}^2_{ij,f,m,t} \right\rangle^{\{l+1\}} + \left\langle \frac{1}{\xi^2_{ij,f,m}} \right\rangle^{\{l\}} \right)^{-1}, \\
\mu^{\{l+1\}}_{\sqrt{\theta}_{ij,f,m}} &= \Sigma^{\{l+1\}}_{\sqrt{\theta}_{ij,f,m}} \left(\sum_{t=1}^{T} \left(\left\langle \tilde{\lambda}_{ij,f,m,t} \right\rangle^{\{l+1\}} x^{(m-1)}_{j,f,t} + < C_t > \right) \right), \\
< C_t > &= f^{(m-1)}_{jt} - \left\langle \boldsymbol{\lambda}_{i,f,m} \right\rangle^{\{l+1\}} \boldsymbol{x}^{(m-1)}_{f,t} - \sum_{k \neq j} x^{(m-1)}_{k,t} \left\langle \tilde{\lambda}_{ik,f,m,t} \right\rangle^{\{l+1\}} \left\langle \sqrt{\theta}_{ik,f,m} \right\rangle^{\{l\}}.
\end{aligned}
$$

B.2.4 Update hyperparameter $r^{\{l+1\}}(a^{\tau}_{i,f})$

For $i = 1, \ldots, K$, the variational distribution $r^{\{l+1\}}(a^{\tau}_{i,f})$ can be computed up to a proportionality constant,

$$
r^{\{l+1\}}(a^{\tau}_{i,f}) \propto \exp\left\{ (K + i - 1) P \left(A \times a^{\tau}_{i,f} - \log \Gamma(a^{\tau}_{i,f}) \right) \right\},
$$

where

$$A = \log(a^{\tau}_{i,f}) + \left\langle \log \nu^2_{i,f} \right\rangle^{\{l+1\}} - \log 2 + < \overline{\log \tau^2_{i,f}} >^{\{l+1\}}$$
$$- \frac{1}{2} \left\langle \nu^2_{i,f} \right\rangle^{\{l+1\}} < \bar{\tau}^2_{i,f} >^{\{l+1\}} - \frac{b^{\tau}}{(K+i-1)P},$$
$$< \overline{\log \tau^2_{i,f}} >^{\{l+1\}} = \frac{1}{(K+i-1)P} \sum_{j=1}^{K+i-1} \sum_{m=1}^{P} \left\langle \log \tau^2_{ij,f,m} \right\rangle^{\{l+1\}}.$$

We can update $r^{\{l+1\}}(a^{\tau}_{i,f})$ with importance sampling as follows:

- For n in $1, \ldots, N$, sample $a^{\tau,n}_{i,f}$ from proposal distribution $q(\cdot)$.
- Compute the weights
$$w(a^{\tau,n}_{i,f}) = \frac{r^{\{l+1\}}(a^{\tau,n}_{i,f})}{q(a^{\tau,n}_{i,f})}.$$
- The variational distribution $r^{\{l+1\}}(a^{\tau}_{i,f})$ can be approximated by
$$r^{\{l+1\}}(a^{\tau}_{i,f}) \approx \frac{\sum_{n=1}^{N} w(a^{\tau,n}_{i,f}) \delta_{a^{\tau}_{i,f}}(a^{\tau}_{i,f})}{\sum_{n=1}^{N} w(a^{\tau,n}_{i,f})}.$$
- For any function $h(a^{\tau}_{i,f})$,
$$E[h(a^{\tau}_{i,f})] = \frac{\sum_{n=1}^{N} w(a^{\tau,n}_{i,f}) h(a^{\tau,n}_{i,f})}{\sum_{n=1}^{N} w(a^{\tau,n}_{i,f})}.$$

B.2.5 Update hyperparameter $r^{\{l+1\}}(a^{\xi}_{i,f})$

For $i = 1, \ldots, K$, the variational distribution $r^{\{l+1\}}(a^{\xi}_{i,f})$ can be computed up to a proportionality constant,

$$r^{\{l+1\}}(a^{\xi}_{i,f}) \propto \exp\left\{(K+i-1)P \left(A \times a^{\xi}_{i,f} - \log \Gamma(a^{\xi}_{i,f})\right)\right\},$$

where

$$\begin{aligned} A = \log(a^{\xi}_{i,f}) + \left\langle \log \kappa^2_{i,f} \right\rangle^{\{l+1\}} - \log 2 + < \overline{\log \xi^2_{i,f}} >^{\{l+1\}} \\ - \frac{1}{2} \left\langle \kappa^2_{i,f} \right\rangle^{\{l+1\}} < \bar{\xi}^2_{i,f} >^{\{l+1\}} - \frac{b^{\xi}}{(K+i-1)P}, \\ < \overline{\log \xi^2_{i,f}} >^{\{l+1\}} = \frac{1}{(K+i-1)P} \sum_{j=1}^{K+i-1} \sum_{m=1}^{P} \left\langle \log \xi^2_{ij,f,m} \right\rangle^{\{l+1\}}. \end{aligned}$$

We can update $r^{\{l+1\}}(a^{\xi}_{i,f})$ with importance sampling. The scheme of importance sampling which is similar as updating $a^{\tau}_{i,f}$ in the previous section.

B.3 Simulation

B.3.1 Simulation (i)

In this section, Figure B.1 shows the plots of true and estimated AR coefficients after lag 2. Figure B.2 shows posterior mean of the log spectral density and squared coherence estimated by Normal Gamma TV-VPARCOR model with variational inference algorithm.

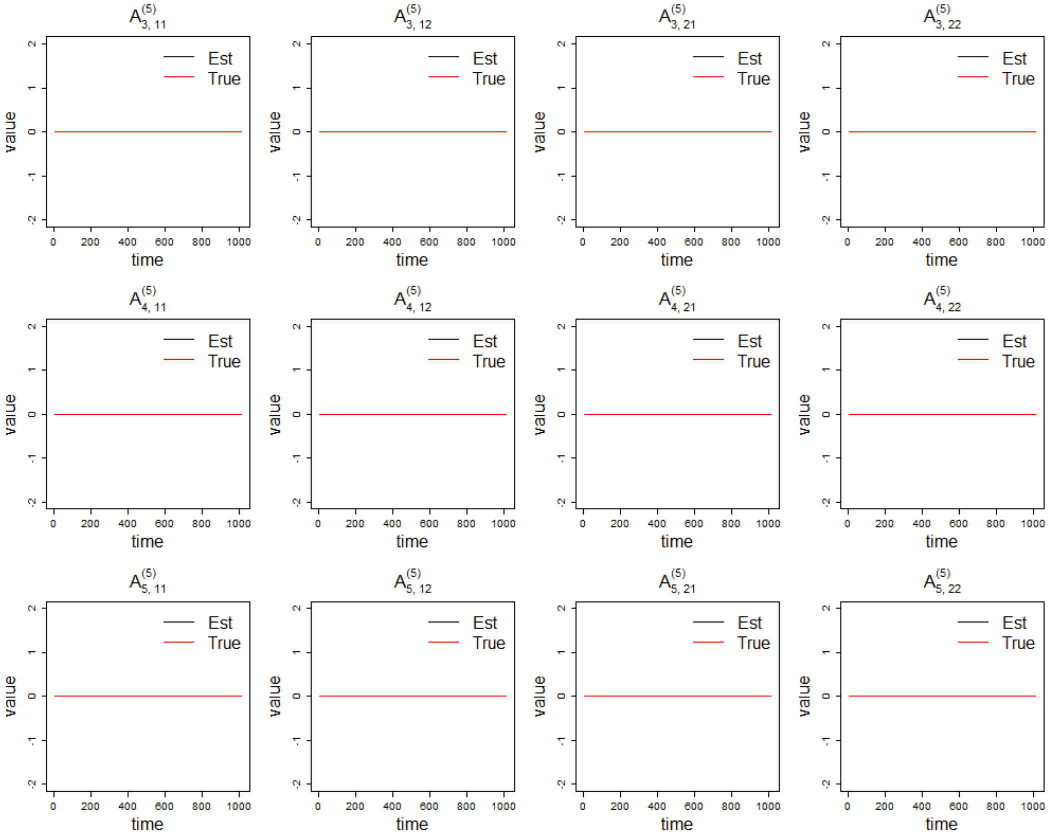

Figure B.1: Simulation scenario (i): The plots of AR coefficients after lag 2. The red lines are the true AR coefficients. The black lines are the estimated AR coefficients.

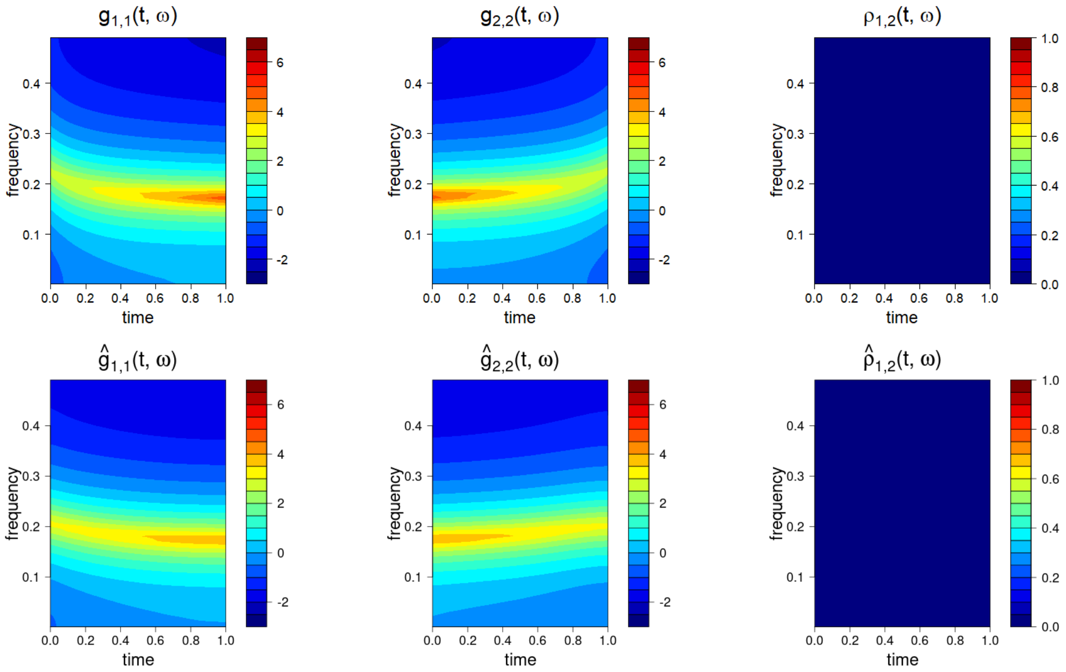

Figure B.2: Simulation scenario (i): Top row: True log spectral density $g_{11}(t, \omega)$ (left); true log spectral density $g_{22}(t, \omega)$ (center); true squared coherence (right). Bottom row: Corresponding estimated log spectral densities (left and center) and estimated squared coherence.

B.3.2 Simulation (ii)

In this section, Figure B.3 shows the plots of true and estimated AR coefficients after lag 2. Figure B.4 shows posterior mean of the log spectral density and squared coherence estimated by Normal Gamma TV-VPARCOR model with variational inference algorithm.

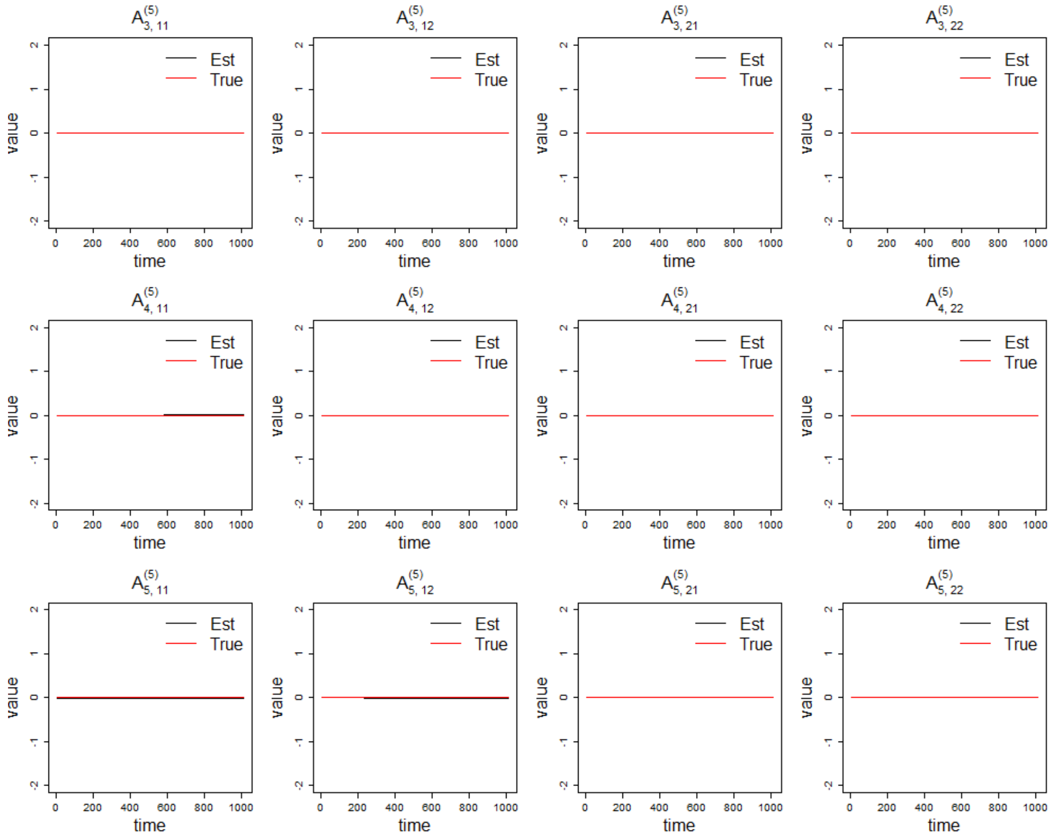

Figure B.3: Simulation scenario (ii): The plots of AR coefficients after lag 2. The red lines are the true AR coefficients. The black lines are the estimated AR coefficients.

B.3.3 Simulation (iii)

In this section, Figure B.5 shows the plots of true and estimated AR coefficients after lag 2. Figure B.6 shows posterior mean of the log spectral density

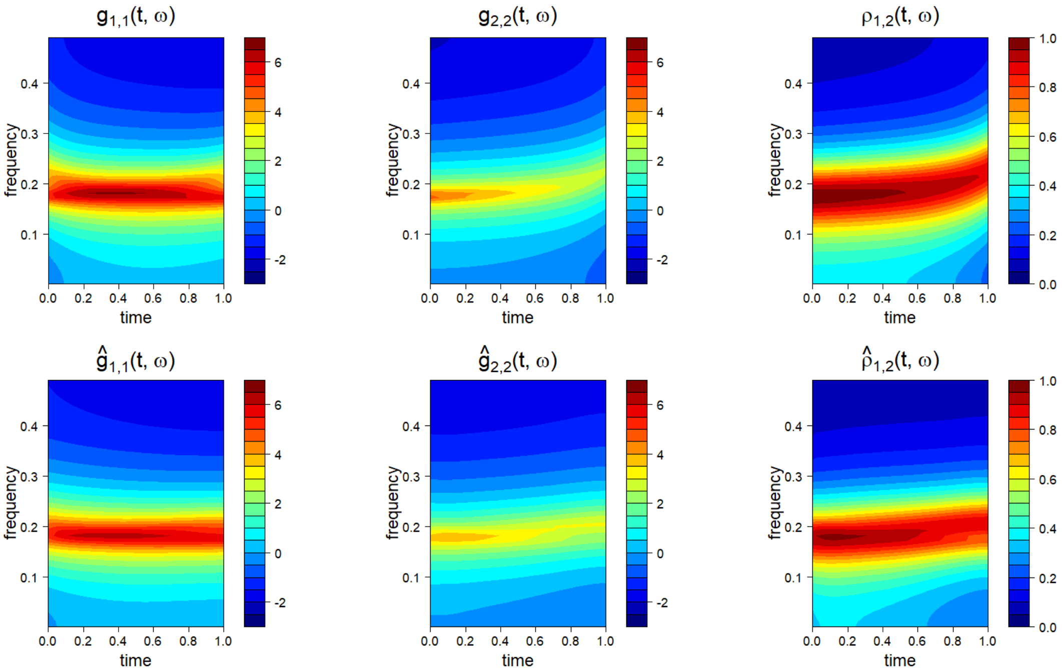

Figure B.4: Simulation scenario (ii): Top row: True log spectral density $g_{11}(t, \omega)$ (left); true log spectral density $g_{22}(t, \omega)$ (center); true squared coherence (right). Bottom row: Corresponding estimated log spectral densities (left and center) and estimated squared coherence.

and squared coherence estimated by Normal Gamma TV-VPARCOR model with variational inference algorithm.

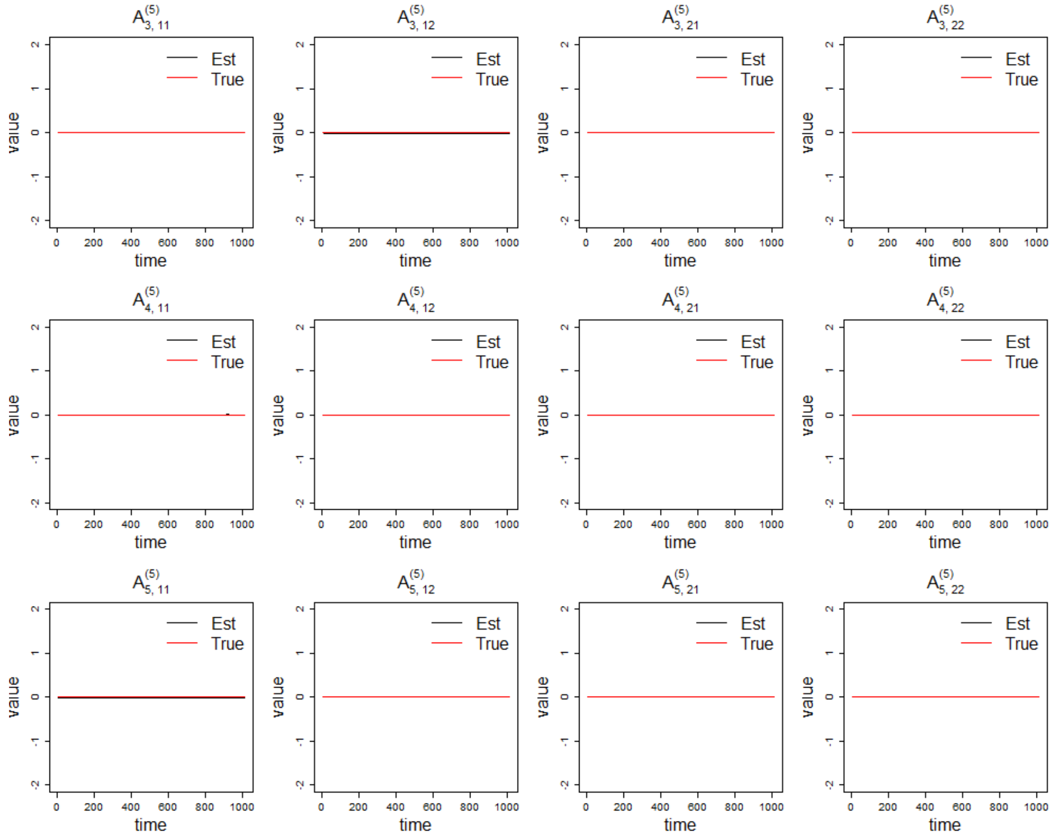

Figure B.5: Simulation scenario (iii): The plots of AR coefficients after lag 2. The red line is the true AR coefficients. The black line is the estimated AR coefficients.

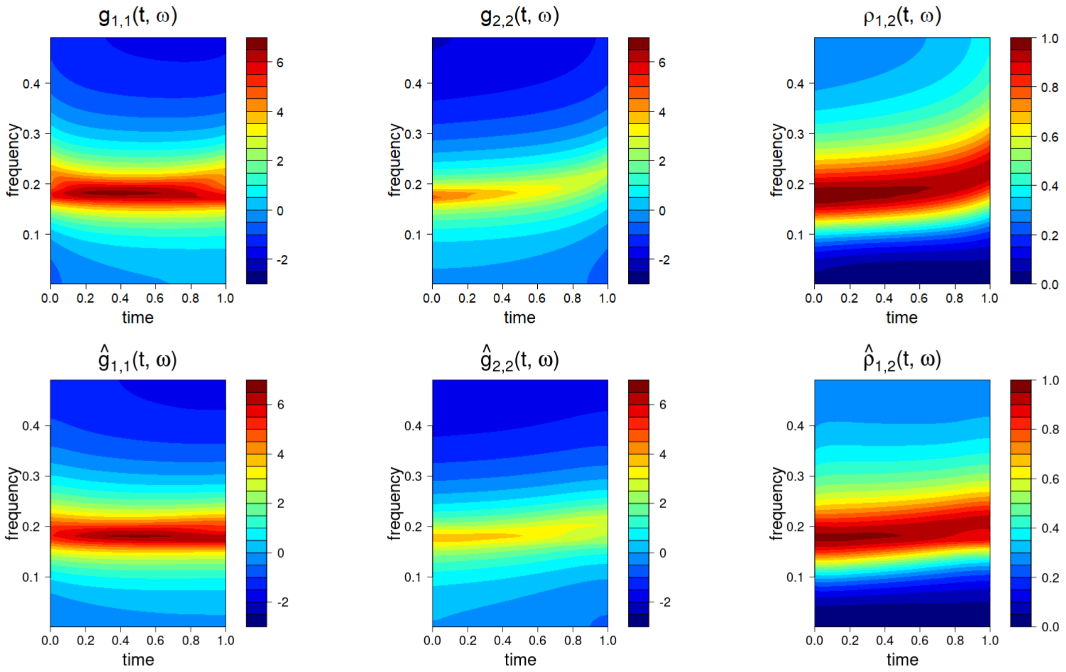

Figure B.6: Simulation scenario (iii): Top row: True log spectral density $g_{11}(t, \omega)$ (left); true log spectral density $g_{22}(t, \omega)$ (center); true squared coherence (right). Bottom row: Corresponding estimated log spectral densities (left and center) and estimated squared coherence.

Bibliography

J. R. Ameen and P. J. Harrison. Normal discount Bayesian models. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting, September 6/10, 1983*, volume 2, page 271. North Holland, 1985.

L. Astolfi, F. Cincotti, D. Mattia, F. De Vico Fallani, A. Tocci, A. Colosimo, S. Salinari, M.G. Marciani, W. Hesse, H. Witte, M. Ursino, M. Zavaglia, and F. Babiloni. Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators. *IEEE Transactions on Biomedical Engineering*, 55 (3):902–913, 2008.

L. A. Baccalá and K. Sameshima. Partial directed coherence: A new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474, 2001.

M. J. Beal. *Variational algorithms for approximate Bayesian inference.* University of London, University College London (United Kingdom), 2003.

M. A. G. Belmonte, G. Koop, and D. Korobilis. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94, 2014.

A. Bhattacharya, D. Pati, N. S. Pillai, and D. B. Dunson. Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490, 2015.

A. Bitto and S. Frühwirth-Schnatter. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97, 2019.

K. J. Blinowska. Review of the methods of determination of directed connectivity from multichannel data. *Medical & biological engineering & computing*, 49(5): 521–529, 2011.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer, New York, second edition, 1991.

S. A. Bruce, M. H. Hall, D. J. Buysse, and R. T. Krafty. Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series. *Biometrics*, 74:260–269, 2018.

A. Cadonna, A. Kottas, and R. Prado. Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, 2019.

C. K. Carter and R.Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

B. L. P. Cheung, B. A. Riedner, G. Tononi, and B. D. Van Veen. Estimation of cortical connectivity from EEG using state-space models. *IEEE Transactions on Biomedical Engineering*, 57(9):2122–2134, 2010.

S. Chiang, M. Guindani, H. J. Yeh, Z. Haneef, J. M. Stern, and M. Vannucci. Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping*, 38: 1311–1332, 2017.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.

S. Frühwirth-Schnatter and H. Wagner. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100, 2010.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

D. Gamerman and H. S. Migon. Dynamic hierarchical models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 629–642, 1993.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC press Boca Raton, FL, third edition, 2014.

C. Gorrostieta, M. Fiecas, H. Ombao, E. Burke, and S. Cramer. Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Frontiers in Computational Neuroscience*, 7:159, 2013.

J.E. Griffin and P.J. Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

M. H. Hayes. *Statistical digital signal processing and modeling*. John Wiley & Sons, 1996.

L. Hu, M. Guindani, N. Fortin, and H. Ombao. A hierarchical Bayesian model for differential connectivity in multi-trial brain signals. *Econometrics and Statistics*, 15:117–135, 2020.

H. Ishwaran and J. S. Rao. Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, 33(2):730–773, 2005.

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87, 2002.

G. Kastner and S. Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423, 2014.

G. Kitagawa and W. Gersch. *Smoothness Priors Analysis of Time Series*. Ima Volumes in Mathematics and Its Applications. Springer New York, 1996.

R. T. Krafty, O. Rosen, D. S. Stoffer, D. J. Buysse, and M. H. Hall. Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association*, 112:1405–1416, 2017.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

R. Kuś, M. Kamiński, and K. J. Blinowska. Determination of EEG activity propagation: Pair-wise versus multichannel estimate. *IEEE Transactions of Biomedical Engineering*, 51(9):1501–1510, 2004.

F. Landim and D. Gamerman. Dynamic hierarchical models: an extension to matrix-variate observations. *Computational statistics & data analysis*, 35(1): 11–42, 2000.

Z. Li and R. T. Krafty. Adaptive Bayesian time-frequency analysis of multivariate time series. *Journal of the American Statistical Association*, 2018. doi: 10.1080/01621459.2017.1415908.

H. F. Lopes, R. E. McCulloch, and R.S. Tsay. Parsimony inducing priors for large scale state-space models. *Bayesian Anal*, 2016.

T. Milde, L. Leistritz, L. Astolfi, W. H. R. Miltner, T. Weiss, F. Babiloni, and H. Witte. 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, 2009.

R. B. Millar and S. McKechnie. A one-step-ahead pseudo-DIC for comparison of bayesian state-space models. *Biometrics*, 70(4):972–980, 2014.

J. Nakajima and M. West. Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164, 2013.

J. Nakajima and M. West. Dynamic network signal processing using latent threshold models. *Digital Signal Processing*, 47:6–15, 2015.

J. Nakajima and M. West. Dynamics and sparsity in latent threshold factor models: A study in multivariate EEG signal processing. *Brazilian Journal of Probability and Statistics*, 31:701–731, 2017.

Automated Surface Observing System (ASOS) User's Guide. NOAA, DoD, FAA, USNavy, 1998.

H. Ombao, J. A. Raz, R. von Sachs, and B. A. Malow. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96(454):543–560, 2001.

A. Omidvarnia, G. Azemi, B. Boashash, J. M. O'Toole, P. B. Colditz, and S. Vanhatalo. Measuring time-varying information flow in scalp EEG signals: Orthogonalized partial directed coherence. *IEEE Transactions on Biomedical Engineering*, 61(3):680–693, 2014.

D. Ostwald, E. Kirilina, L. Starke, and F. Blankenburg. A tutorial on variational bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology*, 60:1–19, 2014.

T. Park and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

S. M. Peterson and D. P. Ferris. Group-level cortical and muscular connectivity during perturbations to walking and standing balance. *NeuroImage*, 198:93–103, 2019.

R. Prado and M. West. Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions. In *Modelling longitudinal and spatially correlated data*, pages 349–361. Springer, 1997.

R. Prado, M. West, and A. D. Krystal. Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(1): 95–109, 2001.

R. Prado, M. A. R. Ferreira, and M. West. *Time Series: Modeling, Computation & Inference*. Chapman & Hall/CRC Press, 2nd edition, 2021.

T. Proietti and A. Giovannelli. A Durbin–Levinson regularized estimator of high-dimensional autocovariance matrices. *Biometrika*, 105(4):783–795, 2018.

V. Rockova and K. McAlinn. Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*, 16(1):233–269, 2021.

C. Schmidt, B. Pester, N. Schmid-Hertel, H. Wittle, A. Wismuller, and L. Leistritz. A multivariate Granger causality concept towards full brain functional connectivity. *PLOS ONE*, 11:1–25, 2016.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples.* Springer texts in statistics. Springer, 2017. ISBN 9780387293172.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

C. Ting, H. Ombao, S. Balqis Samdin, and S. Hussain Salleh. Estimating dynamic connectivity using regime-switching factor models. *IEEE Trans. Med. Imaging*, 37:1011–1023, 2017.

K. Triantafyllopoulos. Covariance estimation for multivariate conditionally Gaussian dynamic linear models. *Journal of Forecasting*, 26(8):551–569, 2007.

R. Tsay. *Multivariate time series analysis: With R and financial applications.* Wiley Series in Probability and Statistics, 2013.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models.* Springer Series in Statistics. Springer New York, second edition, 1997.

M. West, R. Prado, and A. D. Krystal. Evaluation and comparison of eeg traces: Latent structure in nonstationary time series. *Journal of the American Statistical Association*, 94(446):375–387, 1999.

W. H. Yang, S. H. Holan, and C. K. Wikle. Bayesian lattice filters for time-varying autoregression and time–frequency analysis. *Bayesian Analysis*, 11(4): 977–1003, 2016.

Y. Yu and X. Meng. To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

Z. Yu, R. Prado, E. Burke Quinlan, S. C. Cramer, and H. Ombao. Understanding the impact of stroke on brain motor function: A hierarchical Bayesian approach. *Journal of the American Statistical Association*, 111(514):549–563, 2016.

Z. Zhang. *Multivariate time series analysis in climate and environmental research.* Springer, New York, 2017.

G. Zhou. Algorithms for estimation of possibly nonstationary vector time series. *Journal of Time Series Analysis*, 13(2):171–188, 1992.