Bayesian Time-Varying Tensor Vector Autoregressive Models for Dynamic Effective Connectivity

Wei Zhang∗1, Ivor Cribben†2,3, Sonia Petrone‡1, and Michele Guindani§4

1Department of Decision Sciences, Bocconi University
2Department of Accounting and Business Analytics, Alberta School of Business
3Neuroscience and Mental Health Institute, University of Alberta
4Department of Biostatistics, University of California, Los Angeles

May 30, 2024

Abstract

In contemporary neuroscience, a key area of interest is dynamic effective connectivity, which is crucial for understanding the dynamic interactions and causal relationships between different brain regions. Dynamic effective connectivity can provide insights into how brain network interactions are altered in neurological disorders such as dyslexia.
Time-varying vector autoregressive (TV-VAR) models have been employed to draw inferences for this purpose. However, their significant computational requirements pose challenges, since the number of parameters to be estimated increases quadratically with the number of time series. In this paper, we propose a computationally efficient Bayesian time-varying VAR approach. For dealing with large-dimensional time series, the proposed framework employs a tensor decomposition for the VAR coefficient matrices at different lags. Dynamically varying connectivity patterns are captured by assuming that at any given time only a subset of components in the tensor decomposition is active. Latent binary time series select the active components at each time via an innovative and parsimonious Ising model in the time-domain. Furthermore, we propose parsity-inducing priors to achieve global-local shrinkage of the VAR coefficients, determine automatically the rank of the tensor decomposition and guide the selection of the lags of the auto-regression. We show the performances of our model formulation via simulation studies and data from a real fMRI study involving a book reading experiment.

**Keywords**— fMRI experiments, Tensor factorization, Granger Causality, Shrinkage priors, Temporal Ising model, MC³ algorithm

1 Introduction

Investigating the integration among different brain regions to understand how cognitive information is processed and distributed is a central goal of many functional magnetic resonance imaging (fMRI) experiments. Neuroscientists typically distinguish between functional connectivity, which measures the undirected associations, or temporal correlation, between the fMRI time series observed at different locations, and effective connectivity, which estimates the directed influences that one brain region exerts onto other regions [18, 70, 13]. Unlike functional connectivity, which only measures the correlation between brain regions, effective connectivity aims to determine the direction and influence one brain region exerts over another. This offers researchers a way to explore causal relationships in brain activity. Effective connectivity can also provide insights into how brain network interactions are altered in neurological and psychiatric disorders, aiding in the development of diagnostic markers and improving our understanding of disease mechanisms. By identifying normal and abnormal effective connectivity patterns, researchers can develop targeted therapies. For example, for reading disorders such as dyslexia [15, 16], understanding altered brain networks can help in designing more effective
intervention strategies, including neurostimulation or behavioral therapies. The study of effective connectivity can also shed light on how learning and experience shape brain network interactions over time, contributing to our understanding of neural plasticity.

One way to model effective connectivity is via a vector auto-regression (VAR) model, a widely-employed framework for estimating temporal (Granger) casual dependence in fMRI experiments [see, e.g. 21, 8]. Furthermore, it is now recognized that the connectivity patterns may vary dynamically throughout the fMRI experiment, as we anticipate in our motivating dataset [69]. This is supported by recent studies in neuroscience that emphasize the importance of describing changes in brain connectivity in response to stimuli in task-based settings [10, 41, 2] and due to spontaneous fluctuations in resting-state fMRI [26, 57, 47, 65, 69]. A Markov-switching VAR model formulation has been utilized by Samdin et al. [51] and Ombao et al. [42] to characterize dynamic connectivity regimes among a few selected EEG channels. Li et al. [34] developed a stochastic block-model state-space multivariate autoregression to investigate how abnormal neuronal activities start from a seizure onset zone and propagate to otherwise healthy regions using intracranial EEG data.

VAR models can become computationally intensive when analyzing large-dimensional time-series, since the number of parameters to estimate increases quadratically with the number of time-series, often exceeding the number of observed time points. Thus, several approaches have been proposed to enforce sparsity of the VAR coefficient matrix, either by using penalized-likelihood methods [53, 4] or - in a Bayesian setting - by using several types of shrinkage priors [48, 30, 20]. Alternatively, dimension reduction techniques have been employed to reveal and exploit a lower dimensional structure embedded in the parameter space. For example, Velu et al. [60] decompose the VAR coefficient matrix as the product of lower-rank matrices. Wang et al. [63] have proposed an $L_1$-penalized-likelihood approach where a tensor decomposition is employed to represent the elements of the VAR coefficient matrices. Luo and Griffin [35] propose a static Bayesian VAR that decomposes the coefficient matrix into a low-rank tensor where a multiplicative gamma prior and an adaptive inferential scheme are used to estimate the tensor margins and rank parameters. Similarly, Fan et al. [14] developed a Bayesian random effects VAR model for large-dimensional multi-subject panel neuroimaging data where the dimension reduction is achieved by applying a Tucker tensor decomposition. Billio et al. [5] propose a Bayesian autoregressive tensor framework to model the (static) parameters of a time-series regression, allowing for both tensor-valued multi-layer responses and covariates, and employing impulse response analysis to study the propagation of shocks across both the network and the layers.
In this paper, motivated by an experimental study on dynamic effective connectivity patterns arising when reading complex texts, we propose a computationally efficient time-varying Bayesian VAR approach for modeling large-dimensional time series. Similarly as in Wang et al. [63], we assume a tensor decomposition for the VAR coefficient matrices at different lags. A novel feature of the proposed approach is that we capture dynamically varying connectivity patterns by assuming that – at any given time – the VAR coefficient matrices are obtained as a sum of a subset of active components in the tensor decomposition. This sum representation relies on latent indicators of brain activity, that we model through an innovative use of an Ising model on the time-domain, to select what components are active at each time. With respect to Hidden Markov Models – typically employed in the fMRI literature to capture transitions across brain states dynamically over time – the Ising specification models the time-varying activations as a function of only two parameters. The resulting binary time series still maintains a Markovian dependence, but the Ising model naturally assigns a higher probability mass to non-active (zeroed) components to encourage sparsity of the representation and it favors similar selections at two consecutive time points, reflecting the prior belief that the coefficients are changing slowly over time. Furthermore, we show that the Ising model can be represented as the joint distribution of a so-called (new) discrete autoregressive moving average (NDARMA) model [27], a result which is helpful for prior elicitation.

The remaining components of the model are designed to encourage sparsity in the tensor structure and to ascertain model complexity directly from the data through the posterior distribution. Specifically, we employ a multi-way Dirichlet generalized double Pareto prior [23, 5], which allows for global-local shrinkage of the VAR coefficients. This choice enables the model to implicitly determine the effective rank of the tensor decomposition based on the observed data. An additional feature of our approach is that it assumes an increasing-shrinkage prior [31] to guide the selection of the lags of the auto-regression, eliminating the need to rank models of different order based on model selection information criteria. We show how inference on the VAR coefficient matrices can be used to create time-varying Granger causality networks, enabling a detailed analysis of dynamic connections over time.

In our fMRI application involving a book reading experiment, we find that changes in the effective connectivity coincide with key events in the book and the connectivity hubs identified are consistent with the reading literature. Although only controls were studied in this data, the new methodology could be used in a study of neurological and psychiatric disorders (such as dyslexia) where the objective is to understand the differences in network interactions. This study has the potential to aid in the development of diagnostic markers and improve our understanding of such disease mechanisms. The key contributions of the new methodology in terms
of analyzing fMRI data include effective connectivity networks at each experimental time point (which to the best of our knowledge has not been possible before), and Granger causality networks inferred from the BTVT-VAR model.

The posterior distribution in large-dimensional tensor models is often multimodal, and standard MCMC algorithms may struggle to efficiently explore the entire posterior space, potentially getting stuck in local modes. Thus, we obtain posterior inference using an efficient Gibbs sampler with tempering, the Metropolis coupled Markov chain Monte Carlo (MC³) algorithm proposed by Altekar et al. [1]. By running parallel chains at different temperatures and allowing them to swap states, MC³ allows for an improved exploration of the multimodal posterior distribution.

The remainder of the paper is organized as follows. In Section 2, we formulate the time-varying tensor model and elucidate how to obtain dimension reduction via a tensor decomposition into a set of latent base matrices and binary indicators of connectivity patterns over time. In Section 3, we describe the Ising model specification on the temporal transitions, whereas in Section 4 we describe the sparsity-inducing priors on the active elements of the tensor decomposition. In Section 5 we discuss posterior computation and inference. Results of the simulation studies as well as the real data application are shown in Section 6 and Section 7, respectively. Finally, Section 8 provides some concluding remarks and future work.

2 Time-Varying Tensor VAR model for Effective Connectivity

In this section, we introduce the proposed time-varying tensor VAR specification for studying dynamic brain effective connectivity. Let \( y_t \) be an \( N \)-dimensional vector for \( t = 1, \ldots, T \). Each time-series data \( (y_{t1}, \ldots, y_{tN}) \) represents the fMRI BOLD signal recorded at a voxel or region of interest (ROI) \( i, i = 1, \ldots, N \). The time-varying tensor VAR model of order \( P \) assumes that \( y_t \) is a linear combination of the \( P \) lagged signals \( y_{t-1}, \ldots, y_{t-P} \) plus an independent noise term, \( \epsilon_t \in \mathbb{R}^N \),

\[
y_t = \begin{bmatrix} A_{1,t}, A_{2,t}, \ldots, A_{P,t} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-P} \end{bmatrix} + \epsilon_t,
\]

where \( \epsilon_t \sim \mathcal{N}(0, \Sigma) \) and the linear coefficients \( A_{j,t}, j = 1, \ldots, P \) are \( N \times N \) matrices which are allowed to vary across \( t, t = 1, \ldots, T \). We assume that \( \Sigma \) is time-invariant and diagonal, and we focus on the coefficient matrices \( [A_{1,t}, A_{2,t}, \ldots, A_{P,t}] \). If
needed, the assumption on \( \Sigma \) can be appropriately relaxed. The number of coefficients to be estimated is \((T - P) \times N^2 \times P + N\); hence, it is not possible to use the conventional ordinary least square estimator. We propose a model that induces a low-rank sparse representation of the coefficient matrices, implementing several concurrent strategies.

2.1 Time-Varying Latent Component Representation of the VAR coefficient matrices.

To start, we parsimoniously model the dynamic VAR coefficient matrices \([A_{1,t}, \ldots, A_{P,t}]\) as a *time-varying composition* of \(H\) latent static base matrices \([A^*_{1,h}, \ldots, A^*_{P,h}], h = 1, \ldots, H\). Specifically, for \(h = 1, \ldots, H\), we introduce a binary-valued time series, \((\gamma_{h,t})_{t \geq P+1}\). Then, we assume

\[
[A_{1,t}, A_{2,t}, \ldots, A_{P,t}] = \sum_{h=1}^{H} \gamma_{h,t} \left[ A^*_{1,h}, A^*_{2,h}, \ldots, A^*_{P,h} \right],
\]  

(2)

that is, at any given time \(t\), the VAR coefficient matrices are built as a sum of a subset of base matrices \([A^*_{1,h}, \ldots, A^*_{P,h}]\), activated when \(\gamma_{h,t} = 1\). For instance, if at time \(t\), \(\gamma_{h,t} = 1\) for \(h = 1, 2\) and zero for \(h = 3, \ldots, H\), then \(A_{j,t} = A^*_{j,1} + A^*_{j,2}, j = 1, \ldots, P\). However, our primary interest is not in the base matrices \([A^*_{1,h}, \ldots, A^*_{P,h}]\), but rather in accurately recovering the time-varying VAR coefficients. The binary \(\gamma_{h,t}\)’s can be interpreted as indicators of latent individual or experimental conditions. Similarly, \(A^*_{1,h}, A^*_{2,h}, \ldots, A^*_{P,h}\) can be interpreted as *latent base matrices*.

Hidden Markov Models (HMM) with time-invariant transition probabilities have been used in the neuroimaging literature to describe temporal variations in functional connectivity patterns [3, 61, 65]. Our proposed representation assumes that the dynamics of connectivity patterns result from the juxtaposition and cessation of the base connectivity matrices over time. An equivalent HMM would assume \(2^H\) states, leading to a \(2^H \times 2^H\) transition matrix, which would require estimating a prohibitively large number of parameters, even with moderate \(H\). The proposed approach leads to a more parsimonious representation and aligns more closely with current thinking among neuroscience investigators, as it reflects the belief that connectivity states do not undergo sudden changes over time. Instead, they emerge from the aggregation of multiple underlying states. Functional connectivity is thus understood as arising from latent variables derived from the shared variances in connectivity across task states [46, 38].

Compared to estimating \(N^2 \times P\) time-series of length \((T - P)\) in the initial model specification, our formulation (2) requires estimating \(N^2 \times P\) base matrices.
The dynamics of the VAR coefficient matrices are then governed by the temporal dependence between the $\gamma_{h,t}$’s, $h = 1, \ldots, H$, $t = P + 1, \ldots, T$. We assume independence among the $\gamma_{h,t}$’s across $h = 1, \ldots, H$, and model Markovian dependence across different time points through a latent Ising model that depends only on $2 \times H$ parameters (see Section 3).

2.2 Rank-varying PARAFAC decomposition

Despite the reduced dimensionality, (2) remains highly parameterized. Hence, we further propose to stack each set of matrices $[A_{1,h}^*, A_{2,h}^*, \ldots, A_{P,h}^*]$ into a three-way tensor $A_h^*$ of size $N \times N \times P$, and assume it can be decomposed as

$$A_h^* = \alpha_{1,h} \circ \alpha_{2,h} \circ \alpha_{3,h}, \quad h = 1, \ldots, H,$$

where $\circ$ indicates the vector outer product and $\alpha_{1,h}, \alpha_{2,h} \in \mathbb{R}^N, \alpha_{3,h} \in \mathbb{R}^P$ are the tensor margins along the dimensions of the tensor. Thus, the time-varying VAR coefficients in (2) can be stacked and be represented as the tensor

$$A_t = \sum_{h=1}^{H} \gamma_{h,t} A_h^* = \sum_{h=1}^{H} \gamma_{h,t} (\alpha_{1,h} \circ \alpha_{2,h} \circ \alpha_{3,h}).$$

(3)

As the binary latent indicator $\gamma_{h,t}$ may be zero for some $t$, the decomposition (3) effectively defines a time-varying PARAFAC decomposition, with an effective rank that changes dynamically over time and is always less than or equal to $H$. Thus, at each time point, the time-varying tensor VAR coefficients are still represented as a combination of active subsets from the PARAFAC decomposition.

The PARAFAC decomposition, along with the Tucker decomposition, is often employed for tensor dimension reduction due to its straightforward interpretation and implementation. Tensors have been utilized in the neuroimaging literature for detecting activations via tensor regression approaches [73, 23]. However, to the best of our knowledge, their use for studying dynamic effective connectivity within VAR models has not yet been explored. For a brief introduction to the PARAFAC decomposition, we refer to Section S1 of the Supplementary Material [72].

The number of parameters in (3) is $H(T - P) + H(2N + P)$, that is, linear in the observation size $N$, instead of $N^2$ without the tensor reparameterization. The decomposition is invariant under any permutation of the component indices $h$. In other words, for any permutation $\Pi(\cdot)$ of the index set $\{1, 2, \ldots, H\}$, $A_t = \sum_{h=1}^{H} \gamma_{\Pi(h),t} \alpha_{1,\Pi(h)} \circ \alpha_{2,\Pi(h)} \circ \cdots \circ \alpha_{M,\Pi(h)}$. Additionally, $A_t$ is not affected by rescaling. Specifically, $A_t = \sum_{r=1}^{R} \alpha_{1,h}^* \circ \alpha_{2,h}^* \circ \cdots \circ \alpha_{M,h}^*$, where $\alpha_{m,r}^* = \nu_{m,r} \alpha_{m,r}$ for any set of multiplying factors $\nu_{m,r}$ such that $\prod_{m=1}^{M} \nu_{m,r} = 1$. Indeed, it is
important to note that while inference on the individual tensor margins may suffer from identifiability issues, the tensor $A_t$ and the matrices $A_{j,t}$, $j = 1, \ldots, P$ remain identifiable. Specifically, the temporal pattern of the VAR coefficients, obtained from (2)–(3), remains identifiable.

Using tensor algebra, we can rearrange the modes of the tensor decomposition in (2) to express the base matrices as follows:

$$[A^*_1,h, A^*_2,h, \ldots, A^*_P,h] = \alpha^*_{3,h} \otimes (\alpha_{1,h} \circ \alpha_{2,h}), \quad h = 1, \ldots, H,$$

where $\otimes$ denotes the Kronecker product. The matricization operation highlights that the element-by-element ratio between $A^*_1,h$ and $A^*_2,h$ is proportional to the ratio between the first two entries of $\alpha_{3,h}$, and similarly for subsequent lags. After matricization, the set of time-varying VAR coefficient matrices can be expressed as

$$[A_{1,t}, A_{2,t}, \ldots, A_{P,t}] = \sum_{h=1}^{H} \gamma_{h,t} \alpha^*_{3,h} \otimes (\alpha_{1,h} \circ \alpha_{2,h}),$$

where the latent binary time-series $\gamma_{h,t}$'s selects the active components at each time, $h = 1, \ldots, H$. The elements of the tensor margin $\alpha_{3,h} \in \mathbb{R}^P$, corresponding to the lag $P$ of the VAR model, play a crucial role in determining the weighting of the selected components. We expect the influence of past variables to diminish with increasing lags; therefore, it is reasonable to expect that the magnitude of the entries of $\alpha_{3,h}$ would also decrease as the lags increase. In Section 4, we describe prior specifications that promote sparsity in $\alpha_{1,h}$ and $\alpha_{2,h}$ while also enforcing increasing shrinkage in $\alpha_{3,h}$.

Sun and Li [56] have also proposed the stacking of a series of dynamic tensors to form a higher order tensor. In their approach, the data are observed tensors to be clustered over time along the modes generated via the PARAFAC decomposition. Instead, in our approach the data are multivariate time series and the tensor structure is used to construct a lower dimensional parameter space for the unknown VAR coefficients to be estimated. Moreover, Sun and Li [56] achieve smoothness in the parameters through a fusion structure that penalizes discrepancies between neighboring entries in the same tensor margin. We follow a Bayesian approach and further encourage a contiguous structure by means of the Ising model specification detailed in the following section.

## 3 Latent Ising modeling for temporal transitions

The sequence of latent indicators $\gamma_{h,t}$ determines the time-varying activations of the base matrices in the VAR model (1)–(2). To model this latent process, we
propose an innovative use of the Ising model in the time domain. This approach encodes a Markovian dependence but allows modeling the time-varying activations as a function of only two parameters for each \( h = 1, \ldots, H \): one parameter captures general sparsity, and the other captures the strength of dependence between adjacent time points. More specifically, we assume that, independently for each \( h \), the binary state process \((\gamma_{h,t})_{t>P}\) is characterized by joint probability mass functions

\[
\pi(\gamma_{h,P+1}, \ldots, \gamma_{h,T} | \theta_h, \kappa_h) 
\propto \exp \left( \theta_h \gamma_{h,P+1} + \sum_{t=P+2}^{T-1} \theta^*_h \gamma_{h,t} + \theta_h \gamma_{h,T} + \sum_{t=P+1}^{T-1} \kappa_h \gamma_{h,t} \gamma_{h,t+1} \right). \tag{4}
\]

Equation (4) defines an Ising model or an undirected graphical model or a Markov random field involving the binary random vector \( \gamma_h = (\gamma_{h,P+1}, \ldots, \gamma_{h,T}) \in \{0, 1\}^{T-P}, h = 1, \ldots, H \) [see, e.g., 62]. The parameters \( \theta_h \) and \( \theta^*_h \) can be interpreted as sparsity parameters, since they correspond to the probability of activation for component \( h \) at each time \( t \), irrespective of the status at \( t-1 \) and \( t+1 \). Positive values of \( \theta_h \) and \( \theta^*_h \) increase the probability that \( \gamma_{h,t} = 1 \); on the other hand, negative values of \( \theta_h \) and \( \theta^*_h \) increase the probability that \( \gamma_{h,t} = 0, \ t = P + 1, \ldots, T \). The parameter \( \kappa_h \) captures the effect of the interaction between \( \gamma_{h,t} \) and \( \gamma_{h,t+1} \), that is, the persistence of activations over time. In particular, when \( \kappa_h > 0 \), the probability that \( \gamma_{h,t} \) and \( \gamma_{h,t+1} \) are both non-zero is larger.

The Ising model (4) can be seen as a specific instance of a multivariate Bernoulli distribution, as defined by Dai et al. [11]. In particular, in the following, we show that the proposed Ising process is equivalent to a binary discrete autoregressive NDARMA(1) model [27, 36, 28]. For notational simplicity, we focus on a single time series \( \gamma_{h,t} \), and we omit the subscript \( h \) for the remainder of the section. We start by recalling that a NDARMA(1) process is a binary time series that satisfies

\[
\gamma_t = a_t \gamma_{t-1} + (1 - a_t) \epsilon_t, \quad t = 1, \ldots, T, \tag{5}
\]

where \( a_t \overset{i.i.d.}{\sim} \text{Bern}(p_1) \), and \( \epsilon_t \overset{i.i.d.}{\sim} \text{Bern}(p_2) \), with independent \( a_t \) and \( \epsilon_t \). The initial condition assumes \( \gamma_1 \sim \text{Bern}(p_2) \). The NDARMA(1) model has a Markovian dependence structure, with transition probabilities

\[
P(\gamma_t = 1 | \gamma_{t-1}) = p_1 \mathbb{1}(\gamma_{t-1} = 1) + (1 - p_1) p_2,
\]

for \( \gamma_t, \gamma_{t-1} \in \{0, 1\} \). Moreover, marginally \( \gamma_t \sim \text{Bern}(p_2) \). Intuitively, the autocorrelation function at lag 1 of the NDARMA time series is always positive, meaning that \( \gamma_t \) and \( \gamma_{t+1} \) tend to assume the same value, consistent with the contiguous
behavior that the temporal-Ising model (4) encourages when $\kappa$ is assumed to be positive. Then, in a NDARMA(1) model, the joint probability mass function of $\gamma_1, \ldots, \gamma_T$ can be characterized a multivariate Bernoulli distribution [11], as

$$P(\gamma_1, \ldots, \gamma_T) = p_{0\ldots0}^T \gamma_1 (1-\gamma_1) \gamma_2 (1-\gamma_2) \cdots \gamma_T (1-\gamma_T),$$

where $p_{0\ldots0}$ indicates the probability of a zero sequence, that is $p_{0\ldots0} = P(\gamma_1 = 0, \ldots, \gamma_T = 0) = (1 - p_2) \prod_{t=2}^T (p_1 + (1 - p_1)(1 - p_2))$ and, similarly, for the other factors. Let $n = \sum_{t=1}^T \gamma_t$ indicate the total number of active indicators $\gamma_t$'s along the entire time-series. The following proposition maps the parameters $(\theta, \kappa)$ in the Ising model (4) to the parameters $(p_1, p_2)$ in the NDARMA(1) model in (5):

**Proposition 1.** The probability law of the NDARMA(1) model in (5) can be expressed as in (4). In particular, the parameters $(\theta, \kappa)$ are obtained as a function of the parameters $(p_1, p_2)$ in (5) as

$$e^\theta = \frac{p_2 (1 - p_1)}{p_1 + (1 - p_2)(1 - p_1)}, \quad e^\kappa = \frac{p_1 + p_2 (1 - p_2)(1 - p_1)^2}{p_2 (1 - p_2)(1 - p_1)^2},$$

$$\exp(\theta^*) = \frac{p_2 (1 - p_2)(1 - p_1)^2}{(p_1 + (1 - p_2)(1 - p_1))^2} = \frac{e^\theta (e^\theta + 1)}{e^{2\theta + \kappa} + 1}.$$  

Inversely,

$$p_1 = \frac{e^\theta (e^\kappa - 1)}{(e^{2\theta + \kappa} + 1)(e^\theta + 1)}, \quad p_2 = \frac{e^\theta (e^{\theta + \kappa} + 1)}{e^{2\theta + \kappa} + 2e^\theta + 1}.$$

We refer to Section S2 of the supplementary material for the proof [72].

### 3.1 Prior specification

The result in Proposition 1 is helpful in setting the prior distributions for the parameters in (4), as it highlights underlying constraints among the parameters and how the domain of $(\theta, \kappa)$ constrains the admissible range of $(p_1, p_2)$. Indeed, the transformation is bijective, since $\theta^*$ can be expressed as a function of the pair $(\theta, \kappa)$. The parameters $\kappa$ and $p_1$ have the same sign, since they both indicate the strength of the dependence between two neighboring $\gamma_{t-1}$ and $\gamma_t$. It also follows that, given that $\kappa$ is positive, $\theta^*$ is always smaller than $\theta$. As an illustration of the complex dependencies induced by the mapping between $(\theta, \theta^*, \kappa)$ in (4) and $(p_1, p_2)$ in (5), Fig 1 compares the range of $(p_1, p_2)$ for different intervals of values of $(\theta, \kappa)$. For example, if $-2 < \theta < 2$ and $0 < \kappa < 1$, the corresponding set of NDARMA(1) models is limited to a subset of those with $p_1 < 0.25$, whereas if $-8 < \theta < 8$
Figure 1: An illustration of the mapping between the parameters values of 
\((p_1, p_2)\) of the NDARMA(1) model (5) and the parameters \((\theta, \kappa)\) of the Ising 
model (3). The domain of \((\theta, \kappa)\) constrains the admissible range of \((p_1, p_2)\). 
The three panels illustrate the domain of \((p_1, p_2)\) corresponding to increasing 
domains of \((\theta, \kappa)\): (1) \(-2 < \theta < 2, 0 < \kappa < 1\); (2) \(-4 < \theta < 4, 0 < \kappa < 5\); (3) 
\(-8 < \theta < 8, 0 < \kappa < 10\).

and \(0 < \kappa < 10\), the support covers almost all values of \(p_1\) and \(p_2\) between \(0\) and 
1. As the domain of \((\theta, \kappa)\) expands, the set of induced NDARMA(1) models also 
expands. Shrinking \(\gamma_t\) towards zero is desirable for regularization purposes, which 
corresponds to allowing negative values of the parameter \(\theta\). However, excessive 
shrinkage may hinder our ability to identify recurrent latent base patterns, as it 
can lead to very low estimates of \(p_1\) and \(p_2\). Indeed, it is well known that the prior 
specification of the parameters of a Ising model needs to be conducted with care, in 
order to avoid the phenomenon of phase-transition [32, 33]. In statistical physics, 
a phase-transition refers to a sudden change from a disordered (non-magnetic) to 
an ordered (magnetic) state at low temperatures. In Bayesian variable selection, 
the phase-transition has been associated to values of the parameter space that 
lead to selecting either all or none of the tested variables. These considerations 
motivate our suggestion of a proper uniform prior distribution on the parameters 
\((\theta, \kappa)\) over a closed interval in \(\mathbb{R}^2\). More specifically, we assume that, for each 
h = 1, \ldots, H, \(\theta_h\) lies between \([\theta_{h,\text{min}}, \theta_{h,\text{max}}]\) with lower limit \(\theta_{h,\text{min}} < 0\) and upper 
limit \(\theta_{h,\text{max}} > 0\). We have found that choosing \(\theta_{h,\text{min}} = -8\) and \(\theta_{h,\text{max}} = 8\) ensures 
a proper exploration of the parameter space and appears to avoid phase transitions. 
Similarly, for \(\kappa_h\), we encourage a contiguous structure where \(\gamma_{h,t}\) and \(\gamma_{h,t+1}\) are 
simultaneously selected by assuming that \(\kappa_h\) is positive with a uniform prior on 
\(\kappa_h \in [0, \kappa_{h,\text{max}}]\) with \(\kappa_{h,\text{max}} > 0\). Also, in this case, an upper limit \(\kappa_{h,\text{max}} = 10\) 
appears to ensure both reasonably good inference on the time-varying coefficients.
and computational efficiency.

4 Sparsity-inducing priors on the tensor components

In addition to achieving dimension reduction through the PARAFAC decomposition, we further seek shrinkage of the tensor margins' parameters. Hence, we consider priors that shrink the parameters toward zero, enabling a sparse representation of the VAR coefficients and more interpretable estimation of the connectivity patterns. In particular, for the elements of the tensor margins $\alpha_{1,h}$ and $\alpha_{2,h}$ we consider a multi-way Dirichlet generalized double Pareto prior [23], whereas for the lag margin $\alpha_{3,h}$, we consider an increasing shrinkage prior, so that higher-order lags are penalized.

More specifically, we assume that the $N$-dimensional tensor margins $\alpha_{1,h}$ and $\alpha_{2,h}$ are normally distributed with zero mean and variance-covariance matrix $\tau \phi_h W_{l,h}$, with $W_{l,h} = \text{diag}(\{w_{l,h,k}\}_{k=1}^{N})$, $l = 1, 2$. The parameter $\tau$ is a global scale parameter that follows a Gamma distribution, whereas $\phi_1, \ldots, \phi_H$ are local scale parameters that follow a symmetric Dirichlet distribution. The diagonal elements of the covariance matrix have a generalized double Pareto prior. Thus,

$$\alpha_{l,h} | \phi_h, \tau, W_{l,h} \sim \mathcal{N}(0, \phi_h \tau W_{l,h}), \quad l = 1, 2$$

$$w_{l,h,k} | \lambda_{l,h} \sim \text{Exp}\left(\lambda_{l,h}^2/2\right) \quad 1 \leq k \leq N, l = 1, 2$$

$$\lambda_{1,h}, \lambda_{2,h} \overset{i.i.d.}{\sim} \text{Ga}(a_\lambda, b_\lambda)$$

$$\phi_1, \ldots, \phi_H \sim \text{Dirichlet}(a, \ldots, a),$$

$$\tau \sim \text{Ga}(a_\tau, b_\tau).$$

By setting $a_\tau = Ha$, one can obtain tractable full conditionals distributions for $\tau$ and $(\phi_1, \ldots, \phi_H)$, since the full conditional for $\tau$ is a generalized inverse Gaussian distribution and the full conditionals for $(\phi_1, \ldots, \phi_H)$ are normalized generalized inverse Gaussian random variables [23].

Also for the lag mode parameter, $\alpha_{3,h}$, we employ a normal prior with a global-local structure on the covariance, specifically, $\alpha_{3,h} | \phi_h, \tau, W_{3,h} \sim \mathcal{N}(0, \phi_h \tau W_{3,h})$. However, we incorporate a cumulative shrinkage effect on the diagonal entries of $w_{3,h}$ to encourage the estimation of a small number of lags. More specifically, we use a cumulative shrinkage prior [31]. This type of prior induces increasing shrinkage through a series of spike and slab distributions, which assign greater mass to a target spike value as model complexity increases. Let $W_\infty$ indicate the target spike. Here, we set $W_\infty = 0.01$ to prevent the Normal distribution from
degenerating into a point mass and improve computational efficiency [19]. Then, for $1 \leq j \leq P$, we assume that the diagonal elements of $W_{3,h}$, \\

$$w_{3,h,j} \mid z_{h,j} \sim [1 - \mathbb{1}(z_{h,j} \leq j)] \text{InvGa}(a_w, b_w) + \mathbb{1}(z_{h,j} \leq j) \delta_{W_{\infty}},$$ \\

where each $z_{h,j}$ is a draw from a Multinomial random variable with $\text{pr}(z_{h,j} = l \mid \omega_{h,l}) = \omega_{h,l}$, for $l = 1, \ldots, P$. Thus, the probability of selecting the target spike is increasing with the lags $j$, since $P(z_{h,j} \leq j) = \sum_{l=1}^{j} \omega_{h,l}$. The weights $\omega_{h,l}$ are obtained through a stick-breaking construction [52], that is, $\omega_{h,l} = v_{h,l} \prod_{m=1}^{l-1}(1-v_{h,m})$, $v_{h,l} \sim \text{Beta}(\beta_1, \beta_2)$, $1 \leq l \leq P$. Correspondingly, the probability of choosing the Inverse Gamma slab component is $P(z_{h,j} > j) = \prod_{l=1}^{j}(1-v_{h,l})$, which decreases with $j$. Higher sparsity levels for the modes $\alpha_{1,h}, \alpha_{2,h}$ and $\alpha_{3,h}$ are obtained by setting smaller values of $a_\tau$ relative to $b_\tau$ or by setting smaller values of $b_\lambda$ relative to $a_\lambda$. This is evident from the expectations $E(\tau) = a_\tau/b_\tau$ and $E(W_{1,h,k}) = 2b_\lambda^2/[(\alpha_\lambda - 1)(\alpha_\lambda - 2)]$. We discuss these choices in Section 5.

4.1 Rank of the PARAFAC decomposition

A crucial point in the representation (3) is the choice of the rank, $H$. One widely adopted option is to regard this choice as a model selection problem and naturally resort to information criteria such as AIC or BIC [73, 62, 64, 12]. Luo and Griffin [35] use an increasing shrinkage prior for inference on the rank. Here, we propose a different strategy. Drawing from recent Bayesian literature on overfitted mixture models [see, e.g., 37, 49], we set the parameters of the sparsity-inducing hierarchical prior in Section 4 to automatically shrink unnecessary components to zero. More specifically, a small concentration parameter $a$ of the symmetric Dirichlet distribution $(\phi_1, \ldots, \phi_H)$ assigns more probability mass to the edges of the simplex, meaning that more components become redundant. As a result, within the available $H$ components, only a few will be effectively different from zero. Additionally, we can utilize the cumulative shrinkage prior for the VAR lag order $P$ for further regularization, by selecting appropriate values for $\beta_1, \beta_2, a_w$ and $b_w$. For instance, a large value of $\beta_1$ and a small $\beta_2$ encourage a more parsimonious VAR model by assigning little probability to higher orders. Therefore, by accepting a modest increase in computational demand, it is possible to set relatively high values for $H$ (and $P$), and then rely on the implicit regularization of the shrinkage priors to determine the actual number of effective components (and lags), without the need for ranking different models in practice. Fig 2 summarizes the proposed hierarchical model on the $N$-dimensional time series $y_t$ using a directed graph representation.
5 Posterior Computation

To infer the dynamic coefficient matrices and latent indicators in our time-varying VAR model, we utilize Markov Chain Monte Carlo (MCMC) methods. Despite the initially daunting number of parameters, our approach reduces complexity through tensorization, dynamic composition of latent base matrices, and the temporal Ising process. However, an accurate use of MCMC methods is essential due to the complexity, multidimensional and dynamic nature of the time series.

The prior specification allows the use of a blocked Gibbs sampler to draw samples from the posterior distribution. When sampling from the posterior distribution for $\theta_h$ and $\kappa_h$ using a Metropolis-Hastings algorithm, the normalizing constant depends on the sampled parameters, giving rise to a well-known issue of sampling from a doubly-intractable distribution. Thus, we follow the auxiliary variable approach proposed by [39] to obtain the posterior samples from the Ising model. More specifically, the approach introduces an auxiliary variable such that – by adding its full conditional to the Metropolis-Hasting ratio – the normalizing
constant is canceled out. Posterior samples of the latent auxiliary variable are then obtained using the equivalent NDARMA representation which, remarkably, improves efficiency. The use of the auxiliary variable approach is also key for allowing the update of the $\gamma_{h,t}$'s for all $t = P + 1, \ldots , T$ and each $h = 1, \ldots , H$.

In our fMRI data application, to fully explore the geometry of the target posterior distribution, we implement the Metropolis coupled MCMC algorithm on top of the blocked Gibbs sampler. The core idea is to raise the BTVT-VAR likelihood to a sequence of descending powers $\rho$, resulting in multiple different posteriors. When $\rho = 1$, it is the target distribution and when $\rho = 0.1$, it corresponds to the almost flat posterior. The flatter the power posterior is, the easier it is to sample the whole surface. The multiple MCMCs, each assigned with a specific power $\rho$, are then run in parallel. At the end of every pre-fixed iteration, we perform a swap step where we randomly choose one chain with $\rho < 1$, calculate the Metropolis-Hastings ratio between the chosen chain and the target chain based on the last draws and decide whether or not to exchange them using the Metropolis-Hastings updating rule. The details of the Gibbs sampler and the MC$^3$ algorithm are reported in S3 and S4 of the supplementary material [72].

Through MCMC samples from the posterior distribution we can obtain inferences on the dynamic coefficient matrices $[A_{1,t}, A_{2,t}, \ldots , A_{P,t}]$. Posterior expectations at each time point are approximated by the MCMC averages. We summarize inference on the $\gamma_{h,t}$'s by computing the posterior mode, specifically we set $\tilde{\gamma}_{h,t} = 1$ whenever the posterior probability of activation, $P(\gamma_{h,t} = 1 | y_1, \ldots , y_T)$, is over 0.5.

6 Simulation Studies

We illustrate the performances of the proposed model formulation in two simulation studies.

Simulation 1: Synthetic data. We generate multivariate time series from a TVT-VAR model where the coefficients are estimated from the first two sessions of a subject’s fMRI data during the reading experiment detailed in Section 7. Specifically, we obtain the OLS estimates from a static VAR model with $N = 27$ and $P = 3$, fitting the two sessions separately. For each estimated coefficient matrix, we stack its elements and apply a rank-2 PARAFAC decomposition. The PARAFAC decomposition yields $H = 4$ different sets of tensor margins $\alpha_{1,h}, \alpha_{2,h}, \alpha_{3,h}$, which are then used to construct the time-varying coefficients. We generate 30 replicates for each dataset, with specific tensor margins sampled from normal distributions centered on $\alpha_{1,h}, \alpha_{2,h}, \alpha_{3,h}$, obtained from the real fMRI data as outlined above. The sampled tensor margins are further thresholded at 0.3 to induce a desirable
level of sparsity. For each of the 30 replicates, we then generate a multivariate time course of length $T = 400$ from a TV-VAR model with coefficients changing at time points 80, 160, 240, and 320, and assuming $H = 4$, $P = 3$. Within each time interval, the coefficients are assumed to be the sum of a time-invariant subset of the four base matrices calculated from the tensor margins.

The error term $\epsilon_t$ entries have identical standard deviations determined by the estimated signal-to-noise ratio (SNR) in the real data. Several SNR definitions exist in the literature [67, 71, 24]. Following Haslbeck et al. [24], we divide the estimated VAR coefficients’ maximum size by the estimated standard deviation of the error term, obtaining SNR values of 1.8518 and 2.7972 for the two runs. In the data generation, we set the SNR to 2.5. To assess the robustness of the proposed Bayesian model, we also consider two additional scenarios with SNR values of 0.5 and 10. Detailed results for these cases are provided in Section S5 of the supplementary material [72]. For model fitting, we set the hyper-parameter values to be $a_\lambda = 3$, $b_\lambda = 1$, $a = 1/H$, $a_r = Ha = 1$, $b_r = H^4$, $\beta_1 = 1$, $\beta_2 = 5$, $a_w = b_w = 2$, $W_\infty = 0.01$, $a_\sigma = 10$, $b_\sigma = 2$, $\theta_{h,\min} = -8$, $\theta_{h,\max} = 8$, $\kappa_{h,\max} = 10$, to encourage sparsity. A total of 10,000 MCMC samples are drawn, one-third of the output is discarded, and the remaining samples are thinned by a factor of 3 to reduce storage and possible auto-correlation of the chains.

First, we compare the proposed method’s performance under different combinations of $P$ and $H$. We summarize the results of the simulation study by investigating the ability of our model to recover the dynamic coefficient matrices $[A_{1,t}, A_{2,t}, \ldots, A_{P,t}]$ in (2). To assess the performance of the proposed method, we employ the root mean square Frobenius distance between the MCMC estimates of the posterior means $E(A_{j,t} | y_1, \ldots, y_T)$ at each $t$, say $\{\tilde{A}_{j,t}\}_{j=1,\ldots,P; t=P+1,\ldots,T}$, and the true matrices defined by

$$\text{err}(\{\tilde{A}_{j,t}\}) = \sqrt{\frac{\sum_{t=P+1}^{T} \sum_{j=1}^{P} \|\tilde{A}_{j,t} - A_{j,t}\|_F^2}{(T-P) \times N^2 P}}.$$ (6)

We also compare the obtained Bayesian point estimates with those provided by a frequentist time-varying vector auto-regressive model, as implemented in the R package TvReg [7] and a mixed VAR model in the R package mgm [25]. The results for the SNR= 2.5 are shown in Table 1. When $P$ and $H$ increase, performance usually improves. But if $P$ and $H$ are higher than the true values, performance doesn’t improve significantly. For instance, the estimation errors of all entries in the time-varying coefficient matrices $A_{j,t}$ stay between 0.0263 and 0.0317 for different combinations of $P \geq 3$, $H \geq 4$. This observation showcases the effectiveness of the sparsity-inducing priors applied to the tensor margins $\alpha_{1,h}$, $\alpha_{2,h}$, and $\alpha_{3,h}$. The BTVT-VAR model appears to provide an improved estimation of the true dynamic structure of the data with respect to the non-sparse frequentist VAR.
In comparison to mixed VAR estimates obtained through elastic-net regularized neighborhood regression, our method performs better in terms of recovering non-zero entries while maintaining sufficient sparsity over true zero entries. Table 1 also shows the point estimates of the base matrices $[A^*_{1,h}, A^*_{2,h}, \ldots, A^*_{P,h}]$. The error of the MCMC-based estimates of the posterior means say $\{\tilde{A}^*_j\}_{j=1,P}$, is similarly defined as

$$\text{err}(\{\tilde{A}^*_j\}) = \sqrt{\frac{\sum_{j=1}^P \|\tilde{A}^*_j - A^*_j\|^2_F}{N^2 P}}.$$  \hspace{1cm} (7)

In order to compute (7), we first need to address label switching [55], which may affect the estimation of the components in (2). Thus, we develop a procedure to match the posterior means of each component to their true counterparts. Additionally, if the assumed value of $H$ exceeds the true value, some posterior mean estimates $\tilde{A}_j, h^* j = 1, \ldots, P$ might contain redundant elements very close to zero. To identify these essentially “empty” components, we consider the maximum norm $\max_{j=1,\ldots,P} \|\tilde{A}_j, h^*\|_\infty$. If this norm is below a specified threshold, we set the component to 0 and exclude it from further analysis. In the following, we set such threshold to 0.01. Next, to match the remaining posterior mean estimates $\{\tilde{A}^*_j\}_{j=1,\ldots,P}$ with the true components, we rank them based on ascending values of the Frobenius distances. The simulation results of the components indicate that it is sufficient to set $P$ and $H$ to sufficiently large values and let the shrinkage priors automatically select the most suitable $P$ and $H$.

For the inference on the latent binary indicators $(\gamma_{h,P+1}, \ldots, \gamma_{h,T}), h = 1, \ldots, H$, we threshold the estimated posterior probability of activation at each time point for each data set to identify the activated $\gamma_{h,t}$'s from the MCMC samples, as described in Section 5. Table 2 shows the average accuracy, sensitivity, specificity and precision across all 30 data sets. The results show that the model is able to reconstruct the components and their dynamic activation reasonably well.

**Simulation 2.** We consider a time series from a larger $N = 40$ TV-VAR model of order $P = 3$ with dynamic coefficients shown in the left panel of Figure 3. These coefficients are combinations of $H = 3$ components. A total of $T = 300$ observations are simulated; the coefficients matrix of the first 200 observations admit a rank-2 tensor decomposition with time varying mixing components whereas the last 100 observations consist of only one component. The covariance matrix $\Sigma$ of the error term $\epsilon_t$ has diagonal elements $\Sigma[i,i] = i/5$ for $i = 1, \ldots, 25$ and $\Sigma[i,i] = (51 - k)/5$ for $i = 26, \ldots, 40$. For the posterior inference, we run the algorithm for 9,000 iterations, with swap being performed every 300 iterations. For more details on the algorithm, refer to Section 7 and the supplementary material [72]. We choose $H = 4$ and $P = 4$. Even though we include one more component and lag, we expect the extra component as well as the coefficients at lag 4 to be
Table 1: Simulation study 1; SNR = 2.5. Mean and standard deviation (in parentheses) of estimation errors over 30 iterations for posterior means of tensor components $[A^*_{j,h}]$ and dynamic coefficients $[A_{j,t}]$ in the BTVT-VAR model. Compared with frequentist estimates from the \texttt{tvReg} R-package [7] and mixed models from the \texttt{mgm} R-package [25]. See Section 6 for details.

| SNR=2.5 | Latent $A^*_{j,h}$ | VAR Matrices $A_{j,t}$ |
|---------|---------------------|------------------------|
|         | All entries | True non-zeroes | True zeroes | All entries | True non-zeroes | True zeroes |
| P=2 H=3 | 0.0405 (0.0308) | 0.1977 (0.1641) | 0.0170 (0.0314) | 0.0541 (0.0206) | 0.1971 (0.0813) | 0.0216 (0.0153) |
| P=2 H=4 | 0.0348 (0.0360) | 0.1567 (0.1366) | 0.0148 (0.0263) | 0.0475 (0.0165) | 0.1715 (0.0616) | 0.0187 (0.0071) |
| P=2 H=5 | 0.0373 (0.0384) | 0.1651 (0.1519) | 0.0160 (0.0277) | 0.0485 (0.0226) | 0.1737 (0.0832) | 0.0209 (0.0123) |
| P=3 H=3 | 0.0317 (0.0450) | 0.1522 (0.1796) | 0.0137 (0.0380) | 0.0450 (0.0231) | 0.1647 (0.0846) | 0.0194 (0.0109) |
| P=3 H=4 | 0.0190 (0.0300) | 0.0856 (0.1384) | 0.0091 (0.0169) | 0.0317 (0.0233) | 0.1027 (0.0846) | 0.0173 (0.0095) |
| P=3 H=5 | 0.0149 (0.0237) | 0.0622 (0.1013) | 0.0084 (0.0127) | 0.0275 (0.0176) | 0.0875 (0.0728) | 0.0161 (0.0054) |
| P=4 H=3 | 0.0295 (0.0348) | 0.1496 (0.1765) | 0.0120 (0.0248) | 0.0446 (0.0220) | 0.1575 (0.0798) | 0.0199 (0.0106) |
| P=4 H=4 | 0.0188 (0.0288) | 0.0844 (0.1299) | 0.0098 (0.0185) | 0.0308 (0.0193) | 0.1018 (0.0851) | 0.0179 (0.0084) |
| P=4 H=5 | 0.0172 (0.0272) | 0.0722 (0.1176) | 0.0104 (0.0198) | 0.0285 (0.0161) | 0.0942 (0.0769) | 0.0171 (0.0082) |
| P=5 H=3 | 0.0255 (0.0273) | 0.1418 (0.1654) | 0.0095 (0.0174) | 0.0424 (0.0200) | 0.1499 (0.0762) | 0.0193 (0.0109) |
| P=5 H=4 | 0.0159 (0.0227) | 0.0678 (0.1022) | 0.0100 (0.0168) | 0.0263 (0.0119) | 0.0832 (0.0485) | 0.0159 (0.0050) |
| P=5 H=5 | 0.0188 (0.0272) | 0.0769 (0.1144) | 0.0102 (0.0174) | 0.0315 (0.0202) | 0.1067 (0.0927) | 0.0184 (0.0095) |

| SNR=2.5 | \texttt{tvReg} | \texttt{mgm} |
|---------|----------------|-------------|
|         | VAR Matrices $A_{j,t}$ | VAR Matrices $A_{j,t}$ |
|         | All entries | True non-zeroes | True zeroes | All entries | True non-zeroes | True zeroes |
| P=2     | 0.0922 (0.0106) | 0.2801 (0.0514) | 0.0610 (0.0067) | 0.1150 (0.0160) | 0.4479 (0.0386) | 0.0124 (0.0039) |
| P=3     | 0.1070 (0.0116) | 0.3205 (0.0497) | 0.0718 (0.0081) | 0.1165 (0.0161) | 0.4539 (0.0374) | 0.0121 (0.0039) |
| P=4     | 0.1163 (0.0108) | 0.3328 (0.0456) | 0.0825 (0.0073) | 0.1170 (0.0162) | 0.4560 (0.0374) | 0.0118 (0.0033) |
| P=5     | 0.1255 (0.0111) | 0.3402 (0.0444) | 0.0940 (0.0084) | 0.1172 (0.0161) | 0.4571 (0.0374) | 0.0114 (0.0036) |
Table 2: Simulation study 1; SNR = 2.5. Performance evaluation of the posterior estimation of the components’ indicators \((\gamma_{h, P+1}, \ldots, \gamma_{h, T})\), based on average accuracy, sensitivity, specificity and precision across the 30 generated data sets. Standard deviations are indicated in parentheses. See Section 6 for details.

| SNR=2.5 | \((\gamma_{h, P+1}, \ldots, \gamma_{h, T})\) | SNR=2.5 | \((\gamma_{h, P+1}, \ldots, \gamma_{h, T})\) |
|---|---|---|---|
| P=2 H=3 | 0.8669 (0.2110) 0.7069 (0.4334) 0.9812 (0.0933) 0.9579 (0.1499) | P=4 H=3 | 0.8606 (0.2236) 0.7160 (0.4450) 0.9826 (0.0938) 0.9672 (0.1465) |
| P=2 H=4 | 0.9214 (0.1750) 0.8671 (0.3096) 0.9553 (0.2137) 0.9308 (0.2069) | P=4 H=4 | 0.9464 (0.1453) 0.8740 (0.3192) 0.9884 (0.0930) 0.9716 (0.1539) |
| P=2 H=5 | 0.9767 (0.1465) 0.8698 (0.3195) 0.9437 (0.1990) 0.9370 (0.1948) | P=4 H=5 | 0.9516 (0.1355) 0.9064 (0.2769) 0.9679 (0.1438) 0.9623 (0.1536) |
| P=3 H=3 | 0.8731 (0.2123) 0.7198 (0.4451) 0.9751 (0.1373) 0.9722 (0.1322) | P=5 H=3 | 0.8704 (0.2319) 0.7271 (0.4398) 0.9926 (0.0330) 0.9761 (0.1241) |
| P=3 H=4 | 0.9549 (0.1227) 0.8874 (0.3066) 0.9783 (0.1329) 0.9813 (0.0869) | P=5 H=4 | 0.9566 (0.1396) 0.9237 (0.2512) 0.9765 (0.1381) 0.9735 (0.1357) |
| P=3 H=5 | 0.9704 (0.1025) 0.9276 (0.2502) 0.9856 (0.0776) 0.9697 (0.1585) | P=5 H=5 | 0.9548 (0.1331) 0.9018 (0.2889) 0.9732 (0.1415) 0.9621 (0.1724) |

Table 3: Simulation study 2. Posterior means of estimated tensor components \(A^*_{j,h}\) and coefficients \(A_{j,t}\) from the BVT-VAR model, compared with frequentist estimates from the \texttt{tvReg} and \texttt{mgm} R-packages. Tensor components are evaluated using the square root of the average Frobenius norm of the difference between the posterior mean and true components. Columns 2 and 3 show average Euclidean distances for truly non-zero and zero entries, with standard deviations in parentheses. See Section 6 for details.

| | All entries | True non-zeroes | True zeroes |
|---|---|---|---|
| BVT-VAR | Latent \(A^*_{j,h}\) | 0.0141 (0.0016) | 0.0504 (0.0369) | 0.0114 (0.0032) |
| | VAR Matrices \(A_{j,h}\) | 0.0188 | 0.0365 | 0.0158 |
| TvReg | VAR Matrices \(A_{j,t}\) | 0.3798 | 0.2717 | 0.3895 |
| MGM | VAR Matrices \(A_{j,t}\) | 0.1072 | 0.3459 | 0.0090 |
Table 4: Simulation study 2. Performance evaluation of the posterior estimation of the components’ indicators \((\gamma_{h,P+1}, \ldots, \gamma_{h,T})\). The evaluation is based on average accuracy, sensitivity, specificity and precision over the true components.

| \((\gamma_{h,P+1}, \ldots, \gamma_{h,T})\) | Acc | Sens | Spec | Prec |
|-----------------------------------------|-----|------|------|------|
|                                         | 0.9955 | 1.0000 | 0.9911 | 0.9911 |
|                                         | (0.0078) | (0.0000) | (0.0154) | (0.0154) |

close to zero due to the effect of the shrinkage priors. The remaining experiment settings are the same as in the first simulation. We select four estimated coefficients matrices at time point 50, 125, 175 and 250 as in the right panel of Figure 3. Our method identifies three “non-empty” components and it accurately captures the patterns of the true ones. Furthermore, the dynamics of the coefficient matrices are all accurately identified by the model. Table 3 shows an evaluation of the posterior inference on the coefficient matrices in our model versus a frequentist time-varying regression. Once again, our model compares quite favorably. To further verify our method’s ability to detect changing patterns along all the 296 time points, we report the accuracy, sensitivity, specificity and precision in the estimation of the latent activation indicators \(\gamma_{h,t}\)’s in Table 4. In Figure 4 we report the estimated trajectories of the \(\gamma_{h,t}\) along time, for each component \(h = 1, \ldots, H\). The shaded red areas indicate the true component activation, whereas the solid line indicates the posterior modes. One of the trajectories, being constantly zero, identifies an empty component: indeed, we employed \(H = 4\) for model fitting instead of the true number of components. The other three estimated trajectories follow the true activations quite closely, reaching false positive rates of 0%, 2.67% and 0% as well as false negative rate of 0%, 0% and 0%, respectively. Once again, the results illustrate the role of the shrinkage prior specifications, since fixing higher values of \(H\) and \(P\) does not appear to hamper the estimation of the VAR matrices. In particular, we do not need to rely on model selection techniques in order to determine the appropriate values of \(H\) and \(P\). Therefore, in many cases it may be desirable to learn the actual dimensions of the model from the data, by fixing relatively large values of \(H\) and \(P\).

As a final remark, we note that we also considered a simulation scenario with \(N = 100\) multivariate time series. However, the frequentist TvReg approach did show numerical problems with such a large number of times series, after taking 9.44 hours to complete. On the contrary, our method was still able to obtain good inferences for this large dimension case, after taking 3.6 hours to complete.
Figure 3: Comparison between the true BTVT-VAR coefficients $A_{1,t}, A_{2,t}, A_{3,t}$ on the left panel and the MCMC posterior means from the BTV-TVAR model on the right panel in the second simulation study of Section 6. For each truly invariant time window, the estimated matrices at time point 50, 125, 175 and 250 are displayed.

5,000 iterations on a Intel Core i5-6300U CPU at 2.40GHz, with 8GB RAM. For comparison, using our method, each run of the $N = 40$-dimensional case took approximately 56 minutes to complete.

7 Time-varying Effective Connectivity in an fMRI reading experiment

We apply our BTVT-VAR model to a task-based functional magnetic resonance imaging (fMRI) data set [66]. The experiment involved participants aged 18 to 40, who were tasked with reading Chapter 9 of *Harry Potter and the Sorcerer’s Stone* [50]. All subjects had previously read the book or seen the movie. The words of the story were presented in rapid succession, where each word was presented one by one at the center of the screen for 0.5 seconds in black font on a gray background. A Siemens Verio 3.0T scanner was used to acquire the scans, utilizing a T2* sensitive
Figure 4: Estimated trajectories of the latent indicators of activation $\gamma_{h,t}$ for four components $h = 1, \ldots, 4$. The red areas indicate the true activations. The solid lines indicates the estimated values of $\gamma_{h,t}$ based on posterior probabilities of activation greater than 0.5.

echo planar imaging pulse sequence with repetition time (TR) of 2s, time echo (TE) of 29 ms, flip angle (FA) of 79°, 36 number of slices and $3 \times 3 \times 3$ mm$^3$ voxels. Data preprocessing involved realignment, slice timing correction, and co-registration with the subject’s anatomical scan, which was segmented into grey and white matter and cerebro-spinal fluid. The subjects’ scans were then normalized to the Montreal Neurological Institute (MNI) template and smoothed with a $6 \times 6 \times 6$ mm Gaussian kernel smoother. For more details, see [45].

Twenty-seven ROIs defined using the Automated Anatomical Atlas (AAL) brain atlas [59] were extracted from the data set, shown in Table 5. Figure 5, created using R package fabisearch by Ondrus and Cribben [43, 44], shows the locations of the ROIs. These regions contain a variety of voxels that have been previously recognized as important to distinguish between the literary content of two novel text passages based on neural activity while these passages are being read [68].

We expect that the dynamics of effective connectivity inferred through our BTVT-VAR model coincide with different events in the plot. To illustrate, we focus on data from a representative subject and on a time point which includes a thrilling moment in the story: Harry, Ron, and Hermione (the main three characters in the book) arrive in a forbidden corridor, turn around, and come face-to-face with a monstrous three-headed dog. This occurs at approximately time point $t = 1176$ in the experiment. While we are in possession of the event information, our model does not utilize such information. For model fitting, we set $H = 8$, allowing us to capture over 50% of the estimated variability in the sample, as determined by the Frobenious norm of the coefficient matrices estimated in a frequentist multivariate autoregressive integrated moving average (MARIMA) model. Additionally, we set $P = 4$. In applications of VAR models to fMRI data, it’s common to consider only
an AR(1) structure, which is seen as an effective representation of short-range temporal dependencies over a small number of ROIs, typically fewer than ten.

First, we challenge the model to identify possible changes in the effective connectivity that we are not aware of in advance and connect them to events in the story plot. As an example, Figure 6 shows the estimated trajectories of selected coefficients (from ROI F.R to ROI ST.L on the left and from ROI F.R to ROI ST.R on the right) across several time points, with 80% credible intervals, of the representative subject. As explained previously, an important advantage of the BTVT-VAR model is its ability to estimate dynamic effective connectivity at each time point. The two coefficients exhibit similar time-varying patterns: they show large positive values before the encounter with the three-headed monsters and become less active thereafter. Despite the slight change in value, there is still an evident increase in dynamic effective connectivity between the ROIs near the pivotal plot event at time point $t = 1176$. After estimating the $A_t$ coefficients, it is possible to calculate the Frobenius distance $\Delta A_t = \|A_t - A_{t-1}\|_F$ between $A_t$ and $A_{t-1}$ as an indicator of the change in effective connectivity at time $t$. Section S6 in the Supplementary material presents these distances for the selected subject and the other participants in the study, emphasizing the changes in time-varying coefficients, particularly in the vicinity of the pivotal point of the story [72]. Section S7 in the supplementary material further investigates the behavior of the lag-1 mean coefficients for subjects who exhibit more pronounced shifts in their brain
Table 5: Information on the 27 ROIs extracted from the Harry Potter task-based fMRI data set. Each ROI has a left and right hemisphere component apart from the supramarginal gyrus region.

| ROI id | Regions                                      | Label |
|--------|----------------------------------------------|-------|
| 1      | Angular gyrus                                | AG    |
| 2      | Fusiform gyrus                               | F     |
| 3      | Inferior temporal gyrus                      | IT    |
| 4      | Inferior frontal gyrus, opercular part       | IFG 1 |
| 5      | Inferior frontal gyrus, orbital part         | IFG 2 |
| 6      | Inferior frontal gyrus, triangular part      | IFG 3 |
| 7      | Middle temporal gyrus                        | MT    |
| 8      | Occipital lobe                               | O     |
| 9      | Precental gyrus                              | PCG   |
| 10     | Precuneus                                    | PC    |
| 11     | Supplementary motor area                     | SM    |
| 12     | Superior temporal gyrus                      | ST    |
| 13     | Temporal pole                                | TP    |
| 14     | Supramarginal gyrus                          | SG.R  |
dynamic effective connectivity patterns before and after the time point \( t = 1176 \).

In the remainder of this manuscript, we explore the estimation of a Granger causality network, based on the estimated VAR coefficient matrices. Granger causality is a mainstream tool to infer and understand effective brain connectivity [17]. Informally, Granger causality states that if including past values of \( y_{i,t} \), the \( i \)th entry of the observed vector \( y_t \), improves the prediction of \( y_{j,t} \), with respect to only including past values of \( y_{j,t} \), then the \( i \)th entry Granger causes the \( j \)th entry at time \( t \) [22]. It is straightforward to infer Granger causality from the elements of the VAR coefficient matrices: \( A_{p,t}[i,j] = 0 \) for all \( p = 1, \ldots, P \) if and only if, at time \( t \), the \( i \)-th entry does not cause the \( j \)-th entry in a Granger sense. In order to estimate the Granger causalities at each time point, we estimate the probability that the \( i \)th ROI does not Granger cause the \( j \)th ROI by computing

\[
p_{j,i}(t) \equiv P( |A_{p,t}[j,i]| < \delta \text{ for all } p = 1, \ldots, P | y_{1:T})
\]

for some small \( \delta \). This probability can be easily approximated using the MCMC samples.

The analysis of Granger causality networks enables us to create a high-resolution map of time-varying connections. To illustrate, Figure 7 shows the Granger causality networks for the representative subject, as inferred from the BTVT-VAR coefficient matrix estimated at two time points close to the focal point of the story, that is, \( t = 1174 \) and \( t = 1175 \). We focus on the first 100 Granger edges \( (j, i) \) with the highest posterior probabilities \( 1 - p_{j,i}(t) \). Figure a shows edges that are present in both the Granger networks at time point 1174 and 1175. Here the Granger network is sparse and is concentrated around two hub nodes, the right precentral gyrus (PCG.R) and both the left and right precuneus gyrus (PC.L and
Figure 7: Granger causality inferred from the BTVT-VAR coefficient matrix for the representative subject. a shows edges that are present in both the Granger networks at time point 1174 and 1175, b are edges unique at time point 1174 and c are edges unique at time point 1175.

PC.R). The primary motor cortex, which is located in the PCG, is usually active during any task that incorporates movement and has been associated with reading out loud [54]. The PC is involved in a variety of complex functions, which include recollection and memory, integration of information relating to perception of the environment, cue reactivity, mental imagery strategies, episodic memory retrieval [6]. Figure b and c show the unique edges at time points 1174 and 1175. Hence, they reflect the brain effective connectivity dynamics. In Figure b, the right fusiform gyrus (F.R) is a hub node with many connections arising from F.R to several other ROIs. The fusiform gyrus plays important roles in object and face recognition and the visual processing of written words. All of the inferior frontal gyrus areas also have many directed Granger causality connections to and from several other ROIs. The left inferior frontal gyrus has a number of functions including the processing of speech and language in Broca’s area [29]. In Figure c, the Granger causality network exhibits increased dispersion and density. Many regions of interest (ROIs) are linked to the left inferior frontal gyrus, opercular part (IFG.1.L), while the left middle temporal gyrus (MT.L) shows connections both from and to ROIs. The middle temporal gyrus is known for its sensitivity to visual motion. Traditional language processing areas encompass the inferior frontal gyrus (Broca’s area), superior temporal and middle temporal gyri, supramarginal gyrus, and angular gyrus (Wernicke’s area). Recent evidence suggests that structures in the medial temporal lobe also play a role in language processing [58].
8 Discussion

We have introduced a scalable Bayesian time-varying tensor VAR model with a few distinctive features. The model relies on a state-space representation comprising $2^H$ elements. In this representation, each state is derived from a subset of $H$ components shared across the entire time span. The activations of these components at each time point are governed by a latent Ising model, while shrinkage priors enforce sparsity of the tensor representation. These features contribute to a more natural description of connectivity dynamics for fMRI data. Moreover, this representation aligns with the emerging belief within neuroscience that connectivity states evolve gradually, rather than undergoing abrupt changes over time as suggested by traditional models based on hidden Markov models. As a result, our model allows for a finer resolution of time-varying functional connectivity. We can estimate time-varying Granger causality networks and investigate changes in connectivity across subsequent time points.

In our study, we demonstrate the model’s efficacy and interpretability in analyzing effective connectivity in fMRI experiments. Our method extends beyond fMRI to encompass other types of time-varying neuroimaging data, including EEG data, and more generally to any data where a vector autoregressive model is appropriate. In fMRI data, where an AR(1) dependence is often assumed as sufficient [however, see 40, 9, for different perspectives], our data analysis appears to confirm this general suggestion. Higher lags of the coefficient matrices did not exhibit patterns significantly different from zero in our fMRI experiment.

The proposed framework can be extended to obtain scalable inference and describe shared patterns of brain connectivity in multi-subject analyses, by tracking the components active across multiple time intervals and subjects. However, this type of inference would require ensuring the identifiability of the same tensor components across subjects. One approach to achieving this could involve utilizing clustering-inducing Bayesian nonparametric priors, facilitating the borrowing of information across all subjects in estimating the components. Due to the increased computational burden and development required, we defer its exploration to future work.

References

[1] Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, 20(3):407–415.

[2] Anastasiou, A., Cribben, I., and Fryzlewicz, P. (2022). Cross-covariance isolate
detect: A new change-point method for estimating dynamic functional connectivity. *Medical Image Analysis*, 75:102252. 3

[3] Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *eLife*, 3:e01867. 6

[4] Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics*, 43(4):1535–1567. 3

[5] Billio, M., Casarin, R., Iacopini, M., and Kaufmann, S. (2023). Bayesian dynamic tensor regression. *Journal of Business & Economic Statistics*, 41(2):429–439. 3, 4

[6] Borsook, D., Maleki, N., and Burstein, R. (2015). Migraine. In Zigmond, M. J., Rowland, L. P., and Coyle, J. T., editors, *Neurobiology of Brain Disorders*, chapter 42, pages 693–708. Academic Press, San Diego. 26

[7] Casas, I. and Fernández-Casal, R. (2021). *tvReg*: Time-varying coefficients linear regression for single and multi-equations. R package version 0.5.4. 16, 18

[8] Chiang, S., Guindani, M., Yeh, H. J., Haneef, Z., Stern, J. M., and Vannucci, M. (2017). Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping*, 38(3):1311–1332. 3

[9] Corbin, N., Todd, N., Friston, K. J., and Callaghan, M. F. (2018). Accurate modeling of temporal correlations in rapidly sampled fMRI time series. *Human Brain Mapping*, 39(10):3884–3897. 27

[10] Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4):907–920. 3

[11] Dai, B., Ding, S., Wahba, G., et al. (2013). Multivariate Bernoulli distribution. *Bernoulli*, 19(4):1465–1483. 9, 10

[12] Davis, R. A., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25(4):1077–1096. 13

[13] Durante, D. and Guindani, M. (2020). Bayesian methods in brain networks. In Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W., Ruggeri, F., and Teugels, J., editors, *Wiley StatsRef: Statistics Reference Online*. Wiley. 2
[14] Fan, J., Sitek, K., Chandrasekaranb, B., and Sarkara, A. (2022). Bayesian tensor factorized mixed vector autoregressive models for inferring Granger causality patterns from high-dimensional multi-subject panel neuroimaging data. *arXiv preprint arXiv:2206.10757*. 3

[15] Fiecas, M., Cribben, I., Bahktiari, R., and Cummine, J. (2017). A variance components model for statistical inference on functional connectivity networks. *NeuroImage*, 149:256–266. 2

[16] Finn, E. S., Shen, X., Holahan, J. M., Scheinost, D., Lacadie, C., Papademetris, X., Shaywitz, S. E., Shaywitz, B. A., and Constable, R. T. (2014). Disruption of functional networks in dyslexia: a whole-brain, data-driven analysis of connectivity. *Biological psychiatry*, 76(5):397–404. 2

[17] Friston, K., Moran, R., and Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, 23(2):172–178. 25

[18] Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36. 2

[19] George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889. 13

[20] Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451. 3

[21] Gorrostieta, C., Fiecas, M., Ombao, H., Burke, E., and Cramer, S. (2013). Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Frontiers in Computational Neuroscience*, 7:159. 3

[22] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438. 25

[23] Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *The Journal of Machine Learning Research*, 18(1):2733–2763. 4, 7, 12

[24] Haslbeck, J. M., Bringmann, L. F., and Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, 56(1):120–149. 16
[25] Haslbeck, J. M. and Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, 93:1–46. 16, 18

[26] Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378. 3

[27] Jacobs, P. A. and Lewis, P. A. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36. 4, 9

[28] Jentsch, C. and Reichmann, L. (2019). Generalized binary time series models. *Econometrics*, 7(4):47. 9

[29] Kennison, S. M. (2013). *Introduction to Language Development*. Sage Publications. 26

[30] Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203. 3

[31] Legramanti, S., Durante, D., and Dunson, D. B. (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika*, 107(3):745–752. 4, 12

[32] Li, F. and Zhang, N. R. (2010). Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association*, 105(491):1202–1214. 11

[33] Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L., and Coan, J. A. (2015). Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *The Annals of Applied Statistics*, 9(2):687–713. 11

[34] Li, H., Wang, Y., Yan, G., Sun, Y., Tanabe, S., Liu, C.-C., Quigg, M., and Zhang, T. (2021). A Bayesian state-space approach to mapping directional brain networks. *Journal of the American Statistical Association*, 116(536):1637–1647. 3

[35] Luo, Y. and Griffin, J. E. (2022). Bayesian methods of vector autoregressions with tensor decompositions. *arXiv preprint arXiv:2211.01727*. 3, 13

[36] MacDonald, I. L., Zucchini, W., and Langrock, R. (2016). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall/CRC, New York, 2nd edition. 9
[37] Malsiner-Walli, G., Frühwirth-Schnatter, S., and Grün, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing*, 26(1):303–324. 13

[38] McCormick, E. M., Arnemann, K. L., Ito, T., Hanson, S. J., and Cole, M. W. (2022). Latent functional connectivity underlying multiple brain states. *Network Neuroscience*, 6(2):570–590. 6

[39] Möller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458. 14

[40] Monti, M. (2011). Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, 5:1–13. 27

[41] Ofori-Boateng, D., Gel, Y. R., and Cribben, I. (2020). Nonparametric anomaly detection on time series of graphs. *Journal of Computational and Graphical Statistics*, 30:1–26. 3

[42] Ombao, H., Fiecas, M., Ting, C.-M., and Low, Y. F. (2018). Statistical models for brain signals with properties that evolve across trials. *NeuroImage*, 180:609–618. 3

[43] Ondrus, M. and Cribben, I. (2023). FABISEARCH: Change point detection in high-dimensional time series networks. R package version 0.0.4.5. 22

[44] Ondrus, M. and Cribben, I. (2024). FABISEARCH: A package for change point detection in and visualization of the network structure of multivariate high-dimensional time series in R. *Neurocomputing*, page 127321. 22

[45] Ondrus, M., Olds, E., and Cribben, I. (2024). Factorized binary search: change point detection in the network structure of multivariate high-dimensional time series. *arXiv preprint arXiv:2103.06347*. 22

[46] Ou, J., Xie, L., Jin, C., Li, X., Zhu, D., Jiang, R., Chen, Y., Zhang, J., Li, L., and Liu, T. (2014). Characterizing and differentiating brain state dynamics via hidden Markov models. *Brain Topogr.*, 28(5):666–679. 6

[47] Park, H.-J., Friston, K. J., Pae, C., Park, B., and Razi, A. (2018). Dynamic effective connectivity in resting state fMRI. *NeuroImage*, 180:594–608. 3

[48] Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852. 3
[49] Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710. 13

[50] Rowling, J. (2012). *Harry Potter and the Sorcerer’s Stone*. Pottermore Limited. 21

[51] Samdin, S. B., Ting, C.-M., Ombao, H., and Salleh, S.-H. (2016). A unified estimation framework for state-related changes in effective brain connectivity. *IEEE Transactions on Biomedical Engineering*, 64(4):844–858. 3

[52] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650. 13

[53] Shojaie, A. and Michailidis, G. (2010). Discovering graphical granger causality using the truncating lasso penalty. *Bioinformatics*, 26(18):i517–i523. 3

[54] Smelser, N. J., Baltes, P. B., et al. (2001). *International Encyclopedia of the Social & Behavioral Sciences*, volume 11. Elsevier Amsterdam. 26

[55] Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809. 17

[56] Sun, W. W. and Li, L. (2019). Dynamic tensor clustering. *Journal of the American Statistical Association*, 114(528):1894–1907. 8

[57] Taghia, J., Ryali, S., Chen, T., Supekar, K., Cai, W., and Menon, V. (2017). Bayesian switching factor analysis for estimating time-varying functional connectivity in fMRI. *Neuroimage*, 155:271–290. 3

[58] Tracy, J. I. and Boswell, S. B. (2008). Mesial temporal lobe epilepsy: a model for understanding the relationship between language and memory. In *Handbook of the Neuroscience of Language*, pages 319–328. Elsevier. 26

[59] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289. 22

[60] Velu, R. P., Reinsel, G. C., and Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika*, 73(1):105–118. 3
[61] Vidaurre, D., Smith, S. M., and Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832. 6

[62] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305. 9

[63] Wang, D., Zheng, Y., Lian, H., and Li, G. (2021). High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association*, pages 1–19. 3, 4, 13

[64] Wang, Y., Ting, C.-M., and Ombao, H. (2016). Modeling effective connectivity in high-dimensional cortical source signals. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1315–1325. 13

[65] Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., and Vannucci, M. (2018). A Bayesian approach for estimating dynamic functional network connectivity in fMRI data. *Journal of the American Statistical Association*, 113(521):134–151. 3, 6

[66] Welvaert, M. and Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PloS one*, 9(11):e112575. 21

[67] Welvaert, M. and Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PloS one*, 8(11):e77089. 16

[68] Xiong, X. and Cribben, I. (2023). Beyond linear dynamic functional connectivity: A vine copula change point model. *Journal of Computational and Graphical Statistics*, 32(3):853–872. 22

[69] Zarghami, T. S. and Friston, K. J. (2020). Dynamic effective connectivity. *NeuroImage*, 207:116453. 3

[70] Zhang, L., Guindani, M., and Vannucci, M. (2015). Bayesian models for functional magnetic resonance imaging data analysis. *WIREs Computational Statistics*, 7:21–41. 2

[71] Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage*, 95:162–175. 16
[72] Zhang, W., Cribben, I., Petrone, S., and Guindani, M. (2024). Supplement to "bayesian time-varying tensor vector autoregressive models for dynamic effective connectivity". Technical report. 7, 10, 15, 16, 17, 23, 25

[73] Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552. 7, 13