Statistics and Its Interface Volume 16 (2023) 69–79

Hierarchical dynamic PARCOR models for analysis of multiple brain signals

Wenjie Zhao*† and Raquel Prado

We present an efficient hierarchical model for inferring latent structure underlying multiple non-stationary time series. The proposed model describes the time-varying behavior of multiple time series in the partial autocorrelation domain, which results in a lower dimensional representation, and consequently computationally faster inference, than those required by models in the time and/or frequency domains, such as time-varying autoregressive models, which are commonly used in practice. We illustrate the performance of the proposed hierarchical dynamic PARCOR models and 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.

Keywords and phrases: Multiple non-stationary time series, Partial autocorrelation, Hierarchical Bayesian models, Dynamic linear models.

1. INTRODUCTION

Neuroscience studies typically involve simultaneous recording of multiple non-stationary brain signals – e.g., multi-channel electroencephalogram (EEG) recordings or fMRI – on subjects undergoing a treatment or experimental condition. This often results in datasets that consist of multiple time series with a common underlying structure that can be described via hierarchical models. Typically, interest lies in inferring the time-frequency characteristics of the individual time series, as well as those of their latent common structure. For instance, in some 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. Instead, 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 such signals (Section 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 4.2 illustrates this with data from a clinical EEG study).

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 (TVVAR) – have been used to infer latent structure from multiple brain signals [e.g., 16, 18, 13, 12, 5, 6]. 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. Approaches involving multivariate models are often used to infer features related to the behavior of individual time series, as well as those that provide information about relationships across multiple time series such as coherence, partial coherence and/or partial directed coherence [e.g., 5, 6]. Other modeling frameworks such as those based on factor models, focus on discovering the latent structure underlying multiple time series [e.g., 18, 13]. Frequency domain and time-frequency domain hierarchical models are also available to analyze multiple time series [e.g., 10, 3, among others].

The sophisticated modeling approaches described 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) for inference, which limits their use in practical settings that involve simultaneous modeling of a large number of time series. Here we extend the univariate dynamic partial autocorrelation (PARCOR) model of [24] 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

*Corresponding author.
†ORCID: 0000-0002-5102-8703.

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.

The paper is structured as follows. Section 2 discusses the hierarchical PARCOR representation and its connection with TVAR models, as well as the algorithm for posterior inference and tools for model selection. Section 3 presents a simulation study. Section 4 illustrates the performance of the proposed model for analysis of multiple brain signals recorded in two EEG studies. Finally, Section 5 presents a summary and discusses some future directions.

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 time-varying autoregressive (TVAR) model of order P [see e.g., 17] 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.

Yang et al. (2016) [24] develops a Bayesian dynamic linear model (DLM) approach for computationally efficient modeling of univariate non-stationary time series that takes advantage of the partial autocorrelation (PARCOR) representation of the TVAR model. Such approach is based on a lattice structure formulation of the univariate Durbin-Levinson algorithm [see, e.g., 2, 20] used in [8]. More recently, [25] proposed a multivariate version of the model developed in [24]. Here we extend the approach of [24] by adding hierarchical latent structure to the dynamic PARCOR model using the hierarchical dynamic linear structure of [4]. This allows us to jointly model multiple time series (instead of a single time series) to infer the time-frequency characteristics of each time series and also to extract their common underlying structure. We note that unlike the multivariate dynamic PARCOR model of [25], the hierarchical model proposed here is not meant to infer the relationships across multiple time series, but instead focuses on inferring their latent shared characteristics via the hierarchical structure.

One of the main advantages of the PARCOR representation of TVARs and TV-VARs is that it results in lower dimensional models, making posterior inference in the PARCOR domain much more computationally efficient. In the univariate case, the DLM representation of a TVAR model of order P leads to a state-space vector of dimension P, therefore, inference requires inverting matrices of dimension $P \times P$ at each time t for posterior filtering and smoothing. Instead, the PARCOR representation of a TVAR(P) [24] results in $2P$ DLMs (P for the forward and P for the backward PARCOR coefficients), each with a univariate state parameter and so, no matrix inversions are needed for the posterior filtering and smoothing algorithms. This leads to substantial computational gains in the case of univariate and multivariate PARCOR models as shown in [24] and [25]. Similarly, 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,

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

$$b_{it}^{(P)} = x_{it} - \sum_{j=1}^{P} d_{i,j,t}^{(P)} x_{i,t+j}. \tag{2}$$

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

$$f_{it}^{(m-1)} = \alpha_{i,m,t}^{(m)} b_{i,t-m}^{(m-1)} + f_{it}^{(m)}, \quad f_{it}^{(m-1)} \sim \mathcal{N}(0, \sigma_{f,i,m,t}^2), \tag{3}$$

$$b_{it}^{(m-1)} = \beta_{i,m,t}^{(m)} f_{i,t+m}^{(m-1)} + b_{it}^{(m)}, \quad b_{it}^{(m-1)} \sim \mathcal{N}(0, \sigma_{b,i,m,t}^2), \tag{4}$$

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 $m = 1, \ldots, P$. 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-Levison algorithm [see, e.g., Brockwell and Davis, 1991 2] as follows

$$a_{i,j,t}^{(m)} = a_{i,j,t}^{(m-1)} - a_{i,m,t}^{(m)} d_{i,m-j,t}^{(m-1)}, \tag{5}$$

$$d_{i,j,t}^{(m)} = d_{i,j,t}^{(m-1)} - d_{i,m,t}^{(m)} a_{i,m-j,t}^{(m-1)}, \quad j = 1, \ldots, m-1, \tag{6}$$

with $a_{i,m,t}^{(m)} = \alpha_{i,m,t}^{(m)}$ and $d_{i,m,t}^{(m)} = \beta_{i,m,t}^{(m)}$, for $m = 1, \ldots, P$.

2.1 Model specification and inference

The first level of our proposed hierarchical model specification uses (3) and (4) as observational level equations of univariate DLMs [22, 17] on the forward and backward PARCOR time-varying coefficients.

The next level of the hierarchical model requires specifying the structural equations [4]. 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

$$\alpha_{i,m,t}^{(m)} = \mu_{f,m,t}^{(m)} + \gamma_{f,i,m,t}^{(m)} + \nu_{f,i,m,t}, \tag{7}$$

$$\beta_{i,m,t}^{(m)} = \mu_{b,m,t}^{(m)} + \gamma_{b,i,m,t}^{(m)} + \nu_{b,i,m,t}, \tag{8}$$

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,

$$\mu_{f,m,t}^{(m)} = \mu_{f,m,t-1}^{(m)} + w_{\mu,f,m,t}, \tag{9}$$

$$\mu_{b,m,t}^{(m)} = \mu_{b,m,t-1}^{(m)} + w_{\mu,b,m,t}, \tag{10}$$

$$\gamma_{f,i,m,t}^{(m)} = \gamma_{f,i,m,t-1}^{(m)} + w_{\gamma,f,i,m,t}, \ i = 1, \ldots, n-1, \tag{11}$$

$$\gamma_{b,i,m,t}^{(m)} = \gamma_{b,i,m,t-1}^{(m)} + w_{\gamma,b,i,m,t}, \ i = 1, \ldots, n-1, \tag{12}$$

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 [22], however, this is not the case in hierarchical DLMs when the observational, structural and system variance are unknown. A conjugate model structure is available [4, 22], 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 (3) independently follow normal distributions with the observational innovation variance scaled by the parameter $\sigma_{f,m}^2$ at the different stage m. Similarly, the backward prediction errors of all time series (4) are assumed to independently follow normal distributions with the observational innovation variances scaled by $\sigma_{b,m}^2$. In other words we set $\sigma_{f,i,m,t}^2 = \sigma_{f,m}^2$ and $\sigma_{b,i,m,t}^2 = \sigma_{b,m}^2$ for each stage m and all t, and further assume conjugate prior distributions of $\sigma_{f,m}^2$ and $\sigma_{b,m}^2$ as follows:

$$\sigma_{f,m}^2 | \boldsymbol{D}_{f,m,0} \sim \mathcal{IG}\left(\frac{n_{f,0}}{2}, \frac{h_{f,0}}{2}\right), \tag{13}$$

$$\sigma_{b,m}^2 | \boldsymbol{D}_{b,m,0} \sim \mathcal{IG}\left(\frac{n_{b,0}}{2}, \frac{h_{b,0}}{2}\right). \tag{14}$$

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:

$$\boldsymbol{f}_t^{(m-1)} = \boldsymbol{F}_{1,f,m,t}^{(m-1)} \boldsymbol{\theta}_{1,f,m,t}^{(m)} + \boldsymbol{f}_t^{(m)}, \boldsymbol{f}_t^{(m)} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{f,m}^2 \boldsymbol{I}_n\right), \tag{15}$$

$$\boldsymbol{b}_t^{(m-1)} = \boldsymbol{F}_{1,b,m,t}^{(m-1)} \boldsymbol{\theta}_{1,b,m,t}^{(m)} + \boldsymbol{b}_t^{(m)}, \boldsymbol{b}_t^{(m)} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{b,m}^2 \boldsymbol{I}_n\right), \tag{16}$$

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

$$\boldsymbol{\theta}_{1,f,m,t}^{(m)} = \boldsymbol{F}_2 \boldsymbol{\theta}_{2,f,m,t}^{(m)} + \boldsymbol{\nu}_{f,m,t}^{(m)}, \boldsymbol{\nu}_{f,m,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{V}_{2,f,m,t}\right), \tag{17}$$

$$\boldsymbol{\theta}_{1,b,m,t}^{(m-1)} = \boldsymbol{F}_2 \boldsymbol{\theta}_{2,b,m,t}^{(m)} + \boldsymbol{\nu}_{b,m,t}^{(m)}, \boldsymbol{\nu}_{b,m,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{V}_{2,b,m,t}\right), \tag{18}$$

  where

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

  $\boldsymbol{\theta}_{2,\cdot,m,t}^{(m)} = (\mu_{\cdot,m,t}^{(m)}, \gamma_{\cdot,1,m,t}^{(m)}, \ldots, \gamma_{\cdot,n-1,m,t}^{(m)})'$, $\boldsymbol{\nu}_{\cdot,m,t}^{(m)} = (\nu_{\cdot,1,m,t}, \ldots, \nu_{\cdot,n,m,t})'$, $\boldsymbol{V}_{2,\cdot,m,t} = \sigma_{\cdot,m}^2 \boldsymbol{V}_{2,\cdot,m,t}^*$. 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 [22].

- System equations:

(19)
$$\boldsymbol{\theta}^{(m)}_{2,f,m,t} = \boldsymbol{\theta}^{(m)}_{2,f,m,t-1} + \boldsymbol{w}_{f,m,t}, \boldsymbol{w}_{f,m,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{W}_{f,m,t}\right),$$

(20)
$$\boldsymbol{\theta}^{(m)}_{2,b,m,t} = \boldsymbol{\theta}^{(m)}_{2,b,m,t-1} + \boldsymbol{w}_{b,m,t}, \boldsymbol{w}_{b,m,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{W}_{b,m,t}\right),$$

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 (13) and (14) on the innovation variances, we also assume

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

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,

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

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 log-likelihoods resulting from (15) and (16). Details about the selection of discount factors are discussed in Section 2.2.

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

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

where

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

- 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

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

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

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

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

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

$$\boldsymbol{V}^{\Delta}_{2,\cdot,m,t} = \boldsymbol{I}_n + \boldsymbol{F}_{1,\cdot,m,t}\boldsymbol{V}^{*}_{2,\cdot,m,t}\boldsymbol{F}'_{1,\cdot,m,t},$$

and initialized at $t = T$ with $\boldsymbol{m}^{T}_{k,\cdot,m,t} = \boldsymbol{m}_{k,\cdot,m,t}$ and $\boldsymbol{C}^{T}_{k,\cdot,m,t} = \boldsymbol{C}_{k,\cdot,m,t}$. 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 $\boldsymbol{f}^{(0)}_t = \boldsymbol{b}^{(0)}_t = (x_{1t},\ldots,x_{nt})'$, for $t = 1,\ldots,T$.
2. Use $\{\boldsymbol{f}^{(0)}_t\}$ and $\{\boldsymbol{b}^{(0)}_t\}$ as vectors of responses in the observational level equations (15) and (16), respectively, which, combined with structural equations (17) and (18), the random walk system equations (19) and (20) 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 $\{\hat{\boldsymbol{\theta}}^{(1)}_{k,f,1,t}\}$, $\{\hat{\boldsymbol{\theta}}^{(1)}_{k,b,1,t}\}$ 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 (15) and (16) to obtain the new series of forward and backward prediction errors, $\{\boldsymbol{f}^{(1)}_t\}$ and $\{\boldsymbol{b}^{(1)}_t\}$, for $t = 1,\ldots,T$.
4. Repeat steps 2-3 above until $\{\hat{\boldsymbol{\theta}}^{(m)}_{k,f,m,t}\}$, $\{\hat{\boldsymbol{\theta}}^{(m)}_{k,b,m,t}\}$, $\{\hat{\sigma}^2_{f,m}\}$ and $\{\hat{\sigma}^2_{b,m}\}$ have been obtained for all $m = 1,\ldots,P$.
5. Finally, use $\{\hat{\boldsymbol{\theta}}^{(m)}_{1,f,m,t}\}$ and $\{\hat{\boldsymbol{\theta}}^{(m)}_{1,b,m,t}\}$, for $m = 1,\ldots,P$, as well as equations (5) and (6) to obtain estimates of the forward and backward TVAR coefficients $\{\hat{a}^{(m)}_{i,j,t}\}$ and $\{\hat{d}^{(m)}_{i,j,t}\}$, for $i = 1,\ldots,n$, $j = 1,\ldots,m-1$, and $t = 1,\ldots,T$ via the Durbin-Levinson algorithm.
6. (Optional) Use the estimated $\hat{\mu}^{(m)}_{f,m,t}$ and $\hat{\mu}^{(m)}_{b,m,t}$, for $m = 1,\ldots,P$, $t = 1,\ldots,T$, and equations (5) and (6) to obtain the forward and backward baseline TVAR coefficients $\{\bar{a}^{(P)}_{j,t}\}$ and $\{\bar{d}^{(P)}_{j,t}\}$, for $j = 1,\ldots,P$, and $t = 1,\ldots,T$ using the Durbin-Levinson algorithm.

2.2 Model selection and time-frequency representation

[25] develops an approach to select discount factors by maximizing log-likelihood functions derived from the one-step ahead predictive density functions for the case of multivariate PARCOR models. We apply a similar idea in this hierarchical model. 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 (15) 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 (16). Using the selected optimal $\delta_{k,f,1}$ and $\delta_{k,b,1}$, we can obtain the corresponding series $\{\boldsymbol{f}^{(2)}_t\}$ and $\{\boldsymbol{b}^{(2)}_t\}$, 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 $\{\boldsymbol{f}^{(2)}_t\}$ and $\{\boldsymbol{b}^{(2)}_t\}$ 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 $\{\boldsymbol{f}^{(3)}_t\}$ and $\{\boldsymbol{b}^{(3)}_t\}$, 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. This method was used by [24] in the univariate context 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. Another 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, we use scree plots only as a visualization tool and use the model selection criterion below to find an optimal model order, as the methods that do not include a penalization for the number of model parameters do not work well in multivariate settings [25].

Method 2: DIC model selection criterion. We consider an approach based on the deviance information criterion (DIC) to choose the model order [see 21, 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 [11], we compute the one-step-ahead DIC, which uses a

pseudo deviance that conditions on the state at the previous time point, i.e.,

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

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 (15) for $t = 1, \ldots, T$, 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 $s = 1, \ldots, S$, 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 $\widehat{DIC}_m$.

2.3 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 time-frequency representation associated with a time-varying autoregression of order P can be obtained via

(21)
$$s_l(t,\omega) = \frac{\sigma^2_{f,P}}{|1-\sum_{j=1}^{P} a^{(P)}_{l,j,t}\exp(-2\pi i m\omega)|^2}, -1/2 \le \omega \le 1/2,$$

where $i = \sqrt{-1}$ [see, e.g. 9]. Using estimates $\hat{a}^{(P)}_{l,j,t}$ and $\hat{\sigma}^2_{f,P}$, we can obtain the estimated spectral density $\hat{s}_l(t,\omega)$. Note that we can also compute the underlying baseline spectral density, which we denote by $\bar{s}(t,\omega)$, using estimates $\bar{a}^{(P)}_{j,t}, j = 1, \ldots, P$ 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 into the corresponding spectral density using equation (21), 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.

3. SIMULATION STUDY

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

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

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 datasets. 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 $n_{f,0} = n_{b,0} = 1$, $h_{f,0} = h_{b,0} = 10$, $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{f,m,0} = (0,0,0,0,0)'$ and $\boldsymbol{C}^*_{f,m,0} = \boldsymbol{C}^*_{b,m,0} = 10\boldsymbol{I}_5$. The left column of Figure 2 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 1 shows the BLF-scree plots obtained from the hierarchical PARCOR approach for each of the 51 datasets for model orders $m = 1, \ldots, 5$. 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 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.

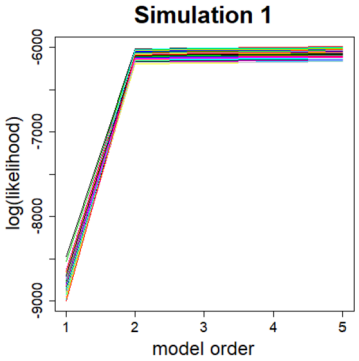

Figure 1. BLF-scree plots of the 51 realizations of simulation.

To illustrate the performance of the hierarchical TV-PARCOR model, we also fit TVAR(2) models to each univariate time series in each dataset to obtain some benchmark results. We computed the mean and standard deviations of the average squared error (ASE) for each of the models and each of the five time series averaging over the 51 datasets. The ASE for each time series l is defined as follows [14]

(22)
$$ASE_l = (TK)^{-1} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\log \hat{s}_{l,k}(t,\omega) - \log s_{l,k}(t,\omega)\right)^2,$$

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

Table 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)

Table 1 summarizes the mean and standard deivations of the ASE based on ASE_l for each simulated time series. Our proposed model outperforms TVAR models for estimating log spectral densities.

Figure 2 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 datasets. 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 3 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 4 shows the estimated 95% posterior interval of

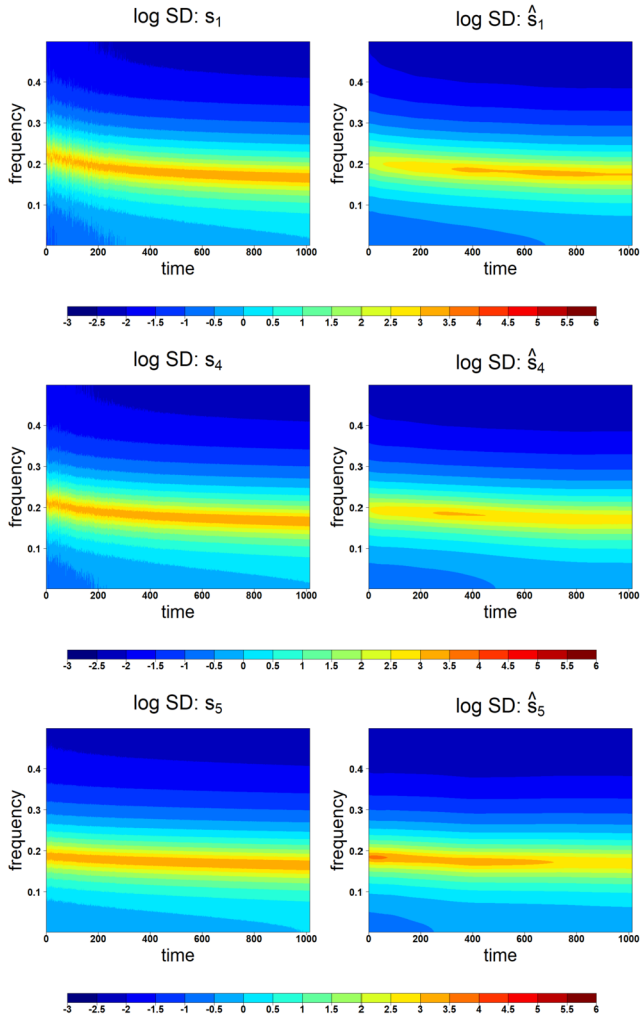

Figure 2. Left: True log-spectral densities $s_1(t,\omega)$ (top), $s_4(t,\omega)$ (middle) and $s_5(t,\omega)$ (bottom). Right: estimated log-spectral densities $\hat{s}_1(t,\omega)$ (top), $\hat{s}_4(t,\omega)$ (middle) and $\hat{s}_5(t,\omega)$ (bottom).

the log-spectral density of the first time series. The main structure of the individual spectral density can be captured in the lower and upper bound. There is more uncertainty at the beginning of process.

4. CASE STUDIES

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

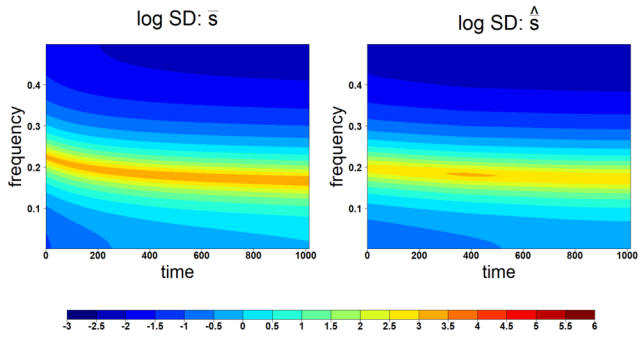

Figure 3. Left: True baseline of log-spectral density $\bar{s}(t,\omega)$. Right: Estimated baseline of log-spectral density $\hat{\bar{s}}(t,\omega)$.

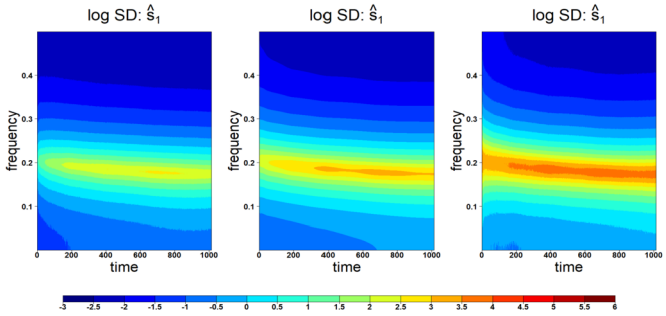

Figure 4. Left: Lower bound of a 95% posterior interval of the log-spectral densities $s_1(t,\omega)$. Middle: Estimated mean of the log-spectral densities $s_1(t,\omega)$. Right: Upper bound of a 95% posterior interval of the log-spectral densities $s_1(t,\omega)$.

males, age 22.5 ± 4.8 years) performing this experiment. Details regarding the data collection methodology and analysis via autoregressive models are available in [15]. The dataset used here is also publicly available [see references in 15].

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 dataset contains time series that correspond to epochs going from -1 s to 2 s, 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 $\pm$ 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 $\boldsymbol{m}_{f,m,0} = \boldsymbol{m}_{b,m,0} = \mathbf{0}$ 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 $\pm$ 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.

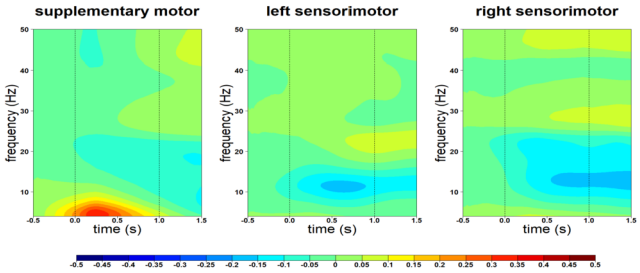

Figure 5. Cortical event-related spectral perturbations (ERSPs) to physical pulls.

Figure 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 -0.5 s 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 -0.5 s 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 [15]. 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 6. The increased spectral power 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 cor-

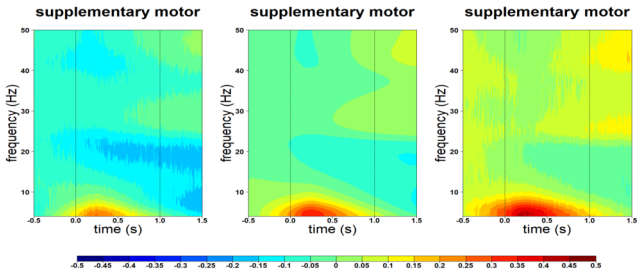

Figure 6. Left: Lower bound of a 95% posterior interval of the log-spectral density of supplementary motor cluster. Middle: Estimated mean of the log-spectral density of supplementary motor cluster. Right: Upper bound of a 95% posterior interval of the log-spectral density of supplementary motor cluster.

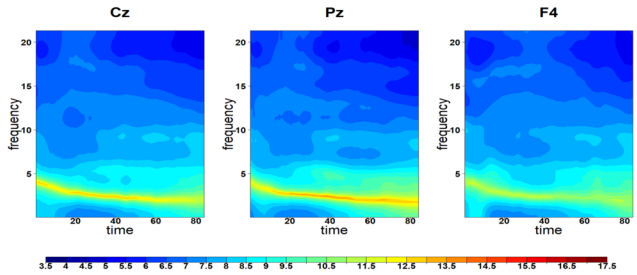

Figure 7. Estimated log-spectral densities for channels Cz, Pz and F4.

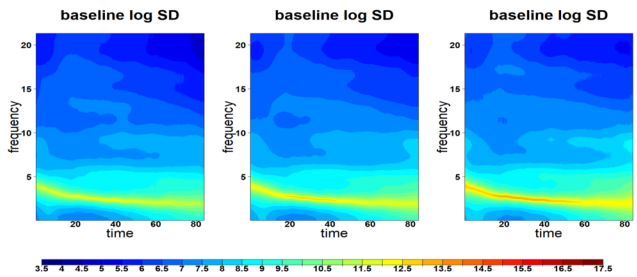

Figure 8. Left: Lower bound of a 95% posterior interval of the baseline log-spectral density. Middle: Estimated mean of baseline log-spectral densities. Right: Upper bound of a 95% posterior interval of the baseline log-spectral density.

responding 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.

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 dataset analyzed in [23] using univariate TVARs and in [18] and using dynamic regression models. [13] presents an analysis of these data latent threshold TV-VAR models. [25] also analyzes these data using a multivariate dynamic PARCOR model, focusing on inferring time-frequency measures of association, such as time-varying coherence and partial-coherence, across multiple channels.

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. 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 subsampled every sixth observation from the highest amplitude portion of the seizure, leading to a set of time series of 3,600 observations (corresponding to 84.375 s) per channel [18].

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 paper 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 7 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 [23] and [25]. 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. Unlike the multivariate model of [25], 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 8 and provides a time-frequency summary of the features underlying the EEG time series.

5. DISCUSSION

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 datasets 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 4. Future developments include incorporating additional prior structures, such as sparsity priors and time-varying sparsity priors [e.g., 1, 12, 13, 19, 7, among other], that can further reduce the dimensionality of the PARCOR model, potentially also in a dynamic and adaptable fashion.

ACKNOWLEDGEMENTS

Raquel Prado and Wenjie Zhao are partially funded by NSF award SES-1853210.

Received 1 April 2021

REFERENCES

[1] Bitto, A. and Frühwirth-Schnatter, S. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210:75–97, 2019. MR3944764

[2] Brockwell, P. J. and Davis, R. A. *Time Series: Theory and Methods*. Springer, New York, second edition, 1991. MR2839251

[3] Cadonna, A., Kottas, A. and Prado, R. Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, 2019. MR4047304

[4] Gamerman, D. and Migon, H. S. Dynamic hierarchical models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 629–642, 1993. MR1223932

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

[6] Hu, L., Guindani, M., Fortin, N. and Ombao, H. A hierarchical Bayesian model for differential connectivity in multi-trial brain signals. *Econometrics and Statistics*, 15:117–135, 2020. MR4124739

[7] Huber, F., Koop, G. and Onorante, L. Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business and Economic Statistics*, 2020. MR4272927

[8] Kitagawa, G. *Introduction to Time Series Modeling*. Chapman & Hall, Boca Raton, FL, 2010. MR2647741

[9] Kitagawa, G. and Gersch, W. *Smoothness Priors Analysis of Time Series*. Ima Volumes in Mathematics and Its Applications. Springer New York, 1996. MR1441074

[10] Krafty, R. T., Rosen, O., Stoffer, D. S., Buysse, D. J. and Hall, M. H. Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association*, 112:1405–1416, 2017. MR3750864

[11] Millar, R. B. and McKechnie, S. A one-step-ahead pseudo-DIC for comparison of bayesian state-space models. *Biometrics*, 70(4):972–980, 2014. MR3295758

[12] Nakajima, J. and West, M. Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164, 2013. MR3055329

[13] Nakajima, J. and West, M. 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. MR3738175

[14] Ombao, H., Raz, J. A., von Sachs, R. and Malow, B. A. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96(454):543–560, 2001. MR1946424

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

[16] Prado, R. and West, M. 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.

[17] Prado, R. and West, M. *Time Series: Modeling, Computation, and Inference*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2010.

[18] Prado, R., West, M. and Krystal, A. D. 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.

[19] Rockova, V. and McAlinn, K. Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*, 16:233–269, 2021. MR4194280

[20] Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and Its Applications: With R Examples*. Springer texts in statistics. Springer, 2017. ISBN 9780387293172. MR3642322

[21] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002. MR1979380

[22] West, M. and Harrison, J. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer New York, 1997. MR1482232

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

[24] Yang, W. H., Holan, S. H. and Wikle, C. K. Bayesian lattice filters for time-varying autoregression and time–frequency analysis. *Bayesian Analysis*, 11(4):977–1003, 2016. MR3545471

[25] Zhao, W. and Prado, R. Efficient Bayesian PARCOR approaches for dynamic modeling of multivariate time series. *Journal of Time Series Analysis*, 41:759–784, 2020. MR4176359

Wenjie Zhao
PhD Candidate
Department of Statistics
University of California Santa Cruz
USA
E-mail address: wzhao24@ucsc.edu

Raquel Prado
Professor
Department of Statistics
University of California Santa Cruz
USA
E-mail address: raquel@soe.ucsc.edu