Bayesian Dynamic Tensor Regression*

Monica Billio[†1], Roberto Casarin[‡1], Sylvia Kaufmann[§2], Matteo Iacopini[¶1,3]

[1]Ca' Foscari University of Venice
[2]Study Center Gerzensee, Foundation of the Swiss National Bank
[3]Université Paris I - Panthéon-Sorbonne

May 8, 2018

Abstract

Multidimensional arrays (i.e. tensors) of data are becoming increasingly available and call for suitable econometric tools. We propose a new dynamic linear regression model for tensor-valued response variables and covariates that encompasses some well known multivariate models such as SUR, VAR, VECM, panel VAR and matrix regression models as special cases. For dealing with the over-parametrization and over-fitting issues due to the curse of dimensionality, we exploit a suitable parametrization based on the parallel factor (PARAFAC) decomposition which enables to achieve both parameter parsimony and to incorporate sparsity effects. Our contribution is twofold: first, we provide an extension of multivariate econometric models to account for both tensor-variate response and covariates; second, we show the effectiveness of proposed methodology in defining an autoregressive process for time-varying real economic networks. Inference is carried out in the Bayesian framework combined with Monte Carlo Markov Chain (MCMC). We show the efficiency of the MCMC procedure on simulated datasets, with different size of the response and independent variables, proving computational efficiency even with high-dimensions of the parameter space. Finally, we apply the model for studying the temporal evolution of real economic networks.

*We thank the conference, workshop and seminar participants at: "CFENetwork 2017" in London, "ESOBE 2016" in Venice, "CFENetwork 2016" in Seville, "ICEEE 2017" in Messina of the Italian Econometric Society, the "3rd Vienna Workshop on High-dimensional Time Series in Macroeconomics and Finance", "BISP10" at Bocconi University of Milan and the "SIS Intermediate Meeting" of the Italian Statistical Society in Florence. Iacopini's research is supported by the "Bando Vinci 2016" grant by the Université Franco-Italienne. This research used the SCSCF multiprocessor cluster system at Ca' Foscari University of Venice.

†e-mail: billio@unive.it

‡e-mail: r.casarin@unive.it

§e-mail: sylvia.kaufmann@szgerzensee.ch

¶e-mail: matteo.iacopini@unive.it

1 Introduction

The increasing availability of large sets of time series data with complex structure, such as EEG (e.g., Li and Zhang (2017)), neuroimaging (e.g., Zhou et al. (2013)), two or multidimensional tables (e.g., ,Balazsi et al. (2015), Carvalho and West (2007)), multilayer networks (e.g., Aldasoro and Alves (2016), Poledna et al. (2015)) has put forward some limitations of the existing multivariate econometric models. In the era of "*Big Data*", mathematical representations of information in terms of vectors and matrices have some non-negligible drawbacks, the most remarkable being the difficulty of accounting for the structure of the data, their nature and the way they are collected (e.g., contiguous pixels in an image, cells of matrix representing a geographical map). As such, if this information is neglected in the modelling the econometric analysis might provide misleading results.

When the data are gathered in the form of matrices (i.e. 2-dimensional arrays), or more generally as tensors, that is multi-dimensional arrays, a statistical modelling approach can rely on vectorizing the object of interest by stacking all its elements in a column vector, then resorting to standard multivariate analysis techniques. The vectorization of an array does not preserve the structural information encrypted in its original format. In other words, the physical characteristics of the data (e.g, the number of dimensions and the length of each of them) matter, since a cell is highly likely to depend on a subset of its contiguous cells. Collapsing the data into a 1-dimensional array does not allow to preserve this kind of information, thus making this statistical approach unsuited for modelling tensors. The development of novel methods capable to deal with 2- or multi-dimensional arrays avoiding their vectorization is still an open challenging question in statistics and econometrics.

Many results for 1-dimensional random variables in the exponential families have been extended to the 2-dimensional case (i.e. matrix-variate, see Gupta and Nagar (1999) for a compelling review). Conversely, tensors have been recently introduced in statistics (see Hackbusch (2012), Kroonenberg (2008), Cichocki et al. (2009)), providing the background for more efficient algorithms in high dimensions especially in handling *Big Data* (e.g. Cichocki (2014)). However, a compelling statistical approach to multi-dimensional random objects is lacking and constitutes a promising field of research.

Recently, the availability of 3-dimensional datasets (e.g., medical data) has fostered the use of tensors in many different fields of theoretical and applied statistics. The main purpose of this article is to contribute to this growing literature by proposing an extension of standard multivariate econometric models to tensor-variate response and covariates.

Matrix models in econometrics have been employed over the past decade, especially in time series analysis where they have been widely used for providing a state space representation (see Harrison and West (1999)). However, only recently the attention of the academic community has moved towards the study of this class of models. Within the time series analysis literature, matrix-variate models have been used for defining dynamic linear models (e.g., Carvalho and West (2007) and Wang and West (2009)), whereas Carvalho et al. (2007) exploited Gaussian graphical models for studying matrix-variate time series. In a different context, matrix models have also been used for classification of longitudinal datasets in Viroli (2011) and Viroli and Anderlucci (2013).

Viroli (2012) the author presented a first generalization of the multivariate regression by introducing a matrix-variate regression where both response and covariate are matrices. Ding and Cook (2016) propose a bilinear multiplicative matrix regression model whose vectorisewd form is a VAR(1) with restrictions on the covariance matrix. The main shortcoming in using bilinear models (either in the additive or multiplicative form) is the difficulty in introducing sparsity constrains. Imposing a zero restriction on a subset of the reduced form coefficients implies a zero restriction on the structural coefficients[1]. Ding and Cook (2016) proposed a generalization of the envelope method of Cook et al. (2010) for achieving sparsity and increasing efficiency of the regression. Further studies which have used matrices as either the response or a covariate include Durante and Dunson (2014), who considered tensors and Bayesian nonparametric frameworks and Hung and Wang (2013), who defined a logistic regression model with a matrix-valued covariate.

Following the model specification strategy available in the existing literature, there are two main research streams. In the first one, Zhou et al. (2013), Zhang et al. (2014) and Xu et al. (2013) propose a linear regression models with a real-valued N-order tensor $\mathcal{X}$ of data to explain a one-dimensional response, by means of the scalar product with a tensor of coefficients $\mathcal{B}$ of the same size. More in detail, Zhang et al. (2014) propose a multivariate model with tensor covariate for longitudinal data analysis; whereas Zhou et al. (2013) uses a generalized linear model with exponential link and tensor covariate for analysing image data. Finally, the approach of Xu et al. (2013) exploits a logistic link function with a tensor covariate to predict a binary scalar response.

In the second and more general stream of the literature (e.g., Hoff (2015) and Li and Zhang (2017)) both response and covariate of a regression model are tensor-valued. From a modelling point of view, there are different strategies. Hoff (2015) regresses a N-order array on an array of the same order but with smaller dimensions by exploiting the Tucker product, and follows the Bayesian approach for the estimation. Furthermore, Bayesian nonparametric approaches for models with a tensor covariate have been formulated by Zhao et al. (2013), Zhao et al. (2014) and Imaizumi and Hayashi (2016). They exploited Gaussian processes with a suitable covariance kernel for regressing a scalar on a multidimensional data array. Conversely, Li and Zhang (2017) defines a model where response and covariates are multidimensional arrays of possibly different order, and subsequently uses the envelope method coupled with an iterative maximum likelihood method for inference.

We propose a new dynamic linear regression modelling framework for tensor-valued response and covariates. We show that our framework admits as special cases Bayesian VAR models (Sims and Zha (1998)), Bayesian panel VAR models (proposed by Canova and Ciccarelli (2004), see Canova and Ciccarelli (2013) for a review) and Multivariate Autoregressive models (i.e. MAR, see Carriero et al. (2016)), as well as univariate and matrix regression models. Furthermore, we exploit the PARAFAC decomposition for reducing the number of parameters to estimate, thus making inference on network models feasible.

We also contribute to the empirical analysis of tensor data in two ways. First, we provide an original study of time-varying economic and financial networks and show that our model can be succesfsfully used to carry out forecast and impulse response analysis in this high-dimensional stting. Few attempts have been made to model time-evolving networks (for example, Holme and Saramäki (2012), Kostakos (2009), Barrat et al. (2013), Anacleto and Queen (2017) and references in Holme and Saramäki (2013)), and this field of research, which stems from physics, has focused on providing a representation and a description of temporally evolving graphs. Second, we show how tensor regression con be applied to macroeconomic panel data, where standard vectorized models cannot be used.

The structure of the paper is as follows. Section 2 is devoted to a brief introduction to tensor calculus and to the presentation of the new modelling framework. The details of the estimation procedure are given in Section 3. In Section 4 we test proposed model on simulated datasets and in Section 5 we present some empirical applications.

[1]The phenomenon is worse in the bilinear multiplicative model, given that each reduced form coefficient is given by the product of those in the structural equation.

2 A Tensor Regression Model

We introduce multi-dimensional arrays (i.e. tensors) and some basic notions of tensor algebra which will be used in this paper. Moreover, we present a general tensor regression model and discuss some special cases.

2.1 Tensor Calculus and Decompositions

The use of tensors is well established in physics and mechanics (see Synge and Schild (1969), Adler et al. (1975), Malvern (1986), Lovelock and Rund (1989), Aris (2012) and Abraham et al. (2012)), but very few references can be found in the literature outside these disciplines. For a general introduction to the algebraic properties of tensor spaces we refer to Hackbusch (2012). A noteworthy introduction to tensors and corresponding operations is in Lee and Cichocki (2016), while we make reference to Kolda and Bader (2009) and Cichocki et al. (2009) for a review on tensor decompositions. In the rest of the paper we will use the terms tensor decomposition and tensor representation interchangeably, even though the latter one is more suited to our approach.

A N-order tensor is a N-dimensional array (whose dimensions are also called modes). The number of dimensions is the *order* of the tensor. Vectors and matrices are examples of first- and second-order tensors, respectively, while one may think about a third order tensor as a series of matrices of the same size put one in front of the other one, forming a parallelepiped. In the rest of the paper we will use lower-case letters for scalars, lower-case bold letters for vectors, capital letters for matrices and calligraphic capital letters for tensors. When dealing with matrices, in order to select a column (or row) we adopt the symbol ":". The same convention is used for tensors when considering all elements of a given mode. For example, let $A \in \mathbb{R}^{m \times n}$ be a matrix and $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ an array of order N, then $A_{i,j}$ and $\mathcal{B}_{i_1,\ldots,i_N}$ indicate the (i,j)-th and $(i_1,\ldots,i_N)$-th element of A and $\mathcal{B}$, respectively, and:

(i) $A_{(i,:)}$ is the i-th row of A, $\forall i \in \{1,\ldots,m\}$;

(ii) $A_{(:,j)}$ is the j-th column of A, $\forall j \in \{1,\ldots,n\}$;

(iii) $\mathcal{B}_{(i_1,\ldots,i_{k-1},:,i_{k+1},\ldots,i_N)}$ is the mode-k fiber of $\mathcal{B}$, $\forall k \in \{1,\ldots,N\}$

(iv) $\mathcal{B}_{(i_1,\ldots,i_{k-1},:,:,i_{k+2},\ldots,i_N)}$ is the mode-$k,k+1$ slice of $\mathcal{B}$, $\forall k \in \{1,\ldots,N-1\}$

The mode-k fiber is the equivalent of rows and columns in a matrix, more precisely it is the vector obtained by fixing all but the k-th index of the tensor. Instead, slices (i.e. bi-dimensional fibers of matrices) or generalizations of them, by keeping fixed all but two or more dimensions (or modes) of the tensor.

The *mode-n matricization* (or unfolding), denoted by $\mathbf{X}_{(n)}$, is the operation of transforming a N-dimensional array $\mathcal{X}$ into a matrix. It consists in re-arranging the mode-n fibers of the tensor to be the columns of the matrix $\mathbf{X}_{(n)}$, which has size $I_n \times \bar{I}_{(-n)}$ with $\bar{I}_{(-n)} = \prod_{i \neq n} I_i$. The mode-$n$ matricization of $\mathcal{X}$ maps the $(i_1,\ldots,i_N)$ element of $\mathcal{X}$ to the (i_n, j) element of $\mathbf{X}_{(n)}$, where:

$$j = 1 + \sum_{m \neq n} (i_m - 1) \prod_{p \neq n}^{m-1} I_p \tag{1}$$

For some numerical examples, see Kolda and Bader (2009) and Appendix A. The mode-1 unfolding is of interest for providing a visual representation of a tensor: for example, when $\mathcal{X}$ be a third-order tensor, its mode-1 unfolding $\mathbf{X}_{(1)}$ is a matrix of size $I_1 \times I_2 I_3$ obtained by horizontally stacking the frontal slices of the tensor. The *vectorization* operator stacks all the elements in direct lexicographic order, forming a vector of length $\bar{I} = \prod_i I_i$. However, notice that other orderings are possible (as for the vectorisation of matrices), since the ordering of the elements is not important as long as it is consistent across the calculations. The mode-n matricization can also be used to vectorize a tensor $\mathcal{X}$, by exploiting this relationship:

$$\operatorname{vec}(\mathcal{X}) = \operatorname{vec}\left(\mathbf{X}_{(1)}\right) , \tag{2}$$

where $\operatorname{vec}\left(\mathbf{X}_{(1)}\right)$ stacks vertically into a vector the columns of he matrix $\mathbf{X}_{(1)}$. Many product operations have been defined for tensors (e.g., see Lee and Cichocki (2016)), but here we constrain ourselves to the operators used in this work. Concerning the basic product operations, the *scalar product* between two tensors $\mathcal{X}, \mathcal{Y}$ of equal order and same dimensions, $I_1, \ldots, I_N$, is defined as:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \ldots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1,\ldots,i_N} \mathcal{Y}_{i_1,\ldots,i_N} = \sum_{i_1,\ldots,i_N} \mathcal{X}_{i_1,\ldots,i_N} \mathcal{Y}_{i_1,\ldots,i_N} . \tag{3}$$

For the ease of notation, we will use the multiple-index summation for indicating the sum over all the corresponding indices.

The *mode-M contracted product* of two tensors $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_M}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ with $I_M = J_M$, denoted $\mathcal{X} \times^M \mathcal{Y}$, yields a tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times \ldots \times I_{M-1} \times J_1 \times \ldots \times J_{N-1}}$ of order $M + N - 2$, with entries:

$$\mathcal{Z}_{i_1,\ldots,i_{M-1},j_1,\ldots,j_{N-1}} = (\mathcal{X} \times^M \mathcal{Y})_{i_1,\ldots,i_{M-1},j_1,\ldots,j_{N-1}} = \sum_{i_M=1}^{I_M} \mathcal{X}_{i_1,\ldots,i_M} \mathcal{Y}_{j_1,\ldots,i_M,\ldots,j_N} . \tag{4}$$

Therefore, it is a generalization of the matrix product. The notation $\times^{1\ldots M}$ is used to denote a sequence of mode-m contracted products, with $m = 1, \ldots, M$.

The *mode-n product* between a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and a matrix $A \in \mathbb{R}^{J \times I_n}$, $1 \leq n \leq N$, is denoted by $\mathcal{X} \bar{\times}_n A$ and yields a tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots, I_{n-i}, J, I_{n+1}, \ldots \times I_N}$ of the same order of $\mathcal{X}$, with the n-th mode's length changed. Each mode-n fiber of the tensor is multiplied by the matrix A, which yields element-wise:

$$\mathcal{Y}_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = (\mathcal{X} \bar{\times}_n A)_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1,\ldots,i_N} A_{j,i_n} . \tag{5}$$

Analogously, the *mode-n product between a tensor and a vector*, i.e. between $\mathcal{X}$ and $\mathbf{v} \in \mathbb{R}^{I_n}$, yields a lower order tensor, since the n-th mode is suppressed as a consequence of the product. It is given, element-wise, by:

$$\mathcal{Y}_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N} = (\mathcal{X} \times_n \mathbf{v})_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1,\ldots,i_n,\ldots,i_N} \mathbf{v}_{i_n} , \tag{6}$$

with $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots, I_{n-i}, I_{n+1}, \ldots \times I_N}$. It is clear that, as for the matrix dot product, the order of the elements in the multiplication matters and both products are not commutative.

The *Hadamard product* $\odot$ is defined in the same usual way as for matrices, i.e. the element-wise multiplication. Formally, for $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$, $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and $\mathcal{Z} \in \mathbb{R}^{I_1^{\times} \ldots \times I_N}$ it holds:

$$\mathcal{Z}_{i_1,\ldots,i_N} = (\mathcal{X} \odot \mathcal{Y})_{i_1,\ldots,i_N} = \mathcal{X}_{i_1,\ldots,i_N} \mathcal{Y}_{i_1,\ldots,i_N} . \tag{7}$$

Finally, let $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_M}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$. The *outer product* $\circ$ of two tensors is the tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_N}$ whose entries are:

$$\mathcal{Z}_{i_1,\ldots,i_M,j_1,\ldots,j_N} = (\mathcal{X} \circ \mathcal{Y})_{i_1,\ldots,i_M,j_1,\ldots,j_N} = \mathcal{X}_{i_1,\ldots,i_M} \mathcal{Y}_{j_1,\ldots,j_N} . \tag{8}$$

For example, the outer product of two vectors is a matrix, while the outer product of two matrices is a fourth order tensor.

Tensor decompositions represent the core of current statistical models dealing with multidimensional variables since many of them allow to represent a tensor as a function of lower dimensional variables, such as matrices of vectors, linked by suitable multidimensional operations. We now define two tensor decompositions, the Tucker and the parallel factor (PARAFAC), which are useful in our applications because the elements of the decomposition are generally low dimensional and easier to handle than the original tensor. Let R be the rank of the tensor $\mathcal{X}$, that is minimum number of rank-1 tensors whose linear combination yields $\mathcal{X}$. A N-order tensor is of rank 1 when it is the outer product of N vectors.

The Tucker decomposition is a higher-order generalization of the Principal Component Analysis (PCA): a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ is decomposed into the product (along the corresponding modes) of a "core" tensor $\mathcal{G} \in \mathbb{R}^{g_1 \times \ldots \times g_N}$ and N factor matrices $A^{(m)} \in \mathbb{R}^{I_m \times J_m}$, $m = 1, \ldots, N$:

$$\mathcal{B} = \mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \ldots \times_N A^{(N)} = \sum_{i_1=1}^{g_1} \sum_{i_2=1}^{g_2} \ldots \sum_{i_N=1}^{g_N} \mathcal{G}_{i_1, i_2, \ldots, i_N} \mathbf{a}_{i_1}^{(1)} \circ \mathbf{a}_{i_2}^{(2)} \circ \ldots \circ \mathbf{a}_{i_N}^{(N)} , \quad (9)$$

where $\mathbf{a}_{i_l}^{(m)} \in \mathbb{R}^{g_m}$ is the m-th column of the matrix $A^{(m)}$. As a result, each entry of the tensor is obtained as:

$$\mathcal{B}_{j_1, \ldots, j_N} = \sum_{i_1=1}^{g_1} \sum_{i_2=1}^{g_2} \ldots \sum_{i_N=1}^{g_N} \mathcal{G}_{i_1, i_2, \ldots, i_N} \cdot A_{i_1, j_1}^{(1)} \cdots A_{i_N, j_N}^{(N)} \quad j_m = 1, \ldots, I_m, \ m = 1, \ldots, N . \quad (10)$$

A special case of the Tucker decomposition, called PARAFAC(R)[2], is obtained when the core tensor is the identity tensor and the factor matrices have all the same number of columns, R. A graphical representation of this decomposition for a third-order tensor is shown in Fig. (1). More precisely, the PARAFAC(R) is a low rank decomposition which represents a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ as a finite sum of R rank-1 tensors obtained as the outer products of N vectors, also called PARAFAC marginals[3] $\boldsymbol{\beta}_j^{(r)} \in \mathbb{R}^{I_j}$, $j = 1, \ldots, N$:

$$\mathcal{B} = \sum_{r=1}^{R} \mathcal{B}_r = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_N^{(r)} . \quad (11)$$

Remark 2.1. *There exists a one-to-one relation between the mode-n product between a tensor and a vector and the vectorisation and matricization operators. Consider a N-order tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ for which is specified a PARAFAC(R) decomposition, a $(N-1)$-order tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots \times I_{N-1}}$ and a vector $\mathbf{x} \in \mathbb{R}^{I_N}$. Then:*

$$\mathcal{Y} = \mathcal{B} \times_N \mathbf{x} \iff \operatorname{vec}(\mathcal{Y}) = \mathbf{B}_{(N)}' \mathbf{x} \iff \operatorname{vec}(\mathcal{Y})' = \mathbf{x}' \mathbf{B}_{(N)} \quad (12)$$

and, denoting $\boldsymbol{\beta}_j^{(r)}$, for $j = 1, \ldots, N$ and $r = 1, \ldots, R$, the marginals of the PARAFAC(R) decomposition of $\mathcal{B}$ we have:

$$\mathbf{B}_{(N)} = \sum_{r=1}^{R} \boldsymbol{\beta}_N^{(r)} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_{N-1}^{(r)} \right)' . \quad (13)$$

[2] See Harshman (1970). Some authors (e.g. Carroll and Chang (1970) and Kiers (2000)) use the term CODE-COMP or CP instead of PARAFAC.

[3] An alternative representation may be used, if all the vectors $\boldsymbol{\beta}_j^r$ are normalized to have unitary length. In this case the weight of each component r is captured by the r-th component of the vector $\boldsymbol{\lambda} \in \mathbb{R}^R$:

$$\mathcal{B} = \sum_{r=1}^{R} \lambda_r \left(\boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_N^{(r)} \right)$$

.

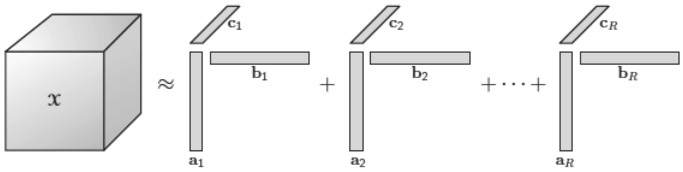

Figure 1: PARAFAC decomposition of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with $\mathbf{a}_r \in \mathbb{R}^{I_1}$, $\mathbf{b}_r \in \mathbb{R}^{I_2}$ and $\mathbf{c}_r \in \mathbb{R}^{I_3}$, $r = 1, \ldots, R$. Figure from Kolda and Bader (2009).

These relations allows to establish a link between operators defined on tensors and operators defined on matrices, for which plenty of properties are known from linear algebra.

Remark 2.2. *For two vectors $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ the following relations hold between the outer product, the Kronecker product $\otimes$ and the vectorisation operator:*

$$\mathbf{u} \otimes \mathbf{v}' = \mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}' \tag{14}$$

$$\mathbf{u} \otimes \mathbf{v} = \mathrm{vec}\left(\mathbf{v} \circ \mathbf{u}\right) . \tag{15}$$

2.2 A General Dynamic Model

The new model we propose, in its most general formulation is:

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{A}_0 + \sum_{j=1}^{p} \mathcal{A}_j \times_{N+1} \mathrm{vec}\left(\mathcal{Y}_{t-j}\right) + \mathcal{B} \times_{N+1} \mathrm{vec}\left(\mathcal{X}_t\right) + \mathcal{C} \times_{N+1} \mathbf{z}_t + \mathcal{D} \times_n W_t + \mathcal{E}_t, \\ \mathcal{E}_t &\overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(0, \Sigma_1, \ldots, \Sigma_N) , \end{aligned} \tag{16}$$

where the tensor response and noise $\mathcal{Y}_t$, $\mathcal{E}_t$ are N-order tensors of sizes $I_1 \times \ldots \times I_N$, while the covariates include a M-order tensor $\mathcal{X}_t$ of sizes $J_1 \times \ldots \times J_M$, a matrix W_t with dimensions $I_n \times K$ and a vector $\mathbf{z}_t$ of length Q.

The coefficients are all tensors of suitable order and sizes: $\mathcal{A}_j$ have dimensions $I_1 \times \ldots \times I_N \times I^*$, with $I^* = \prod_i I_i$, $\mathcal{B}$ has dimensions $I_1 \times \ldots \times I_N \times J^*$, with $J^* = \prod_j J_j$, $\mathcal{C}$ has dimensions $I_1 \times \ldots \times I_N \times Q$ and $\mathcal{D}$ has sizes $I_1 \times \ldots \times I_{n-1} \times K \times I_{n+1} \ldots \times I_N$. The symbol $\times_n$ stands for the mode-n product between a tensor and a vector defined in eq. (6). The reason for the use of tensors coefficients, as opposed to scalars and vectors, is twofold: first, this permits each entry of each covariate to exert a different effect on each entry of the response variable; second, the adoption of tensors allows to exploit the various decompositions, which are fundamental for providing a parsimonious and flexible parametrization of the statistical model.

The noise is assumed to follow a tensor normal distribution (see Ohlson et al. (2013), Manceur and Dutilleul (2013), Arashi (2017)), a generalization of the multivariate normal distribution. Let $\mathcal{X}$ and $\mathcal{M}$ be two N-order tensors of dimensions $I_1, \ldots, I_N$. Define $I^* = \prod_{j=1}^{N} I_j$, $I^*_{-i} = \prod_{j \neq i} I_j$ and let $\times^{1\ldots N}$ be a sequence of mode-j contracted products, $j = 1, \ldots, N$, between the $(K+N)$-order tensor $\mathcal{X}$ and the $(N+M)$-order tensor $\mathcal{Y}$ of conformable dimensions, defined as follows:

$$\left(\mathcal{X} \times^{1\ldots N} \mathcal{Y}\right)_{j_1,\ldots,j_K,h_1,\ldots,h_M} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \mathcal{X}_{j_1,\ldots,j_K,i_1,\ldots,i_N} \mathcal{Y}_{i_N,\ldots,i_1,h_1,\ldots,h_M} . \tag{17}$$

Finally, let $U_j \in \mathbb{R}^{I_j \times I_j}$, $j \in \{1, \ldots, N\}$ be positive definite matrices. The probability density function of a N-order tensor normal distribution with mean array $\mathcal{M}$ and positive definite covariance matrices $U_1, \ldots, U_N$, is given by:

$$f_{\mathcal{X}}(\mathcal{X}) = (2\pi)^{-\frac{d^*}{2}} \prod_{j=1}^{N} |U_j|^{-\frac{I^*_{-j}}{2}} \exp\left\{-\frac{1}{2}(\mathcal{X} - \mathcal{M}) \times^{1\ldots N} \left(\circ_{j=1}^{N} U_j^{-1}\right) \times^{1\ldots N} (\mathcal{X} - \mathcal{M})\right\}. \quad (18)$$

The tensor normal distribution can be rewritten as a multivariate normal distribution with separable covariance matrix for the vectorized tensor, more precisely it holds (see Ohlson et al. (2013)) $\mathcal{X} \sim \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{M}, U_1, \ldots, U_N) \iff \mathrm{vec}(\mathcal{X}) \sim \mathcal{N}_{I_1 \cdots I_N}(\mathrm{vec}(\mathcal{M}), U_N \otimes \ldots \otimes U_1)$. The restriction imposed by the separability assumption allows to reduce the number of parameters to estimate with respect to the unrestricted vectorized from, while allowing both within and between mode dependence.

The unrestricted model in eq. (16) cannot be estimated, as the number of parameters greatly outmatches the available data. We address this issue by assuming a PARAFAC(R) decomposition for the tensor coefficients, which makes the estimation feasible by reducing the dimension of the parameter space. For example, let $\mathcal{B}$ be a N-order tensor of sizes $I_1 \times \ldots \times I_N$ and rank R, then the number of parameters to estimate in the unrestricted case is given by $\prod_{i=1}^{N} I_i$ while in the PARAFAC(R) restricted model is $R \sum_{i=1}^{N} I_i$.

Example 2.1. *For the sake of exposition, consider the model in eq.* (16) *where the response is a third-order tensor* $\mathcal{Y}_t \in \mathbb{R}^{k \times k \times k^2}$ *and the covariates include only a constant term, that is a coefficient tensor* $\mathcal{A}_0$ *of the same size. Define by* $k_{\mathcal{E}}$ *the number of parameters of the noise distribution. As a result, the total number of parameters to estimate in the unrestricted case is given by:*

$$\prod_{i=1}^{3} I_i + k_{\mathcal{E}} = \mathcal{O}(k^4), \quad (19)$$

while assuming a PARAFAC(R) decomposition on $\mathcal{A}_0$ *it reduces to:*

$$\sum_{r=1}^{R} \sum_{i=1}^{3} I_i + k_{\mathcal{E}} = \mathcal{O}(k^2). \quad (20)$$

The magnitude of this reduction is illustrated in Fig. (2)*, for two different values of the rank.*

A well known issue is that a low rank decomposition is not unique. In a statistical model this translates into an identification problem for the PARAFAC marginals $\boldsymbol{\beta}_j^{(r)}$ arising from three sources:

(i) *scale* identification, because replacing $\boldsymbol{\beta}_j^{(r)}$ with $\lambda_{jr}\boldsymbol{\beta}_j^{(r)}$ for $\prod_{j=1}^{N} \lambda_{jr} = 1$ does not alter the outer product;

(ii) *permutation* identification, since for any permutation of the indices $\{1, \ldots, R\}$ the outer product of the original vectors is equal to that of the permuted ones;

(iii) *orthogonal transformation* identification, due to the fact that multiplying two marginals by an orthonormal matrix Q leaves unchanged the outcome $\boldsymbol{\beta}_j^{(r)} Q \circ \boldsymbol{\beta}_k^{(r)} Q = \boldsymbol{\beta}_j^{(r)} \circ \boldsymbol{\beta}_k^{(r)}$.

In our framework these issues do not hamper the inference as our interest is only in the coefficient tensor, which is exactly identified. In fact, we use the PARAFAC decomposition as a practical modelling tool without attaching any interpretation to its marginals.

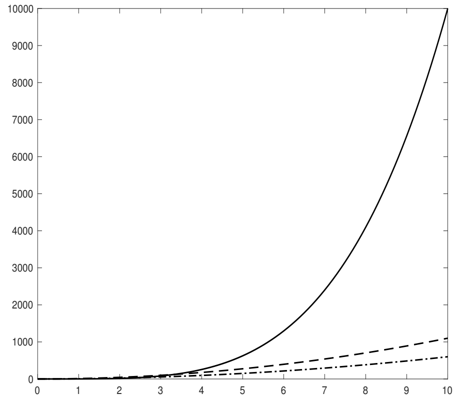

Figure 2: Number of parameters (*vertical axis*) as function of the response dimension (*horizontal axis*) for unconstrained (*solid*) and PARAFAC(R) with $R = 10$ (*dashed*) and $R = 5$ (*dotted*).

2.3 Important special cases

The model in eq. (16) is a generalization of several well-known econometric models, as shown in the following remarks.

Remark 2.3 (Univariate)**.** *If we set $I_j = 1$ for $j = 1, \dots, N$, then the model in eq.* (16) *reduces to a univariate regression:*

$$y_t = \mathcal{A} + \mathcal{B}' \operatorname{vec}(X_t) + \mathcal{C}' \mathbf{z}_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \tag{21}$$

where the coefficients reduce to $\mathcal{A} = \bar{\alpha} \in \mathbb{R}$, $\mathcal{B} = \boldsymbol{\beta} \in \mathbb{R}^Q$ and $\mathcal{C} = \boldsymbol{\gamma} \in \mathbb{R}^J$. See Appendix B for further details.

Remark 2.4 (SUR)**.** *If we set $I_j = 1$ for $j = 2, \dots, N$ and define the unit vector $\boldsymbol{\iota} \in \mathbb{R}^{I_1}$, then the model in eq.* (16) *reduces to a multivariate regression which is interpretable as a Seemingly Unrelated Regression (SUR) model (Zellner (1962)):*

$$\mathbf{y}_t = \mathcal{A} + \mathcal{B} \times_2 \mathbf{z}_t + \mathcal{C} \times_2 \operatorname{vec}(X_t) + \mathcal{D} \times_1 \operatorname{vec}(W_t) + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \overset{iid}{\sim} \mathcal{N}_m(0, \Sigma), \tag{22}$$

where the tensors of coefficients can be expressed as: $\mathcal{A} = \boldsymbol{\alpha} \in \mathbb{R}^m$, $\mathcal{B} = \bar{B} \in \mathbb{R}^{m \times J}$, $\mathcal{C} = C \in \mathbb{R}^{m \times Q}$ and $\mathcal{D} = \mathbf{d} \in \mathbb{R}^m$. See Appendix B for further details.

Remark 2.5 (VARX and Panel VAR)**.** *Consider the setup of the previous Remark 2.4. If we choose $\mathbf{z}_t = \mathbf{y}_{t-1}$ we end up with an (unrestriced) VARX(1) model. Notice that another vector of regressors $\mathbf{w}_t = \operatorname{vec}(W_t) \in \mathbb{R}^q$ may enter the regression* (22) *pre-multiplied (along mode-*3*) by a tensor $\mathcal{D} \in \mathbb{R}^{m \times n \times q}$. Since we are not putting any kind of restrictions on the covariance matrix Σ in* (22)*, the general model* (16) *encompasses as a particular case also the panel VAR models of Canova and Ciccarelli (2004), Canova et al. (2007), Canova and Ciccarelli (2009) and Canova et al. (2012).*

Remark 2.6 (VECM)**.** *It is possible to interpret the model in eq.* (16) *as a generalisation of the Vector Error Correction Model (VECM) widely used in multivariate time series analysis (see Engle and Granger (1987), Schotman and Van Dijk (1991)). A standard K-dimensional VAR(1) model reads:*

$$\mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \Sigma) \,. \tag{23}$$

Defining $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ *and* $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$*, where* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *are* $K \times R$ *matrices of rank* $R < K$*, we obtain the VECM used for studying the cointegration relations between the components of* $\mathbf{y}_t$*:*

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \,. \tag{24}$$

Since $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}' = \sum_{r=1}^{R} \boldsymbol{\alpha}_{:,r}\boldsymbol{\beta}'_{:,r} = \sum_{r=1}^{R} \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)}$*, we can interpret the VECM model in the previous equation as a particular case of the model in eq.* (16) *where the coefficient* $\mathcal{B}$ *is the matrix* $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$*. Furthermore by writing* $\Pi = \sum_{r=1}^{R} \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)}$ *we can interpret this relation as a rank-R PARAFAC decomposition of* Π*. Thus we can interpret the rank of the PARAFAC decomposition for the matrix of coefficients as the cointegration rank and, in presence of cointegrating relations, the vectors* $\tilde{\boldsymbol{\beta}}_1^{(r)}$ *are the mean-reverting coefficients and* $\tilde{\boldsymbol{\beta}}_2^{(r)} = (\tilde{\beta}_{2,1}^{(r)}, \ldots, \tilde{\beta}_{2,K}^{(r)})$ *are the cointegrating vectors. In fact, the PARAFAC(R) decomposition for matrices corresponds to a low rank (R) matrix approximation (see Eckart and Young (1936)). We make reference to Appendix B for further details.*

Remark 2.7 (Tensor AR)**.** *By removing all the covariates from eq.* (16) *except the lags of the dependent variable, we obtain a tensor autoregressive model:*

$$\mathcal{Y}_t = \mathcal{A}_0 + \sum_{j=1}^{p} \mathcal{A}_j \times_{D+1} \mathcal{Y}_{t-j} + \mathcal{E}_t \qquad \mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(0, \Sigma_1, \ldots, \Sigma_N) \,. \tag{25}$$

3 Bayesian Inference

In this section, without loss of generality, we present the inference procedure for a special case of the model in eq. (16), given by:

$$Y_t = \mathcal{B} \times_3 \operatorname{vec}(X_t) + E_t, \quad E_t \overset{iid}{\sim} \mathcal{N}_{I_1,I_2}(\mathbf{0}, \Sigma_1, \Sigma_2) \,, \tag{26}$$

which can also be rewritten in vectorized form as:

$$\operatorname{vec}(Y_t) = \mathbf{B}'_{(3)} \operatorname{vec}(X_t) + \operatorname{vec}(E_t) \,, \quad \operatorname{vec}(E_t) \overset{iid}{\sim} \mathcal{N}_{I_1 I_2}(\mathbf{0}, \Sigma_2 \otimes \Sigma_1) \,. \tag{27}$$

Here $Y_t \in \mathbb{R}^{I_1 \times I_2}$ is a matrix response, $X_t \in \mathbb{R}^{I_1 \times I_2}$ is a covariate matrix of the same size of Y_t and $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_1 I_2}$ is a coefficient tensor. The noise term $E_t \in \mathbb{R}^{I_1 \times I_2}$ is distributed according to a matrix variate normal distribution, with zero mean and covariance matrices $\Sigma_1 \in \mathbb{R}^{I_1 \times I_1}$ and $\Sigma_2 \in \mathbb{R}^{I_2 \times I_2}$ accounting for the covariance between the columns and the rows, respectively. This distribution is a particular case of the tensor Gaussian introduced in eq. (18) whose probability density function is given by:

$$f_X(X) = (2\pi)^{-\frac{I_1 I_2}{2}} |U_2|^{-\frac{I_1}{2}} |U_1|^{-\frac{I_2}{2}} \exp\left\{-\frac{1}{2} U_2^{-1} (X - M)' U_1^{-1} (X - M)\right\} \tag{28}$$

where $X \in \mathbb{R}^{I_1 \times I_2}$, $M \in \mathbb{R}^{I_1 \times I_2}$ is the mean matrix and the covariance matrices are $U_j \in \mathbb{R}^{I_j \times I_j}$, $j = 1, 2$, where index 1 represents the rows and index 2 stands for the columns of the variable X.

The choice the Bayesian approach for inference is motivated by the fact that the large number of parameters may lead to an over-fitting problem, especially when the samples size is rather small. This issue can be addressed by the indirect inclusion of parameter restrictions through a suitable specification of the corresponding prior distribution. Considering the unrestricted model in eq. (26), it would be necessary to define a prior distribution on the three-dimensional array $\mathcal{B}$. The literature on this topic is scarce: though Ohlson et al. (2013) and Manceur and

Dutilleul (2013) presented the family of elliptical array-valued distributions, which include the tensor normal and tensor t, the latter are rather inflexible as imposing some structure on a subset of the entries of the array is very complicated.

We assume a PARAFAC(R) decomposition on the coefficient tensor for achieving two goals: first, by reducing the parameter space this assumption makes estimation feasible; second, the decomposition transforms a multidimensional array into the outer product of vectors, we are left we the choice of a prior distribution on vectors, for which many constructions are available. In particular, we can incorporate sparsity beliefs by specifying a suitable shrinkage prior directly on the marginals of the PARAFAC. Indirectly, this introduces *a priori* sparsity on the coefficient tensor.

3.1 Prior Specification

The choice of the prior distribution on the PARAFAC marginals is crucial for recovering the sparsity pattern of the coefficient tensor and for the efficiency of the inference. In the Bayesian literature the global-local class of prior distributions represent a popular and successful structure for providing shrinkage and regularization in a wide range of models and applications. These priors are based on scale mixtures of normal distributions, where the different components of the covariance matrix produce desirable shrinkage properties of the parameter. By construction, global-local priors are not suited for recovering an exact zero (differently from spike-and-slab priors, see Mitchell and Beauchamp (1988), George and McCulloch (1997), Ishwaran and Rao (2005)), instead they can be recovered via post-estimation thresholding (see Park and Casella (2008)). However, spike-and-slab priors become intractable as the dimensionality of the parameter grows. By contrast, the global-local shrinkage priors have greater scalability and thus represent a desirable choice in high-dimensional models, such as our framework. Motivated by these arguments, we adopt the hierarchical specification forwarded by Guhaniyogi et al. (2017) in order to define adequate global-local shrinkage priors for the marginals[4].

The global-local shrinkage prior for each PARAFAC marginal $\boldsymbol{\beta}_j^{(r)}$ of the coefficient tensor $\mathcal{B}$ is defined as a scale mixture of normals centred in zero, with three components for the covariance. The global component τ is drawn from a gamma distribution[5]. The vector of component-level (shared by all marginals in the r-th component of the decomposition) variances $\boldsymbol{\phi}$ is sampled from a R-dimensional Dirichlet distribution with parameter $\boldsymbol{\alpha} = \alpha\boldsymbol{\iota}_R$, where $\boldsymbol{\iota}_R$ is the vector of ones of length R. Finally, the local component of the variance is a diagonal matrix $W_{j,r} = \text{diag}(\mathbf{w}_{j,r})$ whose entries are exponentially distributed with hyper-parameter $\lambda_{j,r}$. The latter is a key parameter for driving the shrinkage to zero of the marginals and is drawn from a gamma distribution. Summarizing, for $p = 1, \dots, I_j$, $j = 1, \dots, 3$ and $r = 1, \dots, R$ we have the following hierarchical prior structure for each vector of the PARAFAC(R) decomposition in eq. (11):

$$\pi(\alpha) \sim \mathcal{U}(A) \tag{29a}$$

$$\pi(\boldsymbol{\phi}|\alpha) \sim \mathcal{D}ir(\alpha\boldsymbol{\iota}_R) \tag{29b}$$

$$\pi(\tau|\alpha) \sim \mathcal{G}a(a_\tau, b_\tau) \tag{29c}$$

$$\pi(\lambda_{j,r}) \sim \mathcal{G}a(a_\lambda, b_\lambda) \tag{29d}$$

$$\pi(w_{j,r,p}|\lambda_{j,r}) \sim \mathcal{E}xp(\lambda_{j,r}^2/2) \tag{29e}$$

[4]This class of shrinkage priors has been firstly proposed by Bhattacharya et al. (2015) and Zhou et al. (2015).

[5]We use the shape-rate formulation for the gamma distribution:

$$x \sim \mathcal{G}a(a,b) \qquad \iff \qquad f(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \qquad a > 0, b > 0$$

.

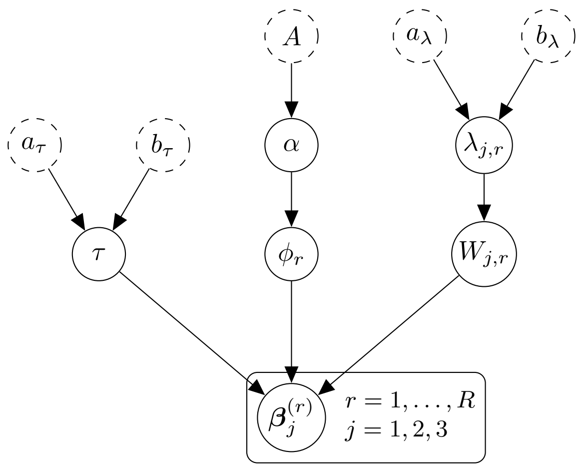

Figure 3: Hierarchical shrinkage prior for the PARAFAC marginals. White circles with continuous border represent the parameters, white circles with dashed border represent fixed hyper-parameters.

$$\pi\left(\boldsymbol{\beta}_j^{(r)} \middle| W_{j,r}, \boldsymbol{\phi}, \tau\right) \sim \mathcal{N}_{I_j}(\mathbf{0}, \tau\phi_r W_{j,r}) \,. \tag{29f}$$

Concerning the covariance matrices for the noise term in eq. (16), the Kronecker structure does not allow to separately identify the scale of the covariance matrices U_n, thus requiring the specification of further restrictions. Wang and West (2009) and Dobra (2015) adopt independent hyper-inverse Wishart prior distributions (Dawid and Lauritzen (1993)) for each U_n, then impose the identification restriction $U_{n,11} = 1$ for $n = 2, \ldots, N$. Instead, Hoff (2011) suggests to introduce dependence between the Inverse Wishart prior distribution $\mathcal{IW}(\nu_n, \gamma\Psi_n)$ of each U_n, $n = 1, \ldots, N$, via a hyper-parameter $\gamma \sim \mathcal{G}a(a, b)$ affecting the scale of each location matrix parameter. Finally, the hard constraint $\Sigma_n = \mathbf{I}_{I_n}$ (where $\mathbf{I}_k$ is the identity matrix of size k), for all but one n, implicitly imposes that the dependence structure within different modes is the same, but there is no dependence between modes. To account for marginal dependence, it is possible to add a level of hierarchy by introducing a hyper-parameter in the spirit of Hoff (2011). Following Hoff (2011), we assume conditionally independent inverse Wishart prior distributions for the covariance matrices of the error term E_t and add a level of hierarchy via the hyper-parameter γ which governs the scale of the covariance matrices:

$$\pi(\gamma) \sim \mathcal{G}a(a_\gamma, b_\gamma) \tag{30a}$$
$$\pi(\Sigma_1|\gamma) \sim \mathcal{IW}_{I_1}(\nu_1, \gamma\Psi_1) \tag{30b}$$
$$\pi(\Sigma_2|\gamma) \sim \mathcal{IW}_{I_2}(\nu_2, \gamma\Psi_2) \,. \tag{30c}$$

Defining the vector of all parameters as $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\phi}, \tau, \Lambda, \mathbf{W}, \mathcal{B}, \Sigma_1, \Sigma_2\}$, with $\Lambda = \{\lambda_{j,r} : j = 1, \ldots, 3,\ r = 1, \ldots, R\}$ and $\mathbf{W} = \{W_{j,r} : j = 1, \ldots, 3,\ r = 1, \ldots, R\}$, the joint prior distribution is given by:

$$\pi(\boldsymbol{\theta}) = \pi(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau)\pi(\mathbf{W}|\Lambda)\pi(\boldsymbol{\phi}|\boldsymbol{\alpha})\pi(\tau|\boldsymbol{\alpha})\pi(\Lambda)\pi(\boldsymbol{\alpha})\pi(\Sigma_1|\gamma)\pi(\Sigma_2|\gamma)\pi(\gamma). \tag{31}$$

The directed acyclic graphs (DAG) of the hierarchical shrinkage prior on the PARAFAC marginals $\boldsymbol{\beta}_j^{(r)}$ and the overall prior structure are given in Figs. 3-4, respectively.

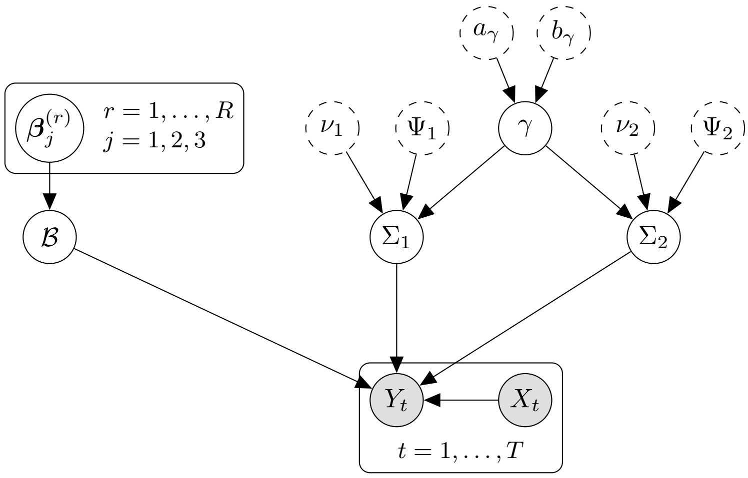

Figure 4: Overall prior structure. Gray circles represent observable variables, white circles with continuous border represent the parameters, white circles with dashed border represent fixed hyper-parameters.

3.2 Posterior Computation

The likelihood function of the model in eq. (26) is given by:

$$L\left(Y_1, \ldots, Y_T | \boldsymbol{\theta}\right) = \prod_{t=1}^{T} (2\pi)^{-\frac{I_1 I_2}{2}} \left|\Sigma_2\right|^{-\frac{I_1}{2}} \left|\Sigma_1\right|^{-\frac{I_2}{2}} \exp\left\{ -\frac{1}{2} \Sigma_2^{-1} (Y_t - \mathcal{B} \times_3 \mathbf{x}_t)' \Sigma_1^{-1} (Y_t - \mathcal{B} \times_3 \mathbf{x}_t) \right\}, \tag{32}$$

where $\mathbf{x}_t = \operatorname{vec}(X_t)$. Since the posterior distribution is not tractable in closed form, we adopt an MCMC procedure based on Gibbs sampling. The computations and technical details of the derivation of the posterior distributions are given in Appendix D. As a consequence of the hierarchical structure of the prior, we can articulate the sampler in three main blocks:

I) sample the hyper-parameters of the global and component-level variance for the marginals, according to:

$$p(\boldsymbol{\alpha}, \boldsymbol{\phi}, \tau | \mathcal{B}, \mathbf{W}) = p(\boldsymbol{\alpha} | \mathcal{B}, \mathbf{W}) p(\boldsymbol{\phi}, \tau | \boldsymbol{\alpha}, \mathcal{B}, \mathbf{W}) \tag{33}$$

(i) sample α from:

$$\mathbb{P}\left(\left\{ \alpha = \alpha_j \right\} | \mathcal{B}, \mathbf{W} \right) = \frac{p(\alpha_j | \mathcal{B}, \mathbf{W})}{\sum_{l=1}^{|A|} p(\alpha_l | \mathcal{B}, \mathbf{W})}, \tag{34}$$

where:

$$p(\alpha | \mathcal{B}, \mathbf{W}) = \pi(\alpha) \frac{1}{M} \sum_{i=1}^{M|A|} \omega_i. \tag{35}$$

(ii) sample independently the auxiliary variable ψ_r, for $r = 1, \ldots, R$, from:

$$p(\psi_r | \mathcal{B}, \mathbf{W}, \alpha) \propto GiG\left(\alpha - \frac{I_0}{2}, 2b_\tau, 2C_r \right) \tag{36}$$

then, for $r = 1, \ldots, R$:

$$\phi_r = \frac{\psi_r}{\sum_{l=1}^{R} \psi_l}. \tag{37}$$

(iii) finally, sample τ from:

$$p(\tau|\mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) \propto GiG\left(a_\tau - \frac{RI_0}{2}, 2b_\tau, 2\sum_{r=1}^{R}\frac{C_r}{\phi_r}\right). \tag{38}$$

II) define $\mathbf{Y} = \{Y_t\}_{t=1}^{T}$, then sample from the posterior of the hyper-parameters of the local component of the variance of the marginals and the marginals themselves, as follows:

$$\begin{aligned} p\left(\boldsymbol{\beta}_j^{(r)}, W_{j,r}, \lambda_{j,r}\middle|\boldsymbol{\phi}, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2\right) &= p\left(\lambda_{j,r}|\boldsymbol{\beta}_j^{(r)}, \phi_r, \tau\right) p\left(w_{j,r,p}|\lambda_{j,r}, \phi_r, \tau, \boldsymbol{\beta}_j^{(r)}\right) \\ &\cdot p\left(\boldsymbol{\beta}_j^{(r)}|\boldsymbol{\beta}_{-j}^{(r)}, \mathcal{B}_{-r}, \phi_r, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2\right) \end{aligned} \tag{39}$$

(i) for $j = 1, 2, 3$ and $r = 1, \ldots, R$ sample independently:

$$p\left(\lambda_{j,r}|\boldsymbol{\beta}_j^{(r)}, \phi_r, \tau\right) \propto \mathcal{G}a\left(a_\lambda + I_j, b_\lambda + \frac{\left\|\boldsymbol{\beta}_j^{(r)}\right\|_1}{\sqrt{\tau\phi_r}}\right). \tag{40}$$

(ii) for $p = 1, \ldots, I_j$, $j = 1, 2, 3$ and $r = 1, \ldots, R$ sample:

$$p\left(w_{j,r,p}|\lambda_{j,r}, \phi_r, \tau, \boldsymbol{\beta}_j^{(r)}\right) \propto GiG\left(\frac{1}{2}, \lambda_{j,r}^2, \frac{\beta_{j,k}^{(r)^2}}{\tau\phi_r}\right) \tag{41}$$

(iii) define $\boldsymbol{\beta}_{-j}^{(r)} = \left\{\boldsymbol{\beta}_i^{(r)} : i \neq j\right\}$ and $\mathcal{B}_{-r} = \{B_i : i \neq r\}$, where $B_r = \boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_N^{(r)}$. For $r = 1, \ldots, R$ sample the PARAFAC marginals from:

$$p\left(\boldsymbol{\beta}_1^{(r)}|\boldsymbol{\beta}_{-1}^{(r)}, \mathcal{B}_{-r}, \phi_r, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2\right) \propto \mathcal{N}_{I_1}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1}, \bar{\Sigma}_{\boldsymbol{\beta}_1}) \tag{42}$$

$$p\left(\boldsymbol{\beta}_2^{(r)}|\boldsymbol{\beta}_{-2}^{(r)}, \mathcal{B}_{-r}, \phi_r, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2\right) \propto \mathcal{N}_{I_2}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2}, \bar{\Sigma}_{\boldsymbol{\beta}_2}) \tag{43}$$

$$p\left(\boldsymbol{\beta}_3^{(r)}|\boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, \phi_r, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2\right) \propto \mathcal{N}_{I_3}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3}, \bar{\Sigma}_{\boldsymbol{\beta}_3}). \tag{44}$$

III) sample the covariance matrices from their posterior:

$$p(\Sigma_1, \Sigma_2, \gamma|\mathcal{B}, \mathbf{Y}) = p(\Sigma_1|\mathcal{B}, \mathbf{Y}, \Sigma_2, \gamma)p(\Sigma_2|\mathcal{B}, \mathbf{Y}, \Sigma_1, \gamma)p(\gamma|\Sigma_1, \Sigma_2) \tag{45}$$

(i) sample the row covariance matrix:

$$p(\Sigma_1|\mathcal{B}, \mathbf{Y}, \Sigma_2, \gamma) \propto \mathcal{IW}_{I_1}(\nu_1 + I_1, \gamma\Psi_1 + S_1) \tag{46}$$

(ii) sample the column covariance matrix:

$$p(\Sigma_2|\mathcal{B}, \mathbf{Y}, \Sigma_1, \gamma) \propto \mathcal{IW}_{I_2}(\nu_2 + I_2, \gamma\Psi_2 + S_2). \tag{47}$$

(iii) sample the scale hyper-parameter:

$$p(\gamma|\Sigma_1, \Sigma_2) \propto \mathcal{G}a\left(\nu_1 I_1 + \nu_2 I_2, \mathrm{tr}\left(\Psi_1\Sigma_1^{-1} + \Psi_2\Sigma_2^{-1}\right)\right). \tag{48}$$

For improving the mixing of the algorithm, it is possible to substitute the draw from the full conditional distributon of the global variance parameter τ or of the PARAFAC marginals with a Hamiltonian Monte Carlo (HMC) step (see Neal (2011)).

4 Simulation Results

We report the results of a simulation study where we have tested the performance of the proposed sampler on synthetic datasets of matrix-valued sequences $\{Y_t, X_t\}_{t=1}^T$, where Y_t, X_t have different size across simulations. The methods described in this paper can be rather computationally intensive, nevertheless thanks to the tensor decomposition we used allows the estimation to be carried out on a laptop. All the simulations were run on an Apple MacBookPro with a 3.1GHz Intel Core i7 processor, RAM 16GB, using MATLAB r2017b with the aid of the Tensor Toolbox v.2.6[6], taking about 30h for the highest-dimensional case (i.e. $I_1 = I_2 = 50$).

For different sizes ($I_1 = I_2$) of the response and covariate matrices, we generated a matrix-variate time series $\{Y_t, X_t\}_{t=1}^T$ by simulating each entry of X_t from:

$$x_{ij,t} - \mu = \alpha_{ij}(x_{ij,t-1} - \mu) + \eta_{ij,t}\,, \qquad \eta_{ij,t} \sim \mathcal{N}(0,1) \tag{49}$$

and a matrix-variate time series $\{Y_t\}_t$ according to:

$$Y_t = \mathcal{B} \times_3 \mathrm{vec}\,(X_t) + E_t\,, \qquad E_t \sim \mathcal{N}_{I_1,I_2}(\mathbf{0}, \Sigma_1, \mathbf{I}_{I_2})\,. \tag{50}$$

where $\mathbb{E}[\eta_{ij,t}\eta_{kl,v}] = 0$, $\mathbb{E}[\eta_{ij,t}E_v] = 0$, $\forall\,(i,j) \neq (k,l)$, $\forall\,t \neq v$, and $\alpha_{ij} \sim \mathcal{U}(-1,1)$. We randomly draw $\mathcal{B}$ by using the PARAFAC representation in eq. (11), with rank $R = 5$ and marginals sampled from the prior distribution in eq. (29f).

The response and covariate matrices in the simulated datasets have the following sizes:

(I) $I_1 = I_2 = I = 10$, for $T = 60$;

(II) $I_1 = I_2 = I = 20$, for $T = 60$;

(III) $I_1 = I_2 = I = 30$, for $T = 60$;

(IV) $I_1 = I_2 = I = 40$, for $T = 60$;

(V) $I_1 = I_2 = I = 50$, for $T = 60$.

We initialized the Gibbs sampler by setting the PARAFAC marginals $\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}$, $r = 1, \ldots, R$ (with $R = 5$), with the output of a simulated annealing algorithm (see Appendix C) and run the algorithm for $N = 10000$ iterations. We present the results for the case $\boldsymbol{\Sigma}_2 = \mathbf{I}_{I_2}$. Since they are similar, we omit the results for unconstrained Σ_2, estimated with the Gibbs in Section 3.

[6]Available at: http://www.sandia.gov/ tgkolda/TensorToolbox/index-2.6.html

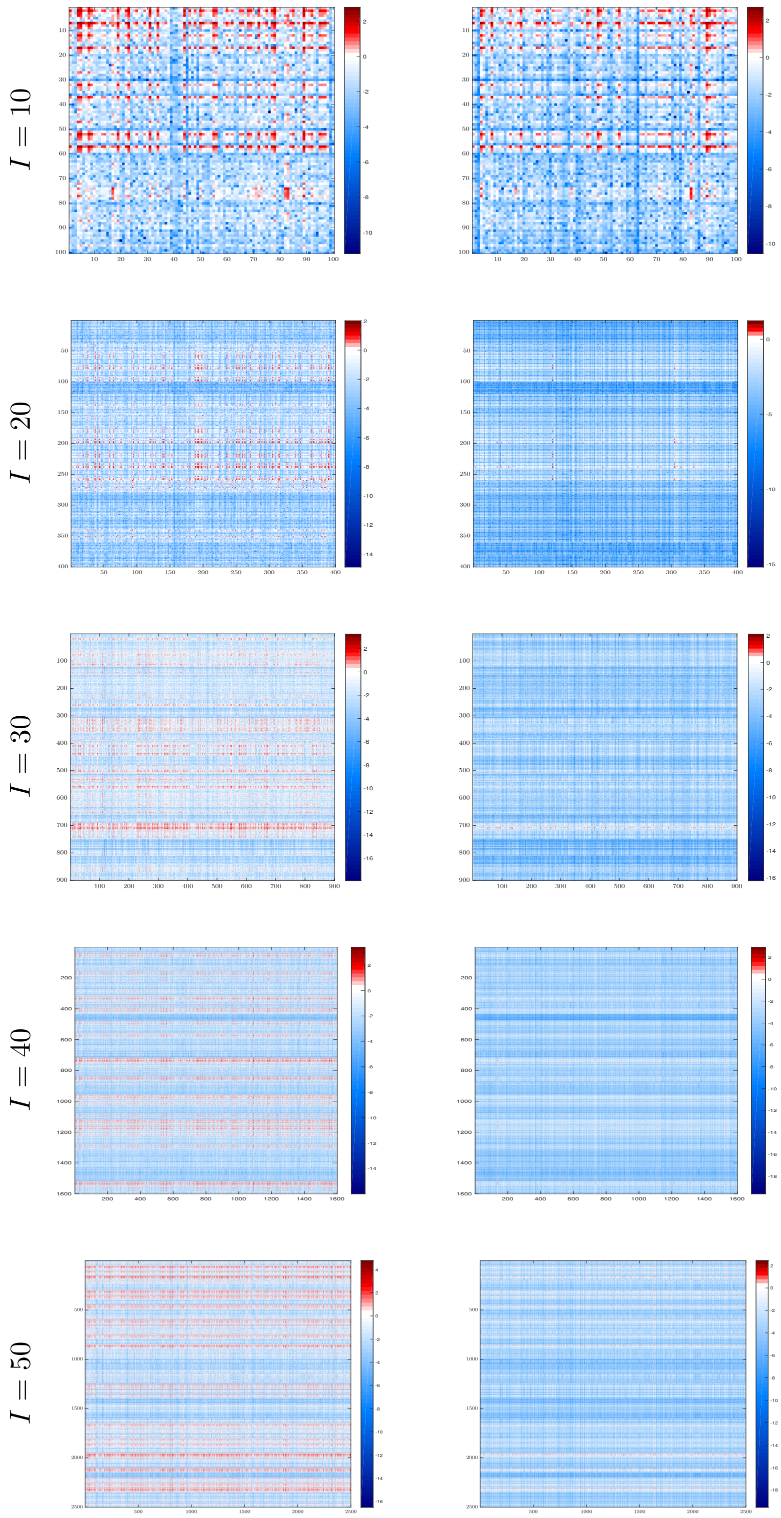

Figure 5: Logarithm of the absolute value of the coefficient tensors: true $\mathcal{B}$ (*left*) and estimated $\hat{\mathcal{B}}$ (*right*).

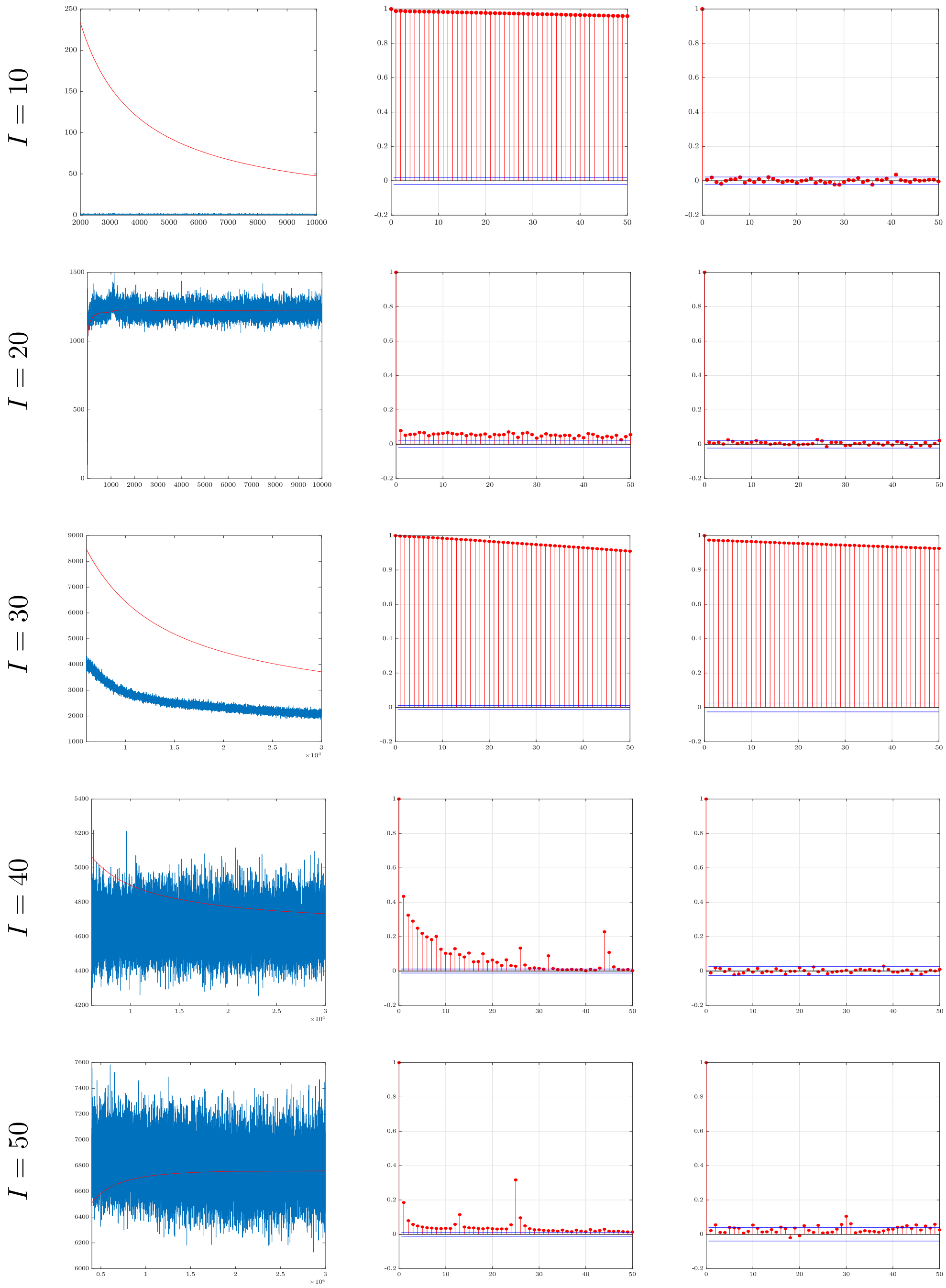

Figure 6: MCMC output (*left*), autocorrelation function for the entire sample (*middle* plot) and after burn-in (*right* plot) of the Frobenious norm of the difference between the true and the estimated covariance matrix Σ_1.

The results are reported in Figs. 5-6, for the different simulated datasets. Fig. 5 shows the good accuracy of the sampler in estimating the coefficient tensor, whose number of entries ranges from 10^4 in the first to 50^4 in the last simulation setting. The estimation error is maily due to the over-shrinking to zero of large signals. This well known drawback of global-local hierarchical prior distributions (e.g., see Carvalho et al. (2010)) is related to its sensitivity to the hyper-parameters setting. Fig. 6 plots the MCMC output of the Frobenious norm (i.e. the L_2 norm) of the covariance matrix of the error term. After a graphical inspection of the trace plots (first column) we chose a burn-in period of 2000 iterations. Due to autocorrelation in the sample (second column plots) we applied thinning and selected every 10th iteration. In most of the cases, after removing burn-in iterations and performing thinning, the autocorrelation wipes out.

We refer the reader to Appendix F for additional details on the simulation experiments, such as trace plots and autocorrelation functions for tensor entries and individual hyper-parameters.

5 Application

5.1 Data description

As put forward by Schweitzer et al. (2009), the analysis of economic networks is one of the most recent and complex challenges that the econometric community is facing nowadays. We contribute to the econometric literature about complex networks by applying the proposed methodology to the study of the temporal evolution of the international trade network (ITN). This economic network has been previously studied by several authors (e.g., see Hidalgo and Hausmann (2009), Fagiolo et al. (2009), Kharrazi et al. (2017), Meyfroidt et al. (2010), Zhu et al. (2014), Squartini et al. (2011)), who have analysed its topological properties and identified its main communities. However, to the best of our knowledge, this is the first attempt to model the temporal evolution of the network as a whole.

The raw trade data come from the United Nations COMTRADE database, a publicly available resource[7]. The particular dataset we use is a subset of the whole COMTRADE database and consists of yearly observations from 1998 to 2016 of total imports and exports between $I_1 = I_2 = I = 10$ countries. In order to remove possible sources of non-linearities in the data, we use a logarithmic transform of the variables of interest. We thus consider the international trade network at each time stamp as one observation from a real-valued matrix-variate stochastic process. Fig. 7 shows the whole network sequence in our dataset.

5.2 Results

We estimate the model setting $X_t = Y_{t-1}$, thus obtaining a matrix-variate autoregressive model. Each matrix Y_t is the $I \times I$ real-valued weighted adjacency matrix of the corresponding international trade network in year t, whose entry (i, j) contains the total exports of country i vis-à-vis country j, in year t. The series $\{Y_t\}_t$, $t = 1, \ldots, T$, has been standardized (over the temporal dimension). We run the Gibbs sampler for $N = 10,000$ iterations. The output is reported below.

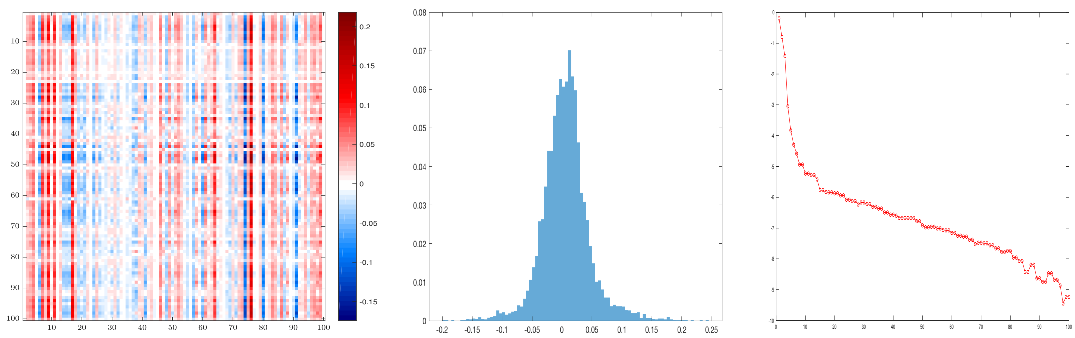

Figure 8: *Left:* Transpose of the mode-3 matricized estimated coefficient tensor, $\hat{\mathbf{B}}'_{(3)}$. *Middle:* distribution of the estimated entries of $\hat{\mathbf{B}}_{(3)}$. *Right:* logarithm of the modulus of the eigenvalues of $\hat{\mathbf{B}}_{(3)}$, in decreasing order.

[7]https://comtrade.un.org

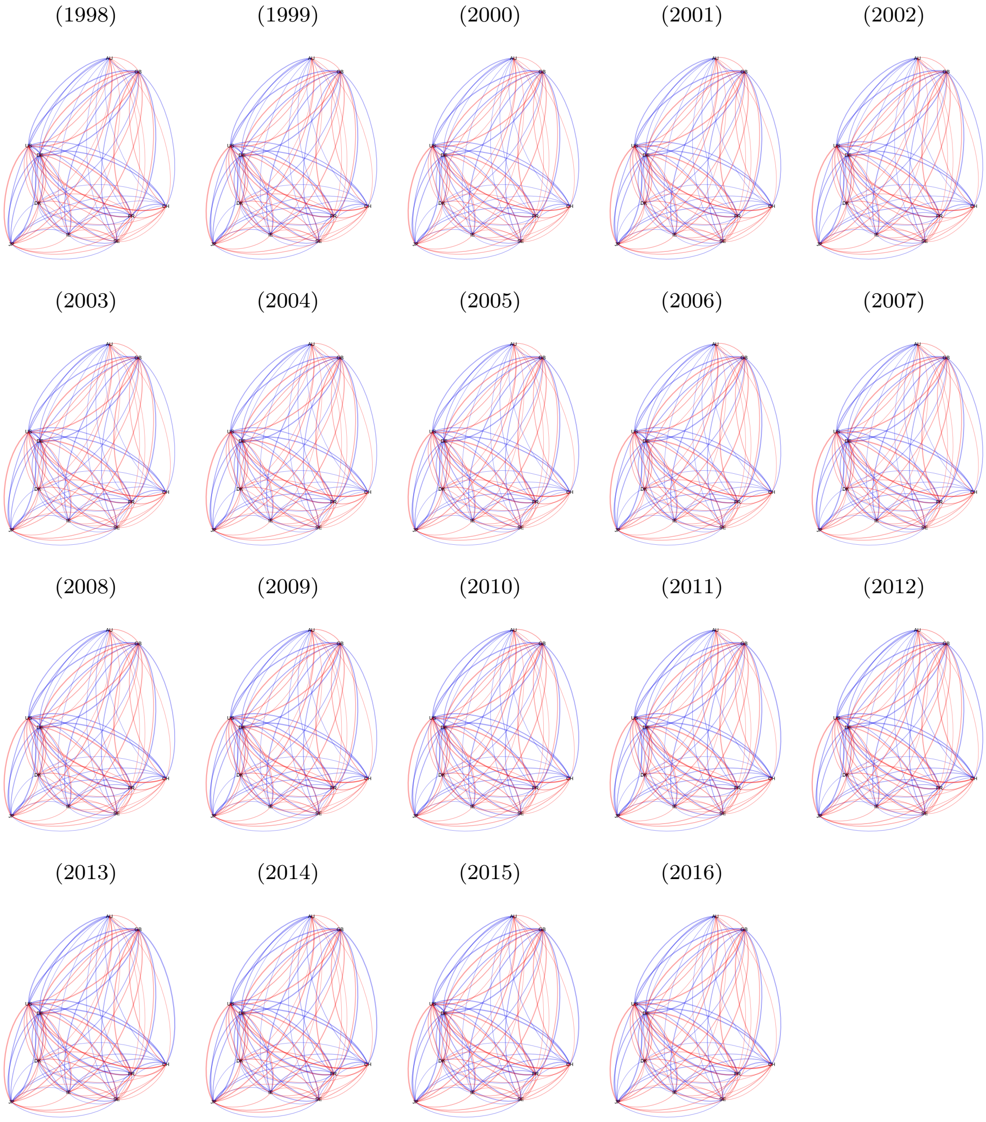

Figure 7: Commercial trade network evolving over time from 1998 (*top left*) to 2016 (*bottom right*). Nodes represent countries, red and blue colored edges stand for exports and imports between two countries, respectively. Edge thickness represents the magnitude of the flow.

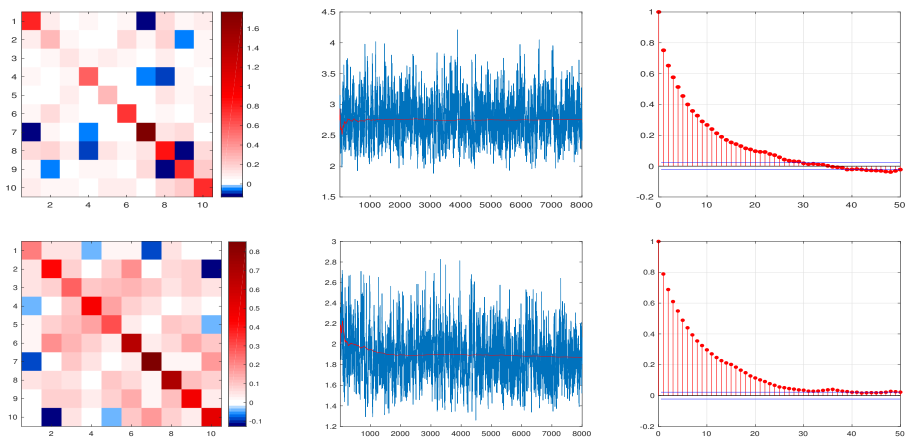

Figure 9: Estimated covariance matrix of the noise term (*first*), posterior distributions (*second*), MCMC output (*third*) and autocorrelation functions (*fourth*) of the Frobenious norm of the covariance matrix of the noise term. *First row*: Σ_1, *second row*: Σ_2.

The mod-3 matricization of the estimated coefficient tensor is shown in the left panel of Fig. 8, each column corresponds to the effects of a lag one edge (horizontal axis) on all the contemporaneous edges (vertical axis). Positive effects in red and negative effects in blue. Fig. 9 shows the estimated covariance matrices of the noise term, that is $\hat{\Sigma}_1, \hat{\Sigma}_2$. As regards the estimated coefficient tensor, we find that:

- the heterogeneity in the estimated coefficients points against parameter pooling assumptions;
- there are patterns, it look like there are groups of edges (bilateral trade flows) with mainly positive (red) or negative (blue) effect on all the other edges. Maybe there are some countries that play a key role for these flows;
- the distribution of the entries of the estimated coefficient tensor (middle panel) confirms the evidence of heterogeneity. The distribution is right-skewed and leptokurtic with mode at zero, which is a consequence of the shrinkage of the estimated coefficient;
- in order to assess the stationarity of the model, we computed the eigenvalues of the mode-3 matricization of the estimated coefficient tensor and the right panel of Fig. 8 plots the logarithm of their modulus. All the eigenvalues are strictly lower than one in modulus, thus indicating that the process describing the evolution of the trade network is stationary.

Moreover, as regards the estimated covariance matrices of the noise term (Fig. 9), we find that:

- in both cases the highest values correspond to individual variances, while the estimated covariances are lower in magnitude and heterogeneous;
- there is evidence of heterogeneity in the dependence structure, since Σ_1, which captures the covariance between exporting countries (i.e., rows), differs from Σ_2, which describes the covariance between importing countries (i.e., columns);
- the dependence between exporting countries is higher on average than between importing countries;
- for assessing the convergence of the MCMC chain, Fig. 9 shows the trace plot and autocorrelation functions (without thinning) of the Frobenious norm of each estimated matrix. Both sets of plots show a good mixing of the chain.

5.3 Impulse response analysis

For understanding the role exerted by the various links of the network, Fig. 10, top panel, shows for each edge the sum of the corresponding positive and negative entries of the estimated coefficient tensor in red and blue, respectively. We find that edges' impact tend to cluster, that is, those with high positive cumulated effects have very low negative cumulated effects and vice-versa. Thus, the bottom panel of Fig. 10 shows the sum of the absolute values of all corresponding entries of the estimated coefficient tensor, which can be interpreted as a measure of the importance of the edge in the network. Based on this statistic, we plot the position of the 10 most and least relevant edges in the network (in red and blue, respectively) in Fig 11. The picture has a heterogeneous structure: first, no single country seems to exert a key role, neither as exporter nor as importer; second, the most and least relevant edges are evenly distributed between the exporting and the importing side.

We study the effects of the propagation of a shock on a single and a group of edges in the network by means of the impulse response function obtained as follows. Define the reverse of the vectorization operator $\mathrm{vec}(\cdot)$ by $\mathrm{vecr}(\cdot)$ and let $\tilde{E}$ be a binary matrix of shocks such that each non zero entry (i,j) of $\tilde{E}$ corresponds to a unitary shock on the edge (i,j). Then the matrix-valued impulse response function is obtained from the recursion:

$$Y_1 = \mathcal{B} \times_3 \mathrm{vec}\left(\tilde{E}\right) = \mathrm{vecr}\left(\mathbf{B}'_{(3)} \cdot \mathrm{vec}\left(\tilde{E}\right)\right) \tag{51}$$

$$Y_2 = \mathcal{B} \times_3 \mathrm{vec}\left(\mathrm{vecr}\left(\mathbf{B}'_{(3)} \cdot \mathrm{vec}\left(\tilde{E}\right)\right)\right) = \mathrm{vecr}\left(\mathbf{B}'_{(3)} \cdot \mathbf{B}'_{(3)} \cdot \mathrm{vec}\left(\tilde{E}\right)\right) \tag{52}$$

$$= \mathrm{vecr}\left([\mathbf{B}'_{(3)}]^2 \cdot \mathrm{vec}\left(\tilde{E}\right)\right), \tag{53}$$

which, for the horizon $h > 0$, generalizes to:

$$Y_h = \mathrm{vecr}\left([\mathbf{B}'_{(3)}]^h \cdot \mathrm{vec}\left(\tilde{E}\right)\right). \tag{54}$$

This equation shows that it is possible to study the joint effect that a contemporaneous shock on a subset of the edges of the network has on the whole network over time.

Fig. 12 and 13, respectively, plot the impulse response function of a unitary shock on the 10 most relevant and the 10 least relevant edges (determined by ranking according to the sum of the absolute values of the entries of the estimated coefficient tensor), for $h = 1, \ldots, 14$ periods. Figs. 14-15 show the effects of a unitary shock to the most and least influential edges, respectively. We find that:

- the effects are remarkably different: both the magnitude and the persistence of the impact of a shock to the most relevant edges is significantly greater than that obtained by hitting the least relevant edges;
- with reference to figs. 14-15, as in the previous case, a shock to the most relevant edge is more persistent than a shock on the least relevant and the magnitude is higher. However, compared to the effects of a shock on 10 edges, both persistence and magnitude are remarkably lower;
- a shock to a single edge affects almost all the others because of the high degree of interconnection of the network, which is responsible for the propagation both in the space (i.e. cross-section) and over time.

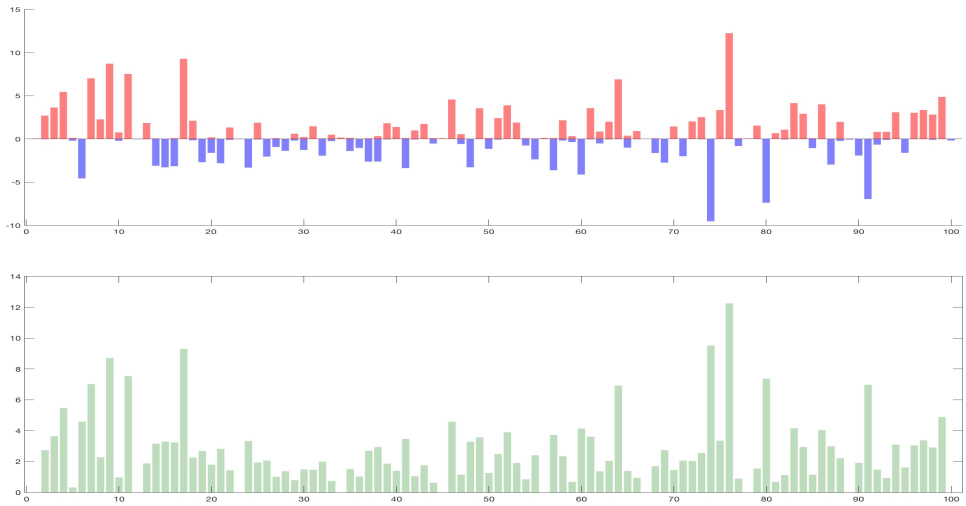

Figure 10: Sum of positive entries (*red,top*), negative entries (*blue,top*) and of absolute values of all entries (*dark green,bottom*) of the estimated coefficient tensor (*y-axis*), per each edge (*x-axis*).

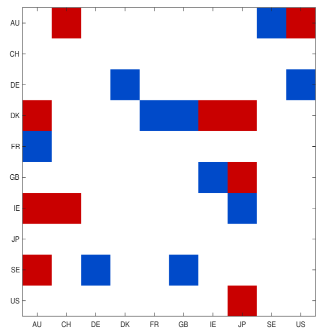

Figure 11: Position in the network of the 10 most relevant (*red*) and least relevant (*blue*) edges, according to the sum of the absolute values. Countries' labels on both axes.

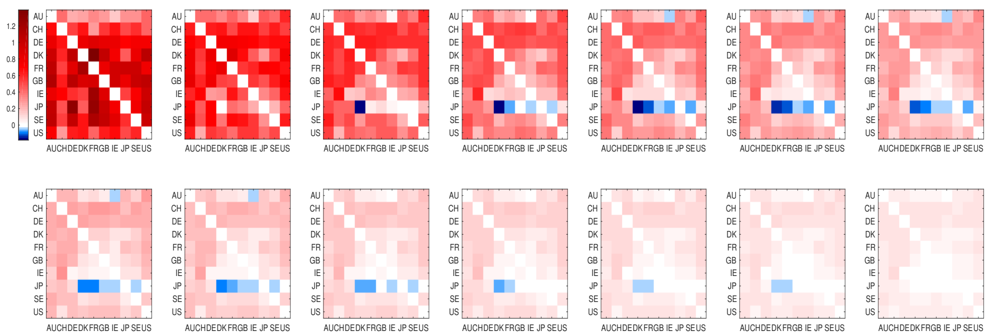

Figure 12: Impulse response for $h = 1, \dots, 14$ periods. Unitary shock on the 10 most relevant edges (sum of absolute values of all coefficients). Countries' labels on both axes.

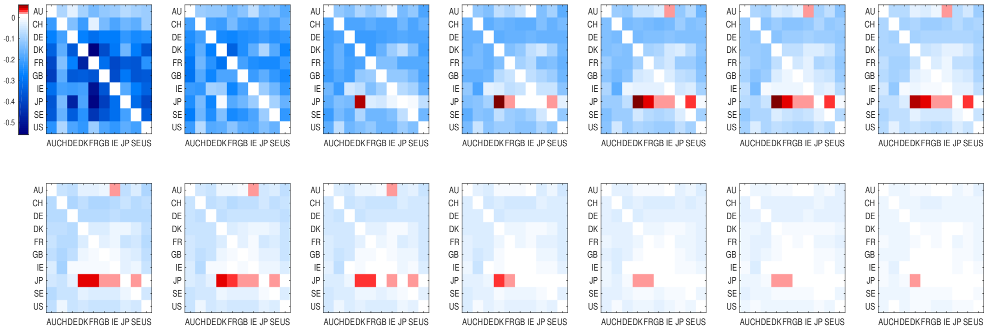

Figure 13: Impulse response for $h = 1, \dots, 14$ periods. Unitary shock on the 10 least relevant edges (sum of absolute values of all coefficients). Countries' labels on both axes.

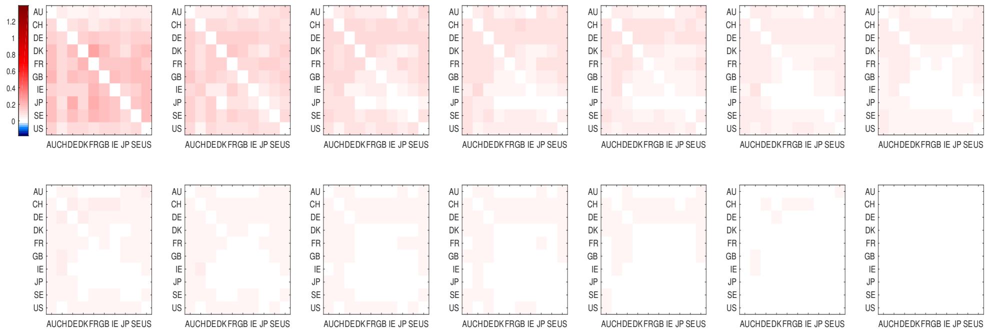

Figure 14: Impulse response for $h = 1, \dots, 14$ periods. Unitary shock on the most relevant edge (sum of absolute values of all coefficients). Countries' labels on both axes.

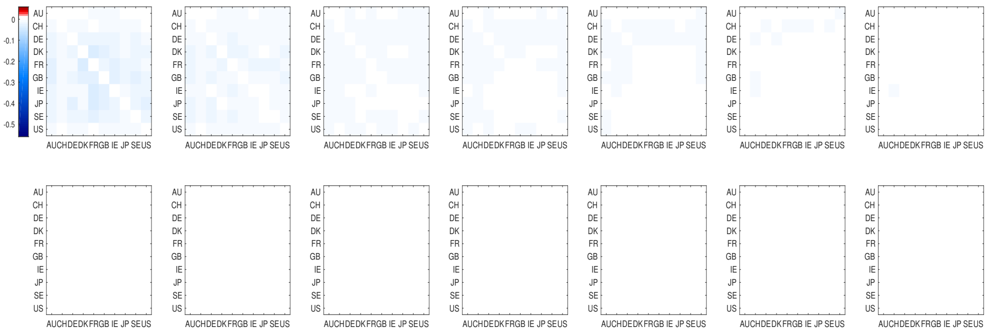

Figure 15: Impulse response for $h = 1, \dots, 14$ periods. Unitary shock on the least relevant edge (sum of absolute values of all coefficients). Countries' labels on both axes.

6 Conclusions

We defined a new statistical framework for dynamic tensor regression. It is a generalisation of many models frequently used in time series analysis, such as VAR, panel VAR, SUR and matrix regression models. The PARAFAC decomposition of the tensor of regression coefficients allows to reduce the dimension of the parameter space but also permits to choose flexible multivariate prior distributions, instead of multidimensional ones. Overall, this allows to encompass sparsity beliefs and to design efficient algorithm for posterior inference.

We tested the Gibbs sampler algorithm on synthetic matrix-variate datasets with matrices of different sizes, obtaining good results in terms of both the estimation of the true value of the parameter and the efficiency.

The proposed methodology has been applied to the analysis of temporal evolution of a subset of the international trade networks. We found evidence of (i) wide heterogeneity in the sign and magnitude of the estimated coefficients; (ii) stationarity of the network process.

Acknowledgements

This research has benefited from the use of the Scientific Computation System of Ca' Foscari University of Venice (SCSCF) for the computational for the implementation of the estimation procedure.

References

Abraham, R., Marsden, J. E., and Ratiu, T. (2012). *Manifolds, tensor analysis, and applications.* Springer Science & Business Media.

Adler, R., Bazin, M., and Schiffer, M. (1975). *Introduction to general relativity.* McGraw-Hill New York.

Aldasoro, I. and Alves, I. (2016). Multiplex interbank networks and systemic importance: an application to European data. *Journal of Financial Stability.*

Anacleto, O. and Queen, C. (2017). Dynamic chain graph models for time series network data. *Bayesian Analysis*, 12(2):491–509.

Arashi, M. (2017). Some theoretical results on tensor elliptical distribution. *arXiv preprint arXiv:1709.00801.*

Aris, R. (2012). *Vectors, tensors and the basic equations of fluid mechanics.* Courier Corporation.

Balazsi, L., Matyas, L., and Wansbeek, T. (2015). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, pages 1–23.

Barrat, A., Fernandez, B., Lin, K. K., and Young, L.-S. (2013). Modeling temporal networks using random itineraries. *Physical Review Letters*, 110(15).

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

Canova, F. and Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics*, 120(2):327–359.

Canova, F. and Ciccarelli, M. (2009). Estimating multicountry VAR models. *International Economic Review*, 50(3):929–959.

Canova, F. and Ciccarelli, M. (2013). *Panel vector autoregressive models: a survey*, volume 32, chapter 12, pages 205–246. Emerald Group Publishing Limited, VAR models in macroeconomics – new developments and applications: essays in honor of Christopher A. Sims edition.

Canova, F., Ciccarelli, M., and Ortega, E. (2007). Similarities and convergence in G-7 cycles. *Journal of Monetary Economics*, 54(3):850–878.

Canova, F., Ciccarelli, M., and Ortega, E. (2012). Do institutional changes affect business cycles? evidence from Europe. *Journal of Economic Dynamics and Control*, 36(10):1520–1533.

Carriero, A., Kapetanios, G., and Marcellino, M. (2016). Structural analysis with multivariate autoregressive index models. *Journal of Econometrics*, 192(2):332–348.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrica*, 35:283–319.

Carvalho, C. M., Massam, H., and West, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659.

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

Carvalho, C. M. and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis*, 2(1):69–97.

Cichocki, A. (2014). Era of Big data processing: a new approach via tensor networks and tensor decompositions. *arXiv preprint arXiv:1403.2048*.

Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S.-i. (2009). *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, 20:927–960.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317.

Ding, S. and Cook, R. D. (2016). Matrix-variate regressions and envelope models. *arXiv preprint arXiv:1605.01485*.

Dobra, A. (2015). *Handbook of Spatial Epidemiology*, chapter Graphical Modeling of Spatial Health Data. Chapman & Hall /CRC, first edition.

Durante, D. and Dunson, D. B. (2014). Nonparametric Bayesian dynamic modelling of relational data. *Biometrika*, 101(4):883–898.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, pages 251–276.

Fagiolo, G., Reyes, J., and Schiavo, S. (2009). World-trade web: topological properties, dynamics, and evolution. *Physical Review E*, 79(3):036115.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *Journal of Machine Learning Research*, 18(79):1–31.

Gupta, A. K. and Nagar, D. K. (1999). *Matrix variate distributions.* CRC Press.

Hackbusch, W. (2012). *Tensor spaces and numerical tensor calculus.* Springer Science & Business Media.

Harrison, J. and West, M. (1999). *Bayesian forecasting & dynamic models.* Springer.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16, 1- 84.*

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26):10570–10575.

Hoff, P. D. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics*, 9(3):1169–1193.

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3):97–125.

Holme, P. and Saramäki, J. (2013). *Temporal networks.* Springer.

Hung, H. and Wang, C.-C. (2013). Matrix variate logistic regression model with application to EEG data. *Biostatistics*, 14(1):189–202.

Imaizumi, M. and Hayashi, K. (2016). Doubly decomposing nonparametric tensor regression. In *International Conference on Machine Learning*, pages 727–736.

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

Kharrazi, A., Rovenskaya, E., and Fath, B. D. (2017). Network structure impacts global commodity trade growth and resilience. *PloS one*, 12(2):e0171184.

Kiers, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chameometrics*, 14(3):105–122.

Kolda, T. G. (2006). Multilinear operators for higher-order decompositions. Technical report, Sandia National Laboratories.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

Kostakos, V. (2009). Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, 388(6):1007–1023.

Kroonenberg, P. M. (2008). *Applied multiway data analysis.* John Wiley & Sons.

Kruijer, W., Rousseau, J., and Van Der Vaart, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electronic Journal of Statistics*, 4:1225–1257.

Lee, N. and Cichocki, A. (2016). Fundamental tensor operations for large-scale data analysis in tensor train formats. *arXiv preprint arXiv:1405.7786*.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146.

Lovelock, D. and Rund, H. (1989). *Tensors, differential forms, and variational principles*. Courier Corporation.

Malvern, L. E. (1986). *Introduction to the mechanics of a continuous medium*. Englewood.

Manceur, A. M. and Dutilleul, P. (2013). Maximum likelihood estimation for the tensor normal distribution: algorithm, minimum sample size, and empirical bias and dispersion. *Journal of Computational and Applied Mathematics*, 239:37–49.

Meyfroidt, P., Rudel, T. K., and Lambin, E. F. (2010). Forest transitions, trade, and the global displacement of land use. *Proceedings of the National Academy of Sciences*, 107(49):20917–20922.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Galin, J. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, chapter 5. Chapman & Hall /CRC.

Ohlson, M., Ahmad, M. R., and Von Rosen, D. (2013). The multilinear normal distribution: introduction and some basic properties. *Journal of Multivariate Analysis*, 113:37–47.

Pan, R. (2014). Tensor transpose and its properties. *arXiv preprint arXiv:1411.1503*.

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

Poledna, S., Molina-Borboa, J. L., Martínez-Jaramillo, S., Van Der Leij, M., and Thurner, S. (2015). The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability*, 20:70–81.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical recipes: the art of scientic computing*. Cambridge University Press.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer.

Schotman, P. and Van Dijk, H. K. (1991). A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics*, 49(1-2):195–238.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., and White, D. R. (2009). Economic networks: the new challenges. *Science*, 325(5939):422–425.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Squartini, T., Fagiolo, G., and Garlaschelli, D. (2011). Randomizing world trade. i. a binary network analysis. *Physical Review E*, 84(4):046117.

Synge, J. L. and Schild, A. (1969). *Tensor calculus*. Courier Corporation.

Viroli, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, 21(4):511–522.

Viroli, C. (2012). On matrix-variate regression analysis. *Journal of Multivariate Analysis*, 111:296–309.

Viroli, C. and Anderlucci, L. (2013). Modelling longitudinal data through matrix-variate normal mixtures. In *Advances in Latent Variables - Methods, Models and Applications*.

Wang, H. and West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, 96(4):821–834.

Xu, T., Yin, Z., Siliang, T., Jian, S., Fei, W., and Yueting, Z. (2013). *Logistic tensor regression for classification*, volume 7751, chapter Intelligent science and intelligent data engineering, pages 573–581. Springer.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zhang, X., Li, L., Zhou, H., and Shen, D. (2014). Tensor generalized estimating equations for longitudinal imaging analysis. *arXiv preprint arXiv:1412.6592*.

Zhao, Q., Zhang, L., and Cichocki, A. (2013). A tensor-variate Gaussian process for classification of multidimensional structured data. In *Twenty-seventh AAAI conference on artificial intelligence*.

Zhao, Q., Zhou, G., Zhang, L., and Cichocki, A. (2014). Tensor-variate Gaussian processes regression and its application to video surveillance. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1265–1269. IEEE.

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.

Zhou, J., Bhattacharya, A., Herring, A. H., and Dunson, D. B. (2015). Bayesian factorizations of big sparse tensors. *Journal of the American Statistical Association*, 110(512):1562–1576.

Zhu, Z., Cerina, F., Chessa, A., Caldarelli, G., and Riccaboni, M. (2014). The rise of China in the international trade network: a community core detection approach. *PloS one*, 9(8):e105496.

A Background material on tensor calculus

A N-order tensor is an element of the tensor product of N vector spaces. Since there exists a isomorphism between two vector spaces of dimensions N and $M < N$, it is possible to define a one-to-one map between their elements, that is, between a N-order tensor and a M-order tensor. We call this tensor reshaping and give its formal definition below.

Definition A.1 (Tensor reshaping). *Let* $V_1, \ldots, V_N$ *and* $U_1, \ldots, U_M$ *be vector subspaces* $V_n, U_m \subseteq \mathbb{R}$ *and* $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N} = V_1 \otimes \ldots \otimes V_N$ *be a* N*-order real tensor of dimensions* $I_1, \ldots, I_N$. *Let* $(\mathbf{v}_1, \ldots, \mathbf{v}_N)$ *be a canonical basis of* $\mathbb{R}^{I_1 \times \ldots \times I_N}$ *and let* Π_S *be the projection defined as:*

$$\begin{aligned} \Pi_S : & V_1 \otimes \ldots \otimes V_N \to V_{s_1} \otimes \ldots \otimes V_{s_k} \\ & \mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_N \mapsto \mathbf{v}_{s_1} \otimes \ldots \otimes \mathbf{v}_{s_k} \,, \end{aligned}$$

with $S = \{s_1, \ldots, s_k\} \subset \{1, \ldots, N\}$. *Let* $(S_1, \ldots, S_M)$ *be a partition of* $\{1, \ldots, N\}$. *The* $(S_1, \ldots, S_M)$ *tensor reshaping of* $\mathcal{X}$ *is defined as:*

$$\begin{aligned} \mathcal{X}_{(S_1, \ldots, S_M)} &= (\Pi_{S_1} \mathcal{X}) \otimes \ldots \otimes (\Pi_{S_M} \mathcal{X}) \\ &\in \left(\bigotimes_{s \in S_1} V_s \right) \otimes \ldots \otimes \left(\bigotimes_{s \in S_M} V_s \right) \\ &= U_1 \otimes \ldots \otimes U_M \,. \end{aligned}$$

It can be proved that the mapping is an isomorphism between $V_1 \otimes \ldots \otimes V_N$ *and* $U_1 \otimes \ldots \otimes U_M$.

The operation of converting a tensor into a matrix can be seen as a particular case of tensor reshaping, where a N-order tensor is mapped to a 2-order tensor. In practice, it consists in choosing the modes of the array to map with the rows and columns of the resulting matrix, then permuting the tensor and reshaping it, accordingly. The formal follows.

Definition A.2. *Let* $\mathcal{X}$ *be a* N *order tensor with dimensions* $I_1, \ldots, I_N$. *Let the ordered sets* $\mathscr{R} = \{r_1, \ldots, r_L\}$ *and* $\mathscr{C} = \{c_1, \ldots, c_M\}$ *be a partition of* $\mathbf{N} = \{1, \ldots, N\}$ *and let* $I_{\mathbf{N}} = \{I_1, \ldots, I_N\}$. *The matricized tensor is specified by:*

$$\mathbf{X}_{(\mathscr{R} \times \mathscr{C} : I_{\mathbf{N}})} \in \mathbb{R}^{J \times K} \qquad J = \prod_{n \in \mathscr{R}} I_n \quad K = \prod_{n \in \mathscr{C}} I_n \,. \tag{A.1}$$

Indices of $\mathscr{R}, \mathscr{C}$ *are mapped to the rows and the columns, respectively. More precisely:*

$$\left(\mathbf{X}_{(\mathscr{R} \times \mathscr{C} : I_{\mathbf{N}})} \right)_{j,k} = \mathcal{X}_{i_1, i_2, \ldots, i_N} \tag{A.2}$$

with:

$$j = 1 + \sum_{l=1}^{L} \left[(i_{r_l} - 1) \prod_{l'=1}^{l-1} I_{r'_l} \right] \qquad k = 1 + \sum_{m=1}^{M} \left[(i_{c_m} - 1) \prod_{m'=1}^{m-1} I_{c'_m} \right] \tag{A.3}$$

We introduce two multilinear operators acting on tensors, see Kolda (2006) for more details.

Definition A.3 (Tucker operator). *Let* $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ *and* $\mathbf{N} = \{1, \ldots, N\}$. *Let* $\{A_n\}_n$ *be a collection of* N *matrices such that* $A_n \in \mathbb{R}^{I_n \times J_n}$ *for* $n \in \mathbf{N}$. *The Tucker operator is defined as:*

$$[\![\mathcal{Y}; A_1, \ldots, A_N]\!] = \mathcal{Y} \times_1 A_1 \times_2 A_2 \ldots \times_N A_N \,, \tag{A.4}$$

and the resulting tensor has size $I_1 \times \ldots \times I_N$.

Definition A.4 (Kruskal operator). *Let* $\mathbf{N} = \{1, \ldots, N\}$ *and* $\{A_n\}_n$ *be a collection of* N *matrices such that* $A_n \in \mathbb{R}^{I_n \times R}$ *for* $n \in \mathbf{N}$. *Let* $\mathcal{I}$ *be the identity tensor of size* $R \times \ldots \times R$, *i.e. a tensor having ones along the superdiagonal and zeros elsewhere. The Kruskal operator is defined as:*

$$\mathcal{X} = [\![A_1, \ldots, A_N]\!] = [\![\mathcal{I}; A_1, \ldots, A_N]\!] \,, \tag{A.5}$$

with $\mathcal{X}$ a tensor of size $I_1 \times \ldots \times I_N$. An alternative representation is obtained by defining $\mathbf{a}_n^{(r)}$ the r-th column of the matrix A_n and using the outer product:

$$\mathcal{X} = [\![A_1, \ldots, A_N]\!] = \sum_{r=1}^{R} \mathbf{a}_1^{(r)} \circ \ldots \circ \mathbf{a}_N^{(r)} . \tag{A.6}$$

By exploiting the Khatri-Rao product $\odot$ (i.e. the columnwise Kronecker product for $A \in \mathbb{R}^{I \times K}$, $B \in \mathbb{R}^{J \times K}$ defined as $A \odot B = [\mathbf{a}_{:,1} \otimes \mathbf{b}_{:,1}, \ldots, \mathbf{a}_{:,K} \otimes \mathbf{b}_{:,K}]$) in combination with the mode-n matricization and the vecotrization operators, we get the following additional representations of $\mathcal{X} = [\![A_1, \ldots, A_N]\!]$:

$$\mathbf{X}_{(n)} = A_n \left(A_N \odot \ldots \odot A_{n+1} \odot A_{n-1} \odot \ldots \odot A_1 \right)' \tag{A.7}$$

$$\operatorname{vec}(\mathcal{X}) = (A_N \odot \ldots \odot A_1) \mathbf{1}_R , \tag{A.8}$$

where $\mathbf{1}_R$ is a vector of ones of length R.

Proposition A.1 (4.3 in Kolda (2006)). *Let $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ and $\mathbf{N} = \{1, \ldots, N\}$ and let $A \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathbf{N}$. If $\mathscr{R} = \{r_1, \ldots, r_L\}$ and $\mathscr{C} = \{c_1, \ldots, c_M\}$ partition $\mathbf{N}$, then:*

$$\mathcal{X} = [\![\mathcal{Y}; A_1, \ldots, A_N]\!] \iff \mathbf{X}_{(\mathscr{R} \times \mathscr{C} : J_{\mathbf{N}})} = \left(A^{(r_L)} \otimes \ldots \otimes A^{(r_1)} \right) \mathbf{Y}_{(\mathscr{R} \times \mathscr{C} : J_{\mathbf{N}})} \left(A^{(c_M)} \otimes \ldots \otimes A^{(c_1)} \right)' \tag{A.9}$$

where $\mathcal{X} = [\![\mathcal{Y}; A_1, \ldots, A_N]\!] = \mathcal{Y} \times_1 A_1 \times_2 A^{(2)} \ldots \times_n A_N$ denotes the Tucker product between the tensor $\mathcal{Y}$ and the collection of matrices $\{A_n\}_{n=1}^N$. The Kruskal operator is a special case of the Tucker operator, obtained when the tensor $\mathcal{Y} = \mathcal{I}$ is an identity tensor of dimensions $R \times \ldots \times R$ and the matrices $\{A_n\}_{n=1}^N$ have dimension $A_n \in \mathbb{R}^{I_n \times R}$. Therefore, we can represent the product using the outer product representation, as follows. Consider the collection of vectors $\{\mathbf{a}^{(n)}\}_{n=1}^N$, of length $\mathbf{a}^{(n)} \in \mathbb{R}^{I_n}$, formed by the columns of the matrices A_n. Then:

$$\mathcal{X} = [\![\mathcal{I}; A_1, \ldots, A_N]\!] = \circ_{n=1}^N \mathbf{a}^{(n)} \iff \mathbf{X}_{(\mathscr{R} \times \mathscr{C} : J_{\mathbf{N}})} = \left(\mathbf{a}^{(r_L)} \otimes \ldots \otimes \mathbf{a}^{(r_1)} \right) \mathbf{I}_{(\mathscr{R} \times \mathscr{C} : J_{\mathbf{N}})} \left(\mathbf{a}^{(c_M)} \otimes \ldots \otimes \mathbf{a}^{(c_1)} \right)' . \tag{A.10}$$

Remark A.1 (Contracted product – vectorization). *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N \times J_{N+1} \times \ldots \times J_{N+P}}$. Let $(\mathscr{S}_1, \mathscr{S}_2)$, with $\mathscr{S}_1 = \{1, \ldots, N\}$, $\mathscr{S}_2 = \{N+1, \ldots, N+P\}$, be a partition of $\{1, \ldots, N+P\}$. The following results hold:*

a) if $P = 0$ and $I_n = J_n$ for $n = 1, \ldots, N$, then:

$$\mathcal{X} \times^{1 \ldots N} \mathcal{Y} = \langle \mathcal{X}, \mathcal{Y} \rangle = \operatorname{vec}(\mathcal{X})' \cdot \operatorname{vec}(\mathcal{Y}) \quad \in \mathbb{R} . \tag{A.11}$$

b) if $P > 0$ and $I_n = J_n$ for $n = 1, \ldots, N$, then:

$$\mathcal{X} \times^{1 \ldots N} \mathcal{Y} = \operatorname{vec}(\mathcal{X}) \times^1 \mathcal{Y}_{(\mathscr{S}_1, \mathscr{S}_2)} \quad \in \mathbb{R}^{j_1 \times \ldots \times j_P} \tag{A.12}$$

$$\mathcal{Y} \times^{1 \ldots N} \mathcal{X} = \mathcal{Y}_{(\mathscr{S}_1, \mathscr{S}_2)} \times^1 \operatorname{vec}(\mathcal{X}) \quad \in \mathbb{R}^{j_1 \times \ldots \times j_P} . \tag{A.13}$$

c) if $P = N$ and $I_n = J_n = J_{N+n}$, $n = 1, \ldots, N$, then:

$$\mathcal{X} \times^{1 \ldots N} \mathcal{Y} \times^{1 \ldots N} \mathcal{X} = \operatorname{vec}(\mathcal{X})' \cdot \mathbf{Y}_{(\mathscr{R} \times \mathscr{C})} \cdot \operatorname{vec}(\mathcal{X}) \quad \in \mathbb{R} . \tag{A.14}$$

Proof. Case a). By definition of contracted product and tensor scalar product:

$$\mathcal{X} \times^{1 \ldots N} \mathcal{Y} = \sum_{i_1=1}^{I_1} \ldots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1, \ldots, i_N} \cdot \mathcal{Y}_{i_1, \ldots, i_N}$$

$$= \sum_{i_1,\dots,i_N} \mathcal{X}_{i_1,\dots,i_N} \cdot \mathcal{Y}_{i_1,\dots,i_N} = \langle \mathcal{X}, \mathcal{Y} \rangle = \mathrm{vec}\,(X)' \cdot \mathrm{vec}\,(Y)\ .$$

Case b). Define $I^* = \prod_{n=1}^{N} I_n$ and $k = 1+\sum_{j=1}^{N}(i_j-1)\prod_{m=1}^{j-1} I_m$. By definition of contracted product and tensor scalar product:

$$\begin{aligned}
\mathcal{X} \times^{1\dots N} \mathcal{Y} &= \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1,\dots,i_N} \cdot \mathcal{Y}_{i_1,\dots,i_N,j_{N+1},\dots,j_{N+P}} \\
&= \sum_{k=1}^{I^*} \mathcal{X}_k \cdot \mathcal{Y}_{k,j_{N+1},\dots,j_{N+P}}\ .
\end{aligned}$$

Notice that the one-to-one correspondence established by the mapping between k and $(i_1,\dots,i_N)$ corresponds to that of the vectorization of a tensor of size N and dimensions $I_1,\dots,I_N$. Moreover, it also corresponds to the mapping established by the tensor reshaping of a tensor of order $N+P$ with dimensions $I_1,\dots,I_N,J_{N+1},\dots,J_{N+P}$ into another tensor of order $1+P$ and dimensions $I^*,J_{N+1},\dots,J_{N+P}$. Define $S=\{1,\dots,N\}$, such that $(S,N+1,\dots,N+P)$ is a partition of $\{1,\dots,N+P\}$. Then:

$$\mathcal{X} \times^{1\dots N} \mathcal{Y} = \mathrm{vec}\,(X) \times^1 \mathcal{Y}_{(S,N+1,\dots,N+P)}\ .$$

Similarly, defining $S=\{P+1,\dots,N+P\}$ yields the second part of the result.

Case c). We follow the same strategy adopted in case b). Define $S_1=\{1,\dots,N\}$ and $S_2=\{N+1,\dots,N+P\}$, such that $(S-1,S_2)$ is a partition of $\{1,\dots,N+P\}$. Let k,k' be defined as in case b). Then:

$$\begin{aligned}
\mathcal{X} \times^{1\dots N} \mathcal{Y} \times^{1\dots N} \mathcal{X} &= \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} \sum_{i'_1=1}^{I_1} \dots \sum_{i'_N=1}^{I_N} \mathcal{X}_{i_1,\dots,i_N} \cdot \mathcal{Y}_{i_1,\dots,i_N,i'_1,\dots,i'_N} \cdot \mathcal{X}_{i'_1,\dots,i'_N} \\
&= \sum_{k=1}^{I^*} \sum_{i'_1=1}^{I_1} \dots \sum_{i'_N=1}^{I_N} \mathcal{X}_k \cdot \mathcal{Y}_{k,i'_1,\dots,i'_N} \cdot \mathcal{X}_{i'_1,\dots,i'_N} \\
&= \sum_{k=1}^{I^*} \sum_{k'=1}^{I^*} \mathcal{X}_k \cdot \mathcal{Y}_{k,k'} \cdot \mathcal{X}_{k'} \\
&= \mathrm{vec}\,(X)' \cdot \mathcal{Y}_{(S_1,S_2)} \cdot \mathrm{vec}\,(X)\ .
\end{aligned}$$

□

Relation between the matricization of a tensor resulting from the outer product of matrices and the Kronecker product.

Remark A.2 (Kronecker - matricization)**.** *Let $X_1,\dots,X_N$ be square matrices of size $I_n \times I_n$, $n=1,\dots,N$ and let $\mathcal{X} = X_1 \circ \dots \circ X_N$ denote the N-order tensor with dimensions $(J_1,\dots J_{2N}) = (I_1,\dots,I_N,I_1,\dots,I_N)$ obtained as the outer product of the matrices $\{U_n\}$. Let $(\mathscr{S}_1,\mathscr{S}_2)$, with $\mathscr{S}_1=\{1,\dots,N\}$ and $\mathscr{S}_2=\{N+1,\dots,N\}$, be a partition of $I_{\mathbf{N}}=\{1,\dots,2N\}$. Then:*

$$\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)} = \mathbf{X}_{(\mathscr{R}\times\mathscr{C}:I_{\mathbf{N}})} = (X_N \otimes \dots \otimes X_1)\,. \tag{A.15}$$

Proof. Use the pair of indices (i_n,i'_n) for the entries of the matrix X_n, $n=1,\dots,N$. By definition of outer product:

$$(X_1 \circ \dots \circ X_N)_{i_1,i_2,\dots,i_N,i'_1,i'_2,\dots,i'_N} = (X_1)_{i_1,i'_1} \cdot (X_2)_{i_2,i'_2} \cdots (X_N)_{i_N,i'_N}\ .$$

From the definition of matricization, $\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)} = \mathbf{X}_{(\mathscr{R}\times\mathscr{C}:I_{\mathbf{N}})}$. Moreover:

$$\left(\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)}\right)_{h,k} = \mathcal{X}_{i_1,\ldots,i_{2N}}$$

with:

$$h = \sum_{p=1}^{N}(i_{S_{1,p}}-1)\prod_{q=1}^{p-1} J_{S_{1,p}} \qquad k = \sum_{p=1}^{N}(i_{S_{2,p}}-1)\prod_{q=1}^{p-1} J_{S_{2,p}}\,.$$

By definition of the Kronecker product we have: that the entry (h',k') of $(X_N \otimes \ldots \otimes X_1)$ is given by:

$$(X_N \otimes \ldots \otimes X_1)_{h',k'} = (X_N)_{i'_N,i'_N} \cdots (X_1)_{i_1,i'_1}$$

where:

$$h' = \sum_{p=1}^{N}(i_{S_{1,p}}-1)\prod_{q=1}^{p-1} J_{S_{1,p}} \qquad k' = \sum_{p=1}^{N}(i_{S_{2,p}}-1)\prod_{q=1}^{p-1} J_{S_{2,p}}\,.$$

Since $h = h'$ and $k = k'$ and the associated elements of $\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)}$ and $(X_N \otimes \ldots \otimes X_1)$ are the same, the result follows. □

Remark A.3. *Let $\mathcal{X}$ be a N-order tensor of dimensions $I_1 \times \ldots \times I_N$ and let $I^* = \prod_{i=1}^N I_i$. Then there exists a vec-permutation (or commutation) matrix $K_{1\to n}$ of size $I^* \times I^*$ such that:*

$$K_{1\to n}\operatorname{vec}(\mathcal{X}) = K_{1\to n}\operatorname{vec}\left(\mathbf{X}_{(1)}\right) = \operatorname{vec}\left(\mathbf{X}_{(n)}\right)\,. \tag{A.16}$$

Moreover, it holds:

$$\operatorname{vec}\left(\mathbf{X}_{(n)}\right) = \operatorname{vec}\left(\mathbf{X}_{(1)}^{T_\sigma}\right) = \operatorname{vec}\left(\mathcal{X}^{T_\sigma}\right)\,, \tag{A.17}$$

where

$$\mathbf{X}_{(1)}^{T_\sigma} = \left(\mathcal{X}^{T_\sigma}\right)_{(1)} = \mathbf{X}_{(n)}\,, \tag{A.18}$$

is the mode-1 matricization of the transposed tensor $\mathcal{X}^{T_\sigma}$ according to the permutation σ which exchanges modes 1 and n, leaving the others unchanged. That is, for $i_j \in \{1,\ldots,I_j\}$ and $j = 1,\ldots,N$:

$$\sigma(i_j) = \begin{cases} 1 & j = n \\ n & j = 1 \\ i_j & j \neq 1, n\,. \end{cases}$$

Remark A.4. *Let $\mathcal{X}$ be a N-order random tensor with dimensions $I_1,\ldots,I_N$ and let $\mathbf{N} = \{1,\ldots,N\}$ be partitioned by the index sets $\mathscr{R} = \{r_1,\ldots,r_m\} \subset \mathbf{D}$ and $\mathscr{C} = \{c_1,\ldots,c_p\} \subset \mathbf{N}$, i.e. $\mathbf{N} = \mathscr{R}\cup\mathscr{C}$, $\mathscr{R}\cap\mathscr{C} = \emptyset$ and $N = m+p$. Then:*

$$\mathcal{X} \sim \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{M}, U_1,\ldots,U_N) \iff \mathbf{X}_{(\mathscr{R}\times\mathscr{C})} \sim \mathcal{N}_{m,p}(\mathbf{M}_{(\mathscr{R}\times\mathscr{C})}, \Sigma_1, \Sigma_2)\,, \tag{A.19}$$

with:

$$\Sigma_1 = U_{r_m} \otimes \ldots \otimes U_{r_1} \qquad \Sigma_2 = U_{c_p} \otimes \ldots \otimes U_{c_1}\,. \tag{A.20}$$

Proof. We demonstrate the statement for $\mathscr{R} = \{n\}$, $n \in \mathbf{N}$, however the results follows from the same steps also in the general case $\#\mathscr{R} > 1$. The strategy it to demonstrate that the probability density functions of the two distributions coincide. To this aim consider separately the exponent and the normalizing constant. Define $I_{-j} = \prod_{i=1,\,n\neq j}^N I_i$ and $I_{\mathbf{N}} = \{I_1,\ldots,I_N\}$, then for the normalizing constant we have:

$$(2\pi)^{-\frac{\prod_i I_i}{2}} |U_1|^{-\frac{I_{-1}}{2}} \cdots |U_n|^{-\frac{I_{-n}}{2}} \cdots |U_N|^{-\frac{I_{-N}}{2}} = \tag{A.21}$$

$$= (2\pi)^{-\frac{\prod_i I_i}{2}} |U_1|^{-\frac{I_{-1}}{2}} \cdots |U_{n-1}|^{-\frac{I_{-(n-1)}}{2}} \cdot |U_{n+1}|^{-\frac{I_{-(n+1)}}{2}} \cdots |U_N|^{-\frac{I_{-N}}{2}} \cdot |U_n|^{-\frac{I_{-n}}{2}}$$

$$= (2\pi)^{-\frac{\prod_i I_i}{2}} |U_N \otimes \ldots \otimes U_{n-1} \otimes U_{n+1} \otimes \ldots \otimes U_N|^{-\frac{n}{2}} \cdot |U_n|^{-\frac{I_{-n}}{2}} . \tag{A.22}$$

Concerning the exponent, let $\mathbf{i} = (i_1, \ldots, i_N)$ and, for ease of notation, define $\mathcal{Y} = \mathcal{X} - \mathcal{M}$ and $\mathcal{U} = (U_N^{-1} \circ \ldots \circ U_1^{-1})$. By the definition of contracted product it holds:

$$\mathcal{Y} \times^{1\ldots N} \mathcal{U} \times^{1\ldots N} \mathcal{Y} = \tag{A.23}$$

$$= \sum_{i_1,\ldots,i_N} \sum_{i'_1,\ldots,i'_N} y_{i_1,\ldots,i_n,\ldots,i_N} \cdot u^{-1}_{i_1,i'_1} \cdots u^{-1}_{i_n,i'_n} \cdots u^{-1}_{i_N,i'_N} \cdot y_{i'_1,\ldots,i_n,\ldots,i'_N} .$$

Define $\mathbf{j} = \sigma(\mathbf{i})$, where σ is the permutation defined above exchanging i_1 with i_n, $n \in \{2, \ldots, N\}$. Then the previous equation can be rewritten as:

$$= \sum_{j_1,\ldots,j_N} \sum_{j'_1,\ldots,j'_N} y_{j_n,\ldots,j_1,\ldots,i_N} \cdot u^{-1}_{j_n,j'_n} \cdots u^{-1}_{j_1,j'_1} \cdots u^{-1}_{i_N,i'_N} \cdot y_{j'_n,\ldots,j'_1,\ldots,i'_N}$$

$$= \mathcal{Y}^\sigma \times^{1\ldots N} \left(U_1^{-1} \circ \ldots \circ U_N^{-1}\right)^\sigma \times^{1\ldots N} \mathcal{Y}^\sigma ,$$

where $\mathcal{Y}^\sigma$ is the transpose tensor of $\mathcal{Y}$ (see Pan (2014)) obtained by permuting the first and the n-th modes and similarly for the 6-order tensor $(U_1^{-1} \circ \ldots \circ U_N^{-1})^\sigma$. Let $(\mathscr{S}_1, \mathscr{S}_2)$, with $\mathscr{S}_1 = \{1, \ldots, N\}$ and $\mathscr{S}_2 = \{N+1, \ldots, 2N\}$, be a partition of $\{1, \ldots, 2N\}$. By vectorizing eq. (A.23) and exploiting the results in A.1 and A.2, we have:

$$\begin{aligned}
\mathcal{Y} \times^{1\ldots N} \mathcal{U} \times^{1\ldots N} \mathcal{Y} &= \mathrm{vec}(\mathcal{Y})' \cdot \mathcal{U}_{(\mathscr{S}_1,\mathscr{S}_2)} \cdot \mathrm{vec}(\mathcal{Y}) &\quad (A.24)\\
&= \mathrm{vec}(\mathcal{Y})' \cdot \left(U_N^{-1} \otimes \ldots \otimes U_n^{-1} \otimes \ldots \otimes U_1^{-1}\right) \cdot \mathrm{vec}(\mathcal{Y})\\
&= \mathrm{vec}(\mathcal{Y}^\sigma)' \cdot \left(U_N^{-1} \otimes \ldots \otimes U_1^{-1} \otimes U_n^{-1}\right) \cdot \mathrm{vec}(\mathcal{Y}^\sigma)\\
&= \mathrm{vec}\left(\mathbf{Y}_{(n)}\right)' \cdot \left(U_N^{-1} \otimes \ldots \otimes U_1^{-1} \otimes U_n^{-1}\right) \cdot \mathrm{vec}\left(\mathbf{Y}_{(n)}\right)\\
&= \mathrm{vec}\left(\mathbf{Y}_{(n)}\right)' \cdot \mathrm{vec}\left(U_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot \left[U_N^{-1} \otimes \ldots \otimes U_1^{-1}\right]\right)\\
&= \mathrm{tr}\left(\mathbf{Y}'_{(n)} \cdot U_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot \left[U_N^{-1} \otimes \ldots \otimes U_1^{-1}\right]\right)\\
&= \mathrm{tr}\left(\left[U_N^{-1} \otimes \ldots \otimes U_1^{-1}\right] \left[\mathbf{X}_{(n)} - \mathbf{M}_{(n)}\right]' \cdot U_n^{-1} \cdot \left[\mathbf{X}_{(n)} - \mathbf{M}_{(n)}\right]\right) . &\quad (A.25)
\end{aligned}$$

Since the term in (A.21) and (A.24) are the normalizing constant and the exponent of the tensor normal distribution, whereas (A.22) and (A.25) are the corresponding expressions for the desired matrix normal distribution, the result is proved for the case $\#\mathscr{R} = 1$. In the general case $\#\mathscr{R} = r > 1$ the proof follows from the same reasoning, by substituting the permutation σ with another permutation σ' which exchanges the modes of the tensor such that the first r modes of the transpose tensor $\mathcal{Y}^{\sigma'}$ correspond to the elements of $\mathscr{R}$. □

B Proofs of the results in Section 2

Proof of result in Remark 2.3. By assuming $I_j = 1$, for $j = 1, \ldots, N$, in mode (16), then:

$$\mathcal{Y}_t, \mathcal{A}, \mathcal{E}_t \in \mathbb{R}^{I_1 \times \ldots \times I_N} \to \mathbb{R} \tag{B.1}$$

$$\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times J} \to \mathbb{R}^J \tag{B.2}$$

$$\mathcal{C} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times Q} \to \mathbb{R}^{Q} \tag{B.3}$$

where the matrix W_t has been removed as covariate. In order to keep it, it would be necessary either to vectorize it (then $\mathcal{D}$ would follow the same change as $\mathcal{B}$) or to assume an inner product (here $\mathcal{D}$ would reduce to a matrix of the same dimension of W_t). Notice that a N-order tensor whose modes have all unitary length is essentially a scalar. As a consequence, the error term distribution reduces to a univariate Gaussian, with 0 mean and variance σ^2. Finally, also the mode-3 product reduces to the standard inner product between vectors.

The PARAFAC(R) decomposition still holds in this case. Consider only $\mathcal{A}$ and $\mathcal{B}$, as the other tensors behave in the same manner. For ease of notation we drop the index t, since it does not affect the result:

$$\mathcal{A} = \sum_{r=1}^{R} \alpha_1^{(r)} \circ \ldots \circ \alpha_N^{(r)} = \sum_{r=1}^{R} \alpha_1^{(r)} \cdot \ldots \cdot \alpha_N^{(r)} = \sum_{r=1}^{R} \tilde{\alpha}_r = \bar{\alpha} \in \mathbb{R}. \tag{B.4}$$

Here, $\tilde{\alpha}_r = \prod_{j=1}^{N} \alpha_j^{(r)}$. Since each mode of $\mathcal{A}$ has unitary length, each of the marginals of the PARAFAC(R) decomposition is a scalar, therefore the outer product reduces to the ordinary product and the outcome is a scalar too (obtained by R sums of D products). Concerning $\mathcal{B}$, we apply the same way of reasoning, with the only exception that in this case one of the modes (the last, in the formulation of eq. (16)) has length $J > 1$, implying that the corresponding marginal is a vector of the same length. The result is a vector, as stated:

$$\begin{aligned} \mathcal{B} &= \sum_{r=1}^{R} \beta_1^{(r)} \circ \ldots \circ \beta_N^{(r)} \circ \boldsymbol{\beta}_{D+1}^{(r)} = \sum_{r=1}^{R} \beta_1^{(r)} \cdot \ldots \cdot \beta_N^{(r)} \cdot \boldsymbol{\beta}_{N+1}^{(r)} && \text{(B.5)} \\ &= \sum_{r=1}^{R} \tilde{\beta}_r \boldsymbol{\beta}_{N+1}^{(r)} = \boldsymbol{\beta} \in \mathbb{R}^{Q}, && \text{(B.6)} \end{aligned}$$

where $\tilde{\beta}_r = \prod_{j=1}^{N} \beta_j^{(r)}$. By an analogous proof, one gets:

$$\mathcal{C} = \boldsymbol{\gamma} \in \mathbb{R}^{J} . \tag{B.7}$$

which completes the proof. □

Proof of result in Remark 2.5. Without loss of generality, let $J_j = 1$, for $j = 2, \ldots, N$ in model (16), then:

$$\begin{aligned} &\mathcal{Y}_t, \mathcal{A}, \mathcal{E}_t \in \mathbb{R}^{I_1 \times \ldots \times I_N} \to \mathbb{R}^{m} && \text{(B.8)} \\ &\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times J} \to \mathbb{R}^{m \times J} && \text{(B.9)} \\ &\mathcal{C} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times Q} \to \mathbb{R}^{m \times Q} && \text{(B.10)} \\ &\mathcal{D} \in \mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times K \times I_{n+1} \ldots \times I_N} \to \mathbb{R}^{m \times K}, && \text{(B.11)} \end{aligned}$$

where it is necessary to assume that $W_t \in \mathbb{R}^{m \times K}$. The two mode-$N+1$ products become mode-2 products and the distribution of the error term reduces to the multivariate (n-dimensional) Gaussian, with a unique covariance matrix ($m \times m$).

As the PARAFAC(R) approximation is concerned, the result for $\mathcal{A}$ follows from the second part of the previous proof and yields $\mathcal{A} = \boldsymbol{\alpha} \in \mathbb{R}^{m}$. For the remaining tensors, it holds (dropping the index for notational ease):

$$\mathcal{B} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \beta_2^{(r)} \circ \ldots \circ \beta_N^{(r)} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \left(\beta_2^{(r)} \cdot \ldots \cdot \beta_N^{(r)} \right) \circ \boldsymbol{\beta}_{N+1}^{(r)} \tag{B.12}$$

$$= \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_{N+1}^{(r)} \cdot \tilde{\beta}_r = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)\prime} \boldsymbol{\beta}_{N+1}^{(r)} \cdot \tilde{\beta}_r \tag{B.13}$$

$$= \sum_{r=1}^{R} B^{(r)} \cdot \tilde{\beta}_r = \bar{B} \in \mathbb{R}^{m \times J} \tag{B.14}$$

where $\tilde{\beta}_r = \prod_{j=2}^{N} \beta_j^{(r)}$. The same result holds for the tensor $\mathcal{C}$, which is equal to $C \in \mathbb{R}^{m \times Q}$, with the last mode's length changed from J to Q. Finally, concerning $\mathcal{D}$:

$$\mathcal{D} = \sum_{r=1}^{R} \delta_1^{(r)} \circ \ldots \circ \delta_{n-1}^{(r)} \circ \boldsymbol{\delta}_n^{(r)} \circ \delta_{n+1}^{(r)} \circ \ldots \circ \delta_N^{(r)} = \sum_{r=1}^{R} \left(\delta_1^{(r)} \cdot \ldots \cdot \delta_{n-1}^{(r)} \right) \cdot \boldsymbol{\delta}_n^{(r)} \cdot \left(\delta_{n+1}^{(r)} \cdot \ldots \cdot \delta_N^{(r)} \right) \tag{B.15}$$

$$= \sum_{r=1}^{R} \boldsymbol{\delta}_n^{(r)} \cdot \left(\delta_1^{(r)} \cdot \ldots \cdot \delta_{n-1}^{(r)} \right) \cdot \left(\delta_{n+1}^{(r)} \cdot \ldots \cdot \delta_N^{(r)} \right) = \sum_{r=1}^{R} \boldsymbol{\delta}_n^{(r)} \cdot \tilde{\delta}_r = \mathbf{d} \in \mathbb{R}^m , \tag{B.16}$$

with $\tilde{\delta}_r = \prod_{j \neq n}^{N} \delta_j^{(r)}$. Notice that the resulting mode-n product reduces to an ordinary dot product between the matrix W and the vector $\bar{\mathbf{d}}$.

It remains to prove that the structure imposed by standard VARX and Panel VAR models holds also in the model of eq. (16). Notice that the latter does not impose any restriction on the coefficients, other than the PARAFAC(R) decomposition. It must be stressed that it is not possible to achieve the desired structure of the coefficients, in terms of the location of the zeros, by means of an accurate choice of the marginals. In fact, the decomposition we are assuming does not allow to create a particular structure on the resulting tensor.

Nonetheless, it is still possible to achieve the desired result by a slight modification of the model in eq. (16). For example, consider the coefficient tensor $\mathcal{B}$, then to create a tensor whose entries are non-zero only in some pre-specified (hence *a-priori* known) cells, it suffices to multiply $\mathcal{B}$ by a binary tensor (i.e. one where all entries are either 0 or 1) via the Hadamard product. In formulas, let $\mathcal{H} \in \{0,1\}^{I_1 \times \ldots \times I_N \times J}$, such that it has 0 only in those cells which are known to be null. Then:

$$\bar{\mathcal{B}} = \mathcal{H} \odot \mathcal{B}$$

will have the desired structure. The same way of reasoning holds for any coefficient tensor as well as for the covariance matrices.

To conclude, in Panel VAR models one generally has as regressors in each equation a function of the endogenous variables (for example their average). Since this does not affect the coefficients of the model, it is possible to re-create it in our framework by simply rearranging the regressors in eq. (16) accordingly. In terms of the model, none of the issues described invalidates the formulation of eq. (16), which is able to encompass all of them by suitable rearrangements of the covariates and/or the coefficients, which are consistent with the general model. □

Remark B.1 (follows from 2.6)**.** *From the VECM in eq.* (24) *and denoting* $\mathbf{y}_{t-1} = \mathrm{vec}\,(Y_{t-1})$ *we can obtain an explicit form for the long run equilibrium (or cointegrating) relations, as follows:*

$$\boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} = \left(\sum_{r=1}^{R} \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)} \circ \tilde{\boldsymbol{\beta}}_3^{(r)} \right) \times_3 \mathbf{y}_{t-1} \tag{B.17a}$$

$$= \sum_{r=1}^{R} \left(\tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)} \right) \cdot \langle \tilde{\boldsymbol{\beta}}_3^{(r)}, \mathbf{y}_{t-1} \rangle \tag{B.17b}$$

$$= \sum_{r=1}^{R} \tilde{B}_{12}^{(r)} \cdot \langle \tilde{\boldsymbol{\beta}}_3^{(r)}, \mathbf{y}_{t-1} \rangle \,, \tag{B.17c}$$

with $\tilde{B}_{12}^{(r)} = \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)}$ being a $K \times K$ matrix of loadings for each $r = 1, \dots, R$, while the inner product $\langle \tilde{\boldsymbol{\beta}}_3^{(r)}, \mathbf{y}_{t-1} \rangle$ defines the cointegrating relations. Notice that for a generic entry $y_{ij,t}$, the previous long run relation is defined in terms of all the entries of the lagged matrix Y_{t-1}, each one having a long run coefficient (in the r-th relation) $\tilde{\beta}_{3,k}^{(r)}$, where k can be obtained from (i,j) via a one-to-one mapping corresponding to the reshaping of the $K \times K$ matrix Y_{t-1} into the $K^2 \times 1$ vector $\mathbf{y}_{t-1}$.

Finally, as the cointegrating relations are not unique, that is $\boldsymbol{\beta}$ in eq. (24) *is not identified, the same is true for the tensor model, as noted in Section 2.*

C Initialisation details

It is well known that the Gibbs sampler algorithm is highly sensitive to the choice of the initial value. From this point of view, the most difficult parameters initialise in the proposed model are the margins of the tensor of coefficients, that is the set of vectors: $\{\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}\}_{r=1}^{R}$. Due to the high complexity of the parameter space, we have chosen to perform an initialisation scheme which is based on the Simulated Annealing (SA) algorithm (see Robert and Casella (2004) and Press et al. (2007) for a thorough discussion). This algorithm is similar to the Metropolis-Hastings one, and the idea behind it is to perform a stochastic optimisation by proposing random moves from the current state which are always accepted when improving the optimum and have positive probability of acceptance even when they are not improving. This is used in order to allow the algorithm to escape from local optima. Denoting the objective function to be minimised by $f(\boldsymbol{\theta})$, the Simulated Annealing method accepts a move from the current state $\boldsymbol{\theta}^(i)$ to the proposed one $\boldsymbol{\theta}^{new}$ with probability given by the Bolzmann-like distribution:

$$p(\Delta f, T) = \exp\left\{ -\frac{\Delta f}{T} \right\} . \tag{C.18}$$

Here $\Delta f = f(\boldsymbol{\theta}^{new}) - f(\boldsymbol{\theta}^{(i)})$ and T is a parameter called temperature. The key of the SA method is in the cooling scheme, which describes the deterministic, decreasing evolution of the temperature over the iterations of the algorithm: it has been proved that under sufficiently slow decreasing schemes, the SA yields a global optimum.

We propose to use the SA algorithm for minimising the objective function:

$$f(\{\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}\}_{r=1}^{R}) = \kappa_N \psi_N + \kappa_3 \psi_3 \,, \tag{C.19}$$

where κ_N is an overall penalty given by the Frobenius norm of the tensor constructed from simulated margins, while κ_3 is the penalty of the sum (over r) of the norms of the marginals $\boldsymbol{\beta}_3^{(r)}$. In formulas:

$$\psi_N = \left\| \mathcal{B}^{SA} \right\|_2 \qquad \psi_3 = \sum_{r=1}^{R} \left\| \boldsymbol{\beta}_3^{(r)} \right\|_2 . \tag{C.20}$$

The proposal distribution for each margin is a normal $\mathcal{N}_{I_j}(\mathbf{0}, \sigma I)$, independent from the current state of the algorithm. Finally, we have chosen a logarithmic cooling scheme which updates the temperature at each iteration of the SA:

$$T_i = \frac{k}{1 + \log(i)} \qquad i = 1, \dots, I \,, \tag{C.21}$$

where $k > 0$ is a tuning parameter, which can be interpreted as the initial value of the temperature. In order to perform the initialisation of the margins, we run the SA algorithm for $I = 1000$ iterations, then we took the vectors which gave the best fit in terms of minimum value of the objective function.

In the tensor case, the initialization of the PARAFAC marginals $\{\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}, \boldsymbol{\beta}_4^{(r)}\}_{r=1}^R$ follows the same line, with ψ_3 in eq. (C.20) replaced by:

$$\psi_4 = \sum_{r=1}^{R} \left\| \boldsymbol{\beta}_4^{(r)} \right\|_2 . \tag{C.22}$$

D Computational details - matrix case

In this section we will follow the convention of denoting the prior distributions with $\pi(\cdot)$. In addition, let $\mathbf{W} = \{W_{j,r}\}_{j,r}$ be the collection of all (local variance) matrices $W_{j,r}$, for $j = 1, 2, 3$ and $r = 1, \ldots, R$; $I_0 = \sum_{j=1}^{3} I_j$ the sum of the length of each mode of the tensor $\mathcal{B}$ and $\mathbf{Y} = \{Y_t, X_t\}_t$ the collection of observed variables.

D.1 Full conditional distribution of $\boldsymbol{\alpha}$

Define $D = [0, 1]^R \times [0, +\infty)$ and recall that $\boldsymbol{\alpha} = \alpha \boldsymbol{\iota}_{|A|}$, where the symbol $|A|$ denotes the cardinality of set A and with $\boldsymbol{\iota}_J$ is a vector of ones of length J. Samples from $p(\boldsymbol{\alpha}|\mathcal{B}, \mathbf{W}) = p(\alpha|\mathcal{B}, \mathbf{W})$, given a uniform discrete prior on the set $A = \{R^{-N}, \ldots, R^{-0.10}\}$, are obtained as follows:

$$p(\alpha|\mathcal{B}, \mathbf{W}) = \int_D p(\alpha, \boldsymbol{\phi}, \tau|\mathcal{B}, \mathbf{W}) \mathrm{d}\boldsymbol{\phi} \mathrm{d}\tau = \int_D \underbrace{p(\alpha|\boldsymbol{\phi}, \tau, \mathcal{B}, \mathbf{W})}_{(a)} \underbrace{p(\boldsymbol{\phi}, \tau|\mathcal{B}, \mathbf{W})}_{(b)} \mathrm{d}\boldsymbol{\phi} \mathrm{d}\tau. \tag{D.23}$$

Developing part (a):

$$\begin{aligned} p(\alpha|\boldsymbol{\phi}, \tau, \mathcal{B}, \mathbf{W}) &= \frac{p(\boldsymbol{\phi}, \tau|\alpha) p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau, \alpha) \pi(\alpha)}{\int_A p(\alpha, \boldsymbol{\phi}, \tau, \mathcal{B}, \mathbf{W}) \mathrm{d}\alpha} = \frac{p(\boldsymbol{\phi}|\alpha) p(\tau|\alpha) p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau) \pi(\alpha)}{p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau) \int_A p(\boldsymbol{\phi}|\alpha) p(\tau|\alpha) \pi(\alpha) \mathrm{d}\alpha} \\ &= \frac{p(\boldsymbol{\phi}|\alpha) p(\tau|\alpha) \pi(\alpha)}{\int_A p(\boldsymbol{\phi}|\alpha) p(\tau|\alpha) \pi(\alpha) \mathrm{d}\alpha} \propto p(\boldsymbol{\phi}|\alpha) p(\tau|\alpha) \pi(\alpha) \end{aligned} \tag{D.24}$$

where the last equality in the first row has been obtained by exploiting the independence relations given by the hierarchical structure of the prior (see Fig. (3)).

As part (b) is concerned:

$$\begin{aligned} p(\boldsymbol{\phi}, \tau|\mathcal{B}, \mathbf{W}) &= \int_A p(\boldsymbol{\phi}, \tau, \alpha|\mathcal{B}, \mathbf{W}) \mathrm{d}\alpha \propto \int_A p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau, \alpha) p(\boldsymbol{\phi}, \tau, \alpha) \mathrm{d}\alpha \\ &= \int_A p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau) p(\boldsymbol{\phi}, \tau|\alpha) \pi(\alpha) \mathrm{d}\alpha \propto \sum_{j=1}^{|A|} p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau) p(\boldsymbol{\phi}, \tau|\alpha_j) \pi(\alpha_j) \end{aligned} \tag{D.25}$$

where the first result in the last row is implied again by the structure of the prior, while the last one is due to the fact that α has discrete support. Since $\pi(\alpha) \sim \mathcal{U}(A)$ is a discrete uniform, it holds $\pi(\alpha_j) = 1/|A| \; \forall j$, where with an abuse of notation we define α_j to be the j-th element of the set A and not the j-th element of the vector $\boldsymbol{\alpha}$. Hence eq. (D.25) becomes:

$$p(\boldsymbol{\phi}, \tau|\mathcal{B}, \mathbf{W}) \propto \sum_{j=1}^{|A|} p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau) p(\boldsymbol{\phi}, \tau|\alpha_j) . \tag{D.26}$$

Substituting eq (D.24) and (D.26) into eq. (D.23) yields:

$$
\begin{aligned}
p(\alpha|\mathcal{B},\mathbf{W}) &= \int_D p(\alpha|\boldsymbol{\phi},\tau,\mathcal{B},\mathbf{W})p(\boldsymbol{\phi},\tau|\mathcal{B},\mathbf{W})\mathrm{d}\boldsymbol{\phi}\mathrm{d}\tau \\
&\propto \int_D \left[p(\boldsymbol{\phi}|\alpha)p(\tau|\alpha)\pi(\alpha)\right]\left[\sum_{j=1}^{|A|} p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi},\tau)p(\boldsymbol{\phi},\tau|\alpha_j)\right]\mathrm{d}\boldsymbol{\phi}\mathrm{d}\tau \\
&= \pi(\alpha)\int_D \left[\sum_{j=1}^{|A|} p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi},\tau)p(\boldsymbol{\phi},\tau|\alpha_j)\right]p(\boldsymbol{\phi},\tau|\alpha)\mathrm{d}\boldsymbol{\phi}\mathrm{d}\tau\,. \qquad \text{(D.27)}
\end{aligned}
$$

We approximate this integral using Monte Carlo approximation by drawing M samples from $p(\boldsymbol{\phi},\tau|\alpha)$, then we compute the posterior distribution of α following the Griddy Gibbs procedure proposed by Ritter and Tanner (1992), thus obtaining:

$$
\begin{aligned}
\tilde{p}(\alpha|\mathcal{B},\mathbf{W}) &\propto \pi(\alpha)\frac{1}{M}\sum_{l=1}^{M}\left[\sum_{j=1}^{|A|} p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi}_l,\tau_l)p(\boldsymbol{\phi}_l,\tau_l|\alpha_j)\right] \\
&= \pi(\alpha)\frac{1}{M}\sum_{l=1}^{M}\sum_{j=1}^{|A|}\omega_{l,j}(\alpha) = \pi(\alpha)\frac{1}{M}\sum_{i=1}^{M|A|}\omega_i(\alpha)\,. \qquad \text{(D.28)}
\end{aligned}
$$

The first equality comes from setting $\omega_{l,j}(\alpha) = p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi}_l,\tau_l)p(\boldsymbol{\phi}_l,\tau_l|\alpha_j)$ and the last by defining a unique index $i = (j-1)M + l$. As a consequence, in order to obtain the correct posterior distribution, it is necessary to normalize eq. (D.28) as follows:

$$
\mathbb{P}\left(\left\{\alpha=\alpha_j\right\}|\mathcal{B},\mathbf{W}\right) = \frac{\tilde{p}(\alpha_j|\mathcal{B},\mathbf{W})}{\sum_l \tilde{p}(\alpha_l|\mathcal{B},\mathbf{W})}\,. \qquad \text{(D.29)}
$$

D.2 Full conditional distribution of $\boldsymbol{\phi}$

In order to derive this posterior distribution, we make use of Lemma 7.9 in Guhaniyogi et al. (2017). Recall that: $a_\tau = \alpha R$, $b_\tau = \alpha(R)^{1/N}$ and $I_0 = \sum_{j=1}^N I_j$. The prior for $\boldsymbol{\phi}$ is $\pi(\boldsymbol{\phi}) \sim \mathcal{D}ir(\boldsymbol{\alpha})$.

$$
p(\boldsymbol{\phi}|\mathcal{B},\mathbf{W}) \propto \pi(\boldsymbol{\phi})p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi}) = \pi(\boldsymbol{\phi})\int_0^{+\infty} p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi},\tau)\pi(\tau)\mathrm{d}\tau\,. \qquad \text{(D.30)}
$$

By plugging in the prior distributions for τ, $\boldsymbol{\phi}$, $\boldsymbol{\beta}_j^{(r)}$ we obtain[8]:

$$
\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{B},\mathbf{W}) &\propto \prod_{r=1}^{R}\phi_r^{\alpha-1}\int_0^{+\infty}\left[\prod_{r=1}^{R}\prod_{j=1}^{N}(\tau\phi_r)^{-I_j/2}\left|W_{j,r}\right|^{-1/2}\exp\left\{-\frac{1}{2\tau\phi_r}\boldsymbol{\beta}_j^{(r)\prime}W_{j,r}^{-1}\boldsymbol{\beta}_j^{(r)}\right\}\right] \\
&\quad\cdot\tau^{a_\tau-1}\exp\left\{-b_\tau\tau\right\}\mathrm{d}\tau \\
&\propto \prod_{r=1}^{R}\phi_r^{\alpha-1}\int_0^{+\infty}\left[\prod_{r=1}^{R}(\tau\phi_r)^{-I_0/2}\exp\left\{-\frac{1}{2\tau\phi_r}\sum_{j=1}^{N}\boldsymbol{\beta}_j^{(r)\prime}W_{j,r}^{-1}\boldsymbol{\beta}_j^{(r)}\right\}\right] \\
&\quad\cdot\tau^{a_\tau-1}\exp\left\{-b_\tau\tau\right\}\mathrm{d}\tau\,. \qquad \text{(D.31)}
\end{aligned}
$$

[8]We have used the property of the determinant: $\det(kA) = k^n\det(A)$, for A square matrix of size n and k scalar.

Define $C_r = \frac{1}{2}\sum_{j=1}^N \boldsymbol{\beta}_j^{(r)'} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)}$, then group together the powers of τ and ϕ_r as follows:

$$\begin{aligned} p(\boldsymbol{\phi}|\mathcal{B},\mathbf{W}) &\propto \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau-1-\frac{RI_0}{2}} \exp\left\{-b_\tau\tau\right\} \left[\prod_{r=1}^R \exp\left\{-\frac{1}{2\tau\phi_r}C_r\right\}\right] \mathrm{d}\tau \\ &= \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau-1-\frac{RI_0}{2}} \exp\left\{-b_\tau\tau - \sum_{r=1}^R \frac{C_r}{2\tau\phi_r}\right\} \mathrm{d}\tau\,. \end{aligned} \tag{D.32}$$

Recall that the probability density function of a Generalized Inverse Gaussian in the parametrization with three parameters ($a>0$, $b>0$, $p\in\mathbb{R}$), with $x\in(0,+\infty)$, is given by:

$$x \sim GiG(a,b,p) \;\Rightarrow\; p(x|a,b,p) = \frac{\left(\frac{a}{b}\right)^{\frac{p}{2}}}{2K_p(\sqrt{ab})} x^{p-1} \exp\left\{-\frac{1}{2}\left(ax+\frac{b}{x}\right)\right\}, \tag{D.33}$$

with $K_p(\cdot)$ a modified Bessel function of the second type. Our goal is to reconcile eq. (E.80) to the kernel of this distribution. Since by definition $\sum_{r=1}^R \phi_r = 1$, it holds that $\sum_{r=1}^R (b_\tau\tau\phi_r) = (b_\tau\tau)\sum_{r=1}^R \phi_r = b_\tau\tau$. This allows to rewrite the exponential as:

$$\begin{aligned} p(\boldsymbol{\phi}|\mathcal{B},\mathbf{W}) &\propto \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{\left(a_\tau-\frac{RI_0}{2}\right)-1} \exp\left\{-\sum_{r=1}^R\left(\frac{C_r}{2\tau\phi_r}+b_\tau\tau\phi_r\right)\right\} \mathrm{d}\tau \\ &= \int_0^{+\infty} \left(\prod_{r=1}^R \phi_r^{\alpha-\frac{I_0}{2}-1}\right) \tau^{\left(\alpha R-\frac{RI_0}{2}\right)-1} \exp\left\{-\sum_{r=1}^R\left(\frac{C_r}{2\tau\phi_r}+b_\tau\tau\phi_r\right)\right\} \mathrm{d}\tau\,, \end{aligned} \tag{D.34}$$

where we expressed $a_\tau = \alpha R$. According to the results in Appendix A and Lemma 7.9 of Guhaniyogi et al. (2017), the function in the previous equation is the kernel of a generalized inverse Gaussian for $\psi_r = \tau\phi_r$, which yields the distribution of ϕ_r after normalization. Hence, for $r=1,\ldots,R$, we first sample :

$$p(\psi_r|\mathcal{B},\mathbf{W},\tau,\alpha) \sim GiG\left(\alpha-\frac{I_0}{2}, 2b_\tau, 2C_r\right) \tag{D.35}$$

then, renormalizing, we obtain:

$$\phi_r = \frac{\psi_r}{\sum_{l=1}^R \psi_l}\,. \tag{D.36}$$

D.3 Full conditional distribution of τ

The posterior distribution of the global variance parameter, τ, is derived by simple application of Bayes' Theorem:

$$\begin{aligned} p(\tau|\mathcal{B},\mathbf{W},\boldsymbol{\phi}) &\propto \pi(\tau)p(\mathcal{B}|\mathbf{W},\boldsymbol{\phi},\tau) \\ &\propto \tau^{a_\tau-1}\exp\left\{-b_\tau\tau\right\}\left[\prod_{r=1}^R (\tau\phi_r)^{-\frac{I_0}{2}} \exp\left\{-\frac{1}{2\tau\phi_r}\sum_{j=1}^N \boldsymbol{\beta}_j^{(r)'}(W_{j,r})^{-1}\boldsymbol{\beta}_j^{(r)}\right\}\right] \\ &\propto \tau^{a_\tau-\frac{RI_0}{2}-1}\exp\left\{-b_\tau\tau-\left(\sum_{r=1}^R \frac{C_r}{\phi_r}\frac{1}{\tau}\right)\right\}. \end{aligned} \tag{D.37}$$

This is the kernel of a generalized inverse Gaussian:

$$p(\tau|\mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) \sim GiG\left(a_\tau - \frac{RI_0}{2}, 2b_\tau, 2\sum_{r=1}^{R}\frac{C_r}{\phi_r}\right). \tag{D.38}$$

D.4 Full conditional distribution of $\lambda_{j,r}$

Start by observing that, for $j = 1, 2, 3$ and $r = 1, \ldots, R$, the prior distribution on the vector $\boldsymbol{\beta}_j^{(r)}$ defined in eq. (29f) implies that each component follows a double exponential distribution:

$$\beta_{j,p}^{(r)} \sim DE\left(0, \frac{\lambda_{j,r}}{\sqrt{\tau\phi_r}}\right) \tag{D.39}$$

with probability density function, for $j = 1, 2, 3$:

$$\pi(\beta_{j,p}^{(r)}|\lambda_{j,r}, \phi_r, \tau) = \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\left\{-\frac{\left|\beta_{j,p}^{(r)}\right|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}}\right\}. \tag{D.40}$$

Then, exploiting the prior distribution $\pi(\lambda_{j,r}) \sim \mathcal{G}a(a_\lambda, b_\lambda)$ and eq. (D.40):

$$\begin{aligned}
p\left(\lambda_{j,r}|\boldsymbol{\beta}_j^{(r)}, \phi_r, \tau\right) &\propto \pi(\lambda_{j,r}) p\left(\boldsymbol{\beta}_j^{(r)}|\lambda_{j,r}, \phi_r, \tau\right) \\
&\propto \lambda_{j,r}^{a_\lambda - 1} \exp\left\{-b_\lambda \lambda_{j,r}\right\} \prod_{p=1}^{I_j} \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\left\{-\frac{\left|\beta_{j,p}^{(r)}\right|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}}\right\} \\
&= \lambda_{j,r}^{a_\lambda - 1} \left(\frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}}\right)^{I_j} \exp\left\{-b_\lambda \lambda_{j,r}\right\} \exp\left\{-\frac{\sum_{p=1}^{I_j}\left|\beta_{j,p}^{(r)}\right|}{\sqrt{\tau\phi_r}/\lambda_{j,r}}\right\} \\
&\propto \lambda_{j,r}^{(a_\lambda + I_j) - 1} \exp\left\{-\left(b_\lambda + \frac{\left\|\boldsymbol{\beta}_j^{(r)}\right\|_1}{\sqrt{\tau\phi_r}}\right)\lambda_{j,r}\right\}.
\end{aligned} \tag{D.41}$$

This is the kernel of a gamma distribution, hence for $j = 1, 2, 3$, $r = 1, \ldots, R$:

$$p(\lambda_{j,r}|\mathcal{B}, \phi_r, \tau) \sim \mathcal{G}a\left(a_\lambda + I_j, b_\lambda + \frac{\left\|\boldsymbol{\beta}_j^{(r)}\right\|_1}{\sqrt{\tau\phi_r}}\right). \tag{D.42}$$

D.5 Full conditional distribution of $w_{j,r,p}$

We sample independently each component $w_{j,r,p}$ of the matrix $W_{j,r} = \mathrm{diag}(\mathbf{w}_{j,r})$, for $p = 1, \ldots, I_j$, $j = 1, 2, 3$ and $r = \ldots, R$, from the full conditional distribution:

$$\begin{aligned}
p\left(w_{j,r,p}|\boldsymbol{\beta}_j^{(r)}, \lambda_{j,r}, \phi_r, \tau\right) &\propto p\left(\beta_{j,p}^{(r)}|w_{j,r,p}, \phi_r, \tau\right) \pi(w_{j,r,p}|\lambda_{j,r}) \\
&= (\tau\phi_r)^{-\frac{1}{2}} w_{j,r,p}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\tau\phi_r}\beta_{j,p}^{(r)^2} w_{j,r,p}^{-1}\right\} \frac{\lambda_{j,r}^2}{2} \exp\left\{-\frac{\lambda_{j,r}^2}{2} w_{j,r,p}\right\}
\end{aligned}$$

$$\propto w_{j,r,p}^{-\frac{1}{2}} \exp\left\{ -\frac{\lambda_{j,r}^2}{2} w_{j,r,p} - \frac{\beta_{j,p}^{(r)^2}}{2\tau\phi_r} w_{j,r,p}^{-1} \right\}, \tag{D.43}$$

where the second row comes from the fact that $w_{j,r,p}$ influences only the p-th component of the vector $\boldsymbol{\beta}_j^{(r)}$. For $p = 1, \ldots, I_j$, $j = 1,2,3$ and $r = 1, \ldots, R$ we get:

$$p\left(w_{j,r,p} | \boldsymbol{\beta}_j^{(r)}, \lambda_{j,r}, \phi_r, \tau \right) \sim \left(\frac{1}{2}, \lambda_{j,r}^2, \frac{\beta_{j,p}^{(r)^2}}{\tau\phi_r} \right). \tag{D.44}$$

D.6 Full conditional distribution of the PARAFAC marginals $\boldsymbol{\beta}_j^{(r)}$, for $j = 1,2,3$

For $r = 1, \ldots, R$ we sample the PARAFAC marginals $(\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)})$ fro their full conditional distribution, since their joint distribution is not available in closed form. First, it is necessary to rewrite the likelihood function in a suitable way. To this aim, for $j = 1,2,3$ and $r = 1, \ldots, R$ define $\boldsymbol{\beta}_{-j}^{(r)} = \left\{ \boldsymbol{\beta}_i^{(r)} : i \neq j \right\}$, $\mathcal{B}_r = \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}$ and $\mathcal{B}_{-r} = \{ \mathcal{B}_i : i \neq r \}$. By properties of the mode-n product:

$$\mathcal{B} \times_3 \mathbf{x}_t = \left(\sum_{r=1}^R \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \times_3 \mathbf{x}_t = \left(\sum_{\substack{s=1 \\ s \neq r}}^R \boldsymbol{\beta}_1^{(s)} \circ \boldsymbol{\beta}_2^{(s)} \circ \boldsymbol{\beta}_3^{(s)} \right) \times_3 \mathbf{x}_t + \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \times_3 \mathbf{x}_t . \tag{D.45}$$

Since our interest is in $\boldsymbol{\beta}_j^{(r)}$ for $j = 1,2,3$, we focus on the second term of eq. (D.45):

$$\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \times_3 \mathbf{x}_t = \sum_{i_3=1}^{I_3} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \right) \cdot \beta_{3,i_3}^{(r)} x_{t,i_3} = \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \right) \cdot \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle . \tag{D.46}$$

The equality comes from the definition of mode-n product given in eq. (6). It holds:

$$\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \times_3 \mathbf{x}_t = \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \right) \cdot \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle = \boldsymbol{\beta}_1^{(r)} \circ \left(\boldsymbol{\beta}_2^{(r)} \cdot \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \right) \tag{D.47}$$

$$= \left(\boldsymbol{\beta}_1^{(r)} \cdot \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \right) \circ \boldsymbol{\beta}_2^{(r)}. \tag{D.48}$$

We exploited the fact that the outcome of the inner product is a scalar, then the result follows by linearity of the outer product.

Given a sample of length T and assuming that the distribution at time $t = 0$ is known (as standard practice in time series analysis), the likelihood function is given by:

$$L\left(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2 \right) = \prod_{t=1}^T \left[(2\pi)^{-\frac{k^2}{2}} |\Sigma_2|^{-\frac{k}{2}} |\Sigma_1|^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\Sigma_2^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right)' \Sigma_1^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right) \right) \right\} \right]$$

$$\propto \exp\left\{ -\frac{1}{2} \sum_{t=1}^T \operatorname{tr}\left(\Sigma_2^{-1} \tilde{E}_t' \Sigma_1^{-1} \tilde{E}_t \right) \right\}, \tag{D.49}$$

with:

$$\tilde{E}_t = \left(Y_t - \mathcal{B}_{-r} \times_3 \mathbf{x}_t - \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \right) \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \right). \tag{D.50}$$

Now, we can focus on a specific r and $j = 1, 2, 3$ and derive the full conditionals of each marginal vector of the tensor $\mathcal{B}$. To make computations clear:

$$L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2\right) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \operatorname{tr}\left(a_{1t} + a_{2t} + b_{1t} + b_{2t} + c_t\right)\right\}, \tag{D.51}$$

where:

$$a_{1t} = -\Sigma_2^{-1} Y_t' \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right) \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \tag{D.52a}$$

$$a_{2t} = -\Sigma_2^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \Sigma_1^{-1} Y_t \tag{D.52b}$$

$$b_{1t} = \Sigma_2^{-1} \left(\mathcal{B}_{-r} \times_3 \mathbf{x}_t\right)' \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right) \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \tag{D.52c}$$

$$b_{2t} = \Sigma_2^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \Sigma_1^{-1} \left(\mathcal{B}_{-r} \times_3 \mathbf{x}_t\right) \tag{D.52d}$$

$$c_t = \Sigma_2^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle . \tag{D.52e}$$

Exploiting linearity of the trace operator and the property $\operatorname{tr}\left(A'\right) = \operatorname{tr}\left(A\right)$, one gets:

$$\begin{aligned} p\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2\right) &\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \left(\operatorname{tr}\left(a_{1t}\right) + \operatorname{tr}\left(a_{2t}\right) + \operatorname{tr}\left(b_{1t}\right) + \operatorname{tr}\left(b_{2t}\right) + \operatorname{tr}\left(c_t\right)\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \left(2\operatorname{tr}\left(a_{1t}\right) + 2\operatorname{tr}\left(b_{1t}\right) + \operatorname{tr}\left(c_t\right)\right)\right\}. \end{aligned} \tag{D.53}$$

Consider now each term in the sum at the exponent, and exploit the property $\operatorname{tr}\left(ABC\right) = \operatorname{tr}\left(CAB\right) = \operatorname{tr}\left(BCA\right)$:

$$\begin{aligned} &\sum_{t=1}^{T} 2\operatorname{tr}\left(-\Sigma_2^{-1} Y_t' \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right) \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) + 2\operatorname{tr}\left(\Sigma_2^{-1} \left(\mathcal{B}_{-r} \times_3 \mathbf{x}_t\right)' \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right) \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \\ &+ \operatorname{tr}\left(\Sigma_2^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \\ &= 2\operatorname{tr}\left(-\Sigma_2^{-1} \left(\sum_{t=1}^{T} Y_t' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)\right) \\ &+ 2\operatorname{tr}\left(\Sigma_2^{-1} \left(\sum_{t=1}^{T} \left(\mathcal{B}_{-r} \times_3 \mathbf{x}_t\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)\right) \\ &+ \operatorname{tr}\left(\Sigma_2^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)' \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right) \left(\sum_{t=1}^{T} \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle^2\right)\right) \\ &= 2\operatorname{tr}\left(-\Sigma_2^{-1} \left(\sum_{t=1}^{T} Y_t' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)\right) \\ &+ 2\operatorname{tr}\left(\Sigma_2^{-1} \left(\sum_{t=1}^{T} \left(\mathcal{B}_{-r} \times_3 \mathbf{x}_t\right)' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle\right) \Sigma_1^{-1} \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right)\right) \end{aligned}$$

$$+\operatorname{tr}\left(\Sigma_2^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)'\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\boldsymbol{\beta}_3^{(r)'}VV'\boldsymbol{\beta}_3^{(r)}\right), \tag{D.54}$$

where $V=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_T]\in\mathbb{R}^{k^2\times T}$. Hence the likelihood function is proportional to:

$$\begin{aligned}
L\left(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2\right)\propto\exp\Bigg\{-\frac{1}{2}\Bigg[&2\operatorname{tr}\left(-\Sigma_2^{-1}\left(\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\right)\\
&+2\operatorname{tr}\left(\Sigma_2^{-1}\left(\sum_{t=1}^T(\mathcal{B}_{-r}\times_3\mathbf{x}_t)'\langle\boldsymbol{\beta}_3,^{(r)}\mathbf{x}_t\rangle\right)\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\right)\\
&+\operatorname{tr}\left(\Sigma_2^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)'\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\boldsymbol{\beta}_3^{(r)'}VV'\boldsymbol{\beta}_3^{(r)}\right)\Bigg]\Bigg\}.
\end{aligned} \tag{D.55}$$

It is now possible to proceed and derive the full conditional distributions of the PARAFAC marginals $\boldsymbol{\beta}_1^{(r)},\boldsymbol{\beta}_2^{(r)},\boldsymbol{\beta}_3^{(r)}$, for fixed r.

D.6.1 Full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$

From eq. (D.55):

$$\begin{aligned}
L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2)\propto\exp\Bigg\{-\frac{1}{2}\Bigg[&-2\operatorname{tr}\left(\Sigma_2^{-1}\left(\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\right)\\
&+2\operatorname{tr}\left(\Sigma_2^{-1}\left(\sum_{t=1}^T(\mathcal{B}_{-r}\times_3\mathbf{x}_t)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\right)\\
&+\operatorname{tr}\left(\Sigma_2^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)'\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\boldsymbol{\beta}_3^{(r)'}VV'\boldsymbol{\beta}_3^{(r)}\right)\Bigg]\Bigg\}\\
=\exp\Bigg\{-\frac{1}{2}\Bigg[&\boldsymbol{\beta}_3^{(r)'}VV'\boldsymbol{\beta}_3^{(r)}\operatorname{tr}\left(\Sigma_2^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)'\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\right)\\
&-2\operatorname{tr}\left(\Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\Sigma_2^{-1}\sum_{t=1}^T\left(Y_t'-(\mathcal{B}_{-r}\times_3\mathbf{x}_t)'\right)\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\Bigg]\Bigg\}.
\end{aligned} \tag{D.56}$$

For the posterior of $\boldsymbol{\beta}_1^{(r)}$ as well as for that of $\boldsymbol{\beta}_2^{(r)}$, define:

$$\begin{aligned}
\tilde{a}&=\boldsymbol{\beta}_3^{(r)'}VV'\boldsymbol{\beta}_3^{(r)}\\
\tilde{E}&=\sum_{t=1}^T\left(Y_t'-(\mathcal{B}_{-r}\times_3\mathbf{x}_t)'\right)\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle.
\end{aligned}$$

In addition, exploit the fact that $\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}=\boldsymbol{\beta}_1^{(r)}\boldsymbol{\beta}_2^{(r)'}$. As a result, eq. (D.56) becomes:

$$L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2)\propto\exp\left\{-\frac{1}{2}\left[\tilde{a}\operatorname{tr}\left(\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\boldsymbol{\beta}_2^{(r)'}\Sigma_2^{-1}\boldsymbol{\beta}_2^{(r)}\boldsymbol{\beta}_1^{(r)'}\right)-2\operatorname{tr}\left(\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\boldsymbol{\beta}_2^{(r)'}\Sigma_2^{-1}\tilde{E}\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\tilde{a}\left(\boldsymbol{\beta}_1^{(r)\prime}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right)\left(\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\boldsymbol{\beta}_2^{(r)}\right) - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}, \tag{D.57}$$

where the last equality comes from the use of the previously mentioned properties of the trace as well as by recognizing that the trace of a scalar is the scalar itself (all the therms in brackets in the last expression are scalars).

Equation (D.57) serves as a basis for the derivation of both the posterior of $\boldsymbol{\beta}_1^{(r)}$ and $\boldsymbol{\beta}_2^{(r)}$. With reference to the first one, the likelihood function in eq. (D.57) can be rearranged as to form the kernel of a Gaussian. For ease of notation define $\tilde{a}_1 = \tilde{a}\left(\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\boldsymbol{\beta}_2^{(r)}\right)$, then from eq. (D.57) it holds:

$$L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2) \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_1^{(r)\prime}\left(\frac{\Sigma_1}{\tilde{a}_1}\right)^{-1}\boldsymbol{\beta}_1^{(r)} - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}. \tag{D.58}$$

By Bayes' Theorem we obtain:

$$\begin{aligned}
&p\left(\boldsymbol{\beta}_1^{(r)}|\boldsymbol{\beta}_{-1}^{(r)},\mathcal{B}_{-r},W_{1,r},\phi_r,\tau,\mathbf{Y},\Sigma_1,\Sigma_2\right) \propto \pi\left(\boldsymbol{\beta}_1^{(r)}|W_{j,r},\phi_r,\tau\right)L\left(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2\right)\\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_1^{(r)\prime}\left(W_{1,r}\phi_r\tau\right)^{-1}\boldsymbol{\beta}_1^{(r)}\right\}\exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_1^{(r)\prime}\left(\frac{\Sigma_1}{\tilde{a}_1}\right)^{-1}\boldsymbol{\beta}_1^{(r)} - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_1^{(r)\prime}\left(\left(W_{1,r}\phi_r\tau\right)^{-1} + \left(\frac{\Sigma_1}{\tilde{a}_1}\right)^{-1}\right)\boldsymbol{\beta}_1^{(r)} - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}.
\end{aligned} \tag{D.59}$$

This is the kernel of a normal distribution, therefore for $r = 1,\dots,R$:

$$p\left(\boldsymbol{\beta}_1^{(r)}|\boldsymbol{\beta}_{-1}^{(r)},\mathcal{B}_{-r},W_{1,r},\phi_r,\tau,\Sigma_1,\Sigma_2,\mathbf{Y}\right) \sim \mathcal{N}_{I_1}\left(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1},\bar{\Sigma}_{\boldsymbol{\beta}_1}\right), \tag{D.60}$$

where:

$$\begin{aligned}
\bar{\Sigma}_{\boldsymbol{\beta}_1} &= \left[\left(W_{1,r}\phi_r\tau\right)^{-1} + \tilde{a}_1\Sigma_1^{-1}\right]^{-1}\\
\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1} &= \bar{\Sigma}_{\boldsymbol{\beta}_1}\Sigma_1^{-1}\tilde{E}'\Sigma_2^{-1}\boldsymbol{\beta}_2^{(r)}.
\end{aligned}$$

D.6.2 Full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$

Consider the likelihood function in eq. (D.57) and define $\tilde{a}_2 = \tilde{a}\left(\boldsymbol{\beta}_1^{(r)\prime}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right)$. By algebraic manipulation we obtain the proportionality relation:

$$L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2) \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_2^{(r)\prime}\left(\frac{\Sigma_2}{\tilde{a}_2}\right)^{-1}\boldsymbol{\beta}_2^{(r)} - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}. \tag{D.61}$$

Then, Bayes' theorem yields:

$$\begin{aligned}
&p\left(\boldsymbol{\beta}_2^{(r)}|\boldsymbol{\beta}_{-2}^{(r)},\mathcal{B}_{-r},W_{2,r},\phi_r,\tau,\mathbf{Y},\Sigma_1,\Sigma_2\right) \propto \pi\left(\boldsymbol{\beta}_1^{(r)}|W_{j,r},\phi_r,\tau\right)L\left(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2\right)\\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}\left(W_{2,r}\phi_r\tau\right)^{-1}\boldsymbol{\beta}_2^{(r)}\right\}\exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_2^{(r)\prime}\left(\frac{\Sigma_2}{\tilde{a}_2}\right)^{-1}\boldsymbol{\beta}_2^{(r)} - 2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}
\end{aligned}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_2^{(r)\prime}\left(\left(W_{2,r}\phi_r\tau\right)^{-1}+\left(\frac{\Sigma_2}{\tilde{a}_2}\right)^{-1}\right)\boldsymbol{\beta}_2^{(r)}-2\boldsymbol{\beta}_2^{(r)\prime}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}. \tag{D.62}$$

Which, for $r = 1, \ldots, R$, is the kernel of a normal distribution:

$$p\left(\boldsymbol{\beta}_2^{(r)}|\boldsymbol{\beta}_{-2}^{(r)}, \mathcal{B}_{-r}, W_{2,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \mathbf{Y}\right) \sim \mathcal{N}_{I_2}\left(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2}, \bar{\Sigma}_{\boldsymbol{\beta}_2}\right), \tag{D.63}$$

where:

$$\begin{aligned}
\bar{\Sigma}_{\boldsymbol{\beta}_2} &= \left[\left(W_{2,r}\phi_r\tau\right)^{-1}+\tilde{a}_2\Sigma_2^{-1}\right]^{-1} \\
\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2} &= \bar{\Sigma}_{\boldsymbol{\beta}_2}\Sigma_2^{-1}\tilde{E}\Sigma_1^{-1}\boldsymbol{\beta}_1^{(r)}.
\end{aligned}$$

D.6.3 Full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$

For ease of notation, define:

$$\begin{aligned}
A &= \Sigma_1^{-1}\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)\Sigma_2^{-1} \\
\tilde{A} &= A\left(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right)'.
\end{aligned}$$

Define $\tilde{V} = V \cdot (\mathrm{tr}(\tilde{A}))^{\frac{1}{2}}$, then eq. (D.55) becomes:

$$\begin{aligned}
&L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2) \propto \\
&\propto \exp\left\{-\frac{1}{2}\left[-2\,\mathrm{tr}\left(A\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)+2\,\mathrm{tr}\left(A\sum_{t=1}^T\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)+\boldsymbol{\beta}_3^{(r)\prime}VV'\boldsymbol{\beta}_3^{(r)}\,\mathrm{tr}\left(\tilde{A}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[-2\,\mathrm{tr}\left(A\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)+2\,\mathrm{tr}\left(A\sum_{t=1}^T\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)+\boldsymbol{\beta}_3^{(r)\prime}\tilde{V}\tilde{V}'\boldsymbol{\beta}_3^{(r)}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_3^{(r)\prime}\tilde{V}\tilde{V}'\boldsymbol{\beta}_3^{(r)}-2\,\mathrm{tr}\left(A\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle-A\sum_{t=1}^T\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\right]\right\}.
\end{aligned} \tag{D.64}$$

Then, focus on the second term in square brackets:

$$\begin{aligned}
&\mathrm{tr}\left(A\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle-A\sum_{t=1}^T\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right) \\
&=\mathrm{tr}\left(A\left(\sum_{t=1}^T Y_t'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle-\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right)\right)=\mathrm{tr}\left(A\sum_{t=1}^T\left(Y_t'-\left(\mathcal{B}_{-r}\times_3\mathbf{x}_t\right)'\right)\langle\boldsymbol{\beta}_3^{(r)},\mathbf{x}_t\rangle\right).
\end{aligned} \tag{D.65}$$

For ease of notation, define $\tilde{\tilde{Y}}_t = Y_t' - (\mathcal{B}_{-r} \times_3 \mathbf{x}_t)'$, then by linearity of the trace operator:

$$
\begin{aligned}
&= \operatorname{tr}\left(A \sum_{t=1}^{T} \tilde{\tilde{Y}}_t' \langle \boldsymbol{\beta}_3^{(r)}, \mathbf{x}_t \rangle \right) = \operatorname{tr}\left(\sum_{t=1}^{T} \left(A \tilde{\tilde{Y}}_t' \right) \left(\boldsymbol{\beta}_3^{(r)\prime} \mathbf{x}_t \right) \right) = \sum_{t=1}^{T} \operatorname{tr}\left(A \tilde{\tilde{Y}}_t \right) \left(\boldsymbol{\beta}_3^{(r)\prime} \mathbf{x}_t \right) \\
&= \sum_{t=1}^{T} \tilde{y}_t \left(\boldsymbol{\beta}_3^{(r)\prime} \mathbf{x}_t \right) = \tilde{\mathbf{y}}' V' \boldsymbol{\beta}_3^{(r)} \,,
\end{aligned} \tag{D.66}
$$

where we defined $\tilde{y}_t = \operatorname{tr}(A \tilde{\tilde{Y}}_t)$. As a consequence, rewrite eq. (D.64) as:

$$
L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2) \propto \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_3^{(r)\prime} (\tilde{V}\tilde{V}') \boldsymbol{\beta}_3^{(r)} - 2\tilde{\mathbf{y}}' V' \boldsymbol{\beta}_3^{(r)} \right] \right\} . \tag{D.67}
$$

We can now recover the full conditional posterior distribution of $\boldsymbol{\beta}_3^{(r)}$ by applying Bayes' Theorem:

$$
\begin{aligned}
& p\left(\boldsymbol{\beta}_3^{(r)} | \boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, W_{3,r}, \phi_r, \tau, \mathbf{Y}, \Sigma_1, \Sigma_2 \right) \propto \pi\left(\boldsymbol{\beta}_3^{(r)} | W_{3,r}, \phi_r, \tau \right) L\left(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2 \right) \\
& \propto \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}_3^{(r)\prime} \left(W_{3,r}\phi_r\tau \right)^{-1} \boldsymbol{\beta}_3^{(r)} \right\} \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_3^{(r)\prime} (\tilde{V}\tilde{V}') \boldsymbol{\beta}_3^{(r)} - 2\tilde{\mathbf{y}}' V' \boldsymbol{\beta}_3^{(r)} \right] \right\} \\
& \propto \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_3^{(r)\prime} \left(\left(W_{3,r}\phi_r\tau \right)^{-1} + \tilde{V}\tilde{V}' \right) \boldsymbol{\beta}_3^{(r)} - 2\tilde{\mathbf{y}}' V' \boldsymbol{\beta}_3^{(r)} \right] \right\} ,
\end{aligned} \tag{D.68}
$$

which is the kernel of a normal distribution. As a consequence, defining:

$$
\begin{aligned}
\bar{\Sigma}_{\boldsymbol{\beta}_3} &= \left[\left(W_{3,r}\phi_r\tau \right)^{-1} + \tilde{V}\tilde{V}' \right]^{-1} \\
\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3} &= \bar{\Sigma}_{\boldsymbol{\beta}_3} V \tilde{\mathbf{y}} \,,
\end{aligned}
$$

we get, for $r = 1, \dots, R$:

$$
p\left(\boldsymbol{\beta}_3^{(r)} | \boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, W_{j,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \mathbf{Y} \right) \sim \mathcal{N}_{I_1 \cdot I_2}\left(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3}, \bar{\Sigma}_{\boldsymbol{\beta}_3} \right) . \tag{D.69}
$$

D.7 Full conditional distribution of Σ_1

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_1 is conjugate:

$$
\begin{aligned}
& p(\Sigma_1|\mathcal{B}, \mathbf{Y}, \Sigma_2, \gamma) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_2, \Sigma_1) \pi(\Sigma_1) \\
& \propto |\Sigma_1|^{-\frac{TI_2}{2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}\left(\Sigma_2^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right)' \Sigma_1^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right) \right) \right\} |\Sigma_1|^{-\frac{\nu_1 + I_1 + 1}{2}} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + TI_2 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\operatorname{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \operatorname{tr}\left(\sum_{t=1}^{T} \left(\left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right) \Sigma_2^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right)' \right) \Sigma_1^{-1} \right) \right] \right\} .
\end{aligned} \tag{D.70}
$$

The last row comes from exploiting two ties the linearity of the trace operator. For ease of notation, define $S_1 = \sum_{t=1}^{T} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right) \Sigma_2^{-1} \left(Y_t - \mathcal{B} \times_3 \mathbf{x}_t \right)'$, obtaining:

$$
p(\Sigma_1|\mathcal{B}, \mathbf{Y}, \Sigma_2, \gamma) \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + TI_2 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\operatorname{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \operatorname{tr}\left(S_1 \Sigma_1^{-1} \right) \right] \right\}
$$

$$\propto |\Sigma_1|^{-\frac{(\nu_1+TI_2)+I_1+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\left(\gamma\Psi_1+S_1\right)\Sigma_1^{-1}\right)\right\}, \tag{D.71}$$

where we have used again the linearity of the trace operator. As a consequence:

$$p(\Sigma_1|\mathcal{B},\mathbf{Y},\Sigma_2,\gamma) \sim \mathcal{IW}_{I_1}\left(\nu_1+TI_2,\gamma\Psi_1+S_1\right). \tag{D.72}$$

D.8 Full conditional distribution of Σ_2

By the same reasoning of Σ_1, the posterior full conditional distribution of Σ_2 is conjugate and follows from:

$$\begin{aligned}
&p(\Sigma_2|\mathcal{B},\mathbf{Y},\Sigma_1,\gamma) \propto L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2)\pi(\Sigma_2|\gamma)\\
&\propto |\Sigma_2|^{-\frac{TI_1}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\operatorname{tr}\left(\Sigma_2^{-1}\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)'\Sigma_1^{-1}\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)\right)\right\}|\Sigma_2|^{-\frac{\nu_2+I_2+1}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1+1}{2}}\exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\operatorname{tr}\left(\sum_{t=1}^{T}\left(\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)'\Sigma_1^{-1}\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)\right)\Sigma_2^{-1}\right)\right]\right\}.
\end{aligned} \tag{D.73}$$

The last row comes from exploiting two ties the linearity of the trace operator. For ease of notation, define $S_2 = \sum_{t=1}^{T}\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)'\Sigma_1^{-1}\left(Y_t-\mathcal{B}\times_3\mathbf{x}_t\right)$, obtaining:

$$\begin{aligned}
p(\Sigma_2|\mathcal{B},\mathbf{Y},\Sigma_1,\gamma) &\propto |\Sigma_2|^{-\frac{\nu_2+I_2+I_1k+1}{2}}\exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\operatorname{tr}\left(S_2\Sigma_2^{-1}\right)\right]\right\}\\
&\propto |\Sigma_2|^{-\frac{(\nu_2+TI_1)+I_2+1}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\left(\gamma\Psi_2+S_2\right)\Sigma_2^{-1}\right)\right\},
\end{aligned} \tag{D.74}$$

where we have used again the linearity of the trace operator. As a consequence:

$$p(\Sigma_2|\mathcal{B},\mathbf{Y},\Sigma_1) \sim \mathcal{IW}_{I_2}\left(\nu_2+TI_1,\gamma\Psi_2+S_2\right). \tag{D.75}$$

D.9 Full conditional distribution of γ

Using a gamma prior distribution we have:

$$\begin{aligned}
p(\gamma|\Sigma_1,\Sigma_2) &\propto p(\Sigma_1,\Sigma_2|\gamma)\pi(\gamma)\\
&\propto |\gamma\Psi_1|^{-\frac{\nu_1}{2}}|\gamma\Psi_2|^{-\frac{\nu_2}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_1\Sigma_1^{-1}\right)\right\}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)\right\}\gamma^{a_\gamma-1}\exp\{-b_\gamma\gamma\}\\
&\propto \gamma^{-\frac{\nu_1I_1+\nu_2I_2}{2}}\exp\left\{-\frac{1}{2}\gamma\operatorname{tr}\left(\Psi_1\Sigma_1^{-1}+\Psi_2\Sigma_2^{-1}\right)-b_\gamma\gamma\right\}\gamma^{a_\gamma-1}\\
&\propto \gamma^{a_\gamma-\frac{\nu_1I_1+\nu_2I_2}{2}-1}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Psi_1\Sigma_1^{-1}+\Psi_2\Sigma_2^{-1}\right)-b_\gamma\gamma\right\},
\end{aligned} \tag{D.76}$$

thus:

$$p(\gamma|\Sigma_1,\Sigma_2) \sim \mathcal{G}a\left(a_\gamma+\frac{1}{2}(\nu_1I_1+\nu_2I_2), b_\gamma+\frac{1}{2}\operatorname{tr}\left(\Psi_1\Sigma_1^{-1}+\Psi_2\Sigma_2^{-1}\right)\right). \tag{D.77}$$

E Computational details - tensor case

In this section we will follow the convention of denoting the prior distributions with $\pi(\cdot)$. In addition, let $\mathbf{W} = \{W_{j,r}\}_{j,r}$ be the collection of all (local variance) matrices $W_{j,r}$, for $j = 1, 2, 3, 4$ and $r = 1, \dots, R$; $I_0 = \sum_{j=1}^{4} I_j$ the sum of the length of each mode of the tensor $\mathcal{B}$ and $\mathbf{Y} = \{\mathcal{Y}_t, \mathcal{X}_t\}_t$ the collection of observed variables.

E.1 Full conditional distribution of $\boldsymbol{\phi}$

In order to derive this posterior distribution, we make use of Lemma 7.9 in Guhaniyogi et al. (2017). Recall that: $a_\tau = \alpha R$, $b_\tau = \alpha(R)^{1/N}$ and $I_0 = \sum_{j=1}^{N} I_j$. The prior for $\boldsymbol{\phi}$ is $\pi(\boldsymbol{\phi}) \sim \mathcal{D}ir(\boldsymbol{\alpha})$.

$$p(\boldsymbol{\phi}|\mathcal{B}, \mathbf{W}) \propto \pi(\boldsymbol{\phi}) p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}) = \pi(\boldsymbol{\phi}) \int_0^{+\infty} p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau)\pi(\tau) \mathrm{d}\tau \,. \tag{E.78}$$

By plugging in the prior distributions for τ, $\boldsymbol{\phi}$, $\boldsymbol{\beta}_j^{(r)}$ we obtain[9]:

$$\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{B}, \mathbf{W}) &\propto \prod_{r=1}^{R} \phi_r^{\alpha-1} \int_0^{+\infty} \left[\prod_{r=1}^{R} \prod_{j=1}^{N} (\tau\phi_r)^{-I_j/2} \left| W_{j,r} \right|^{-1/2} \exp\left\{ -\frac{1}{2\tau\phi_r} \boldsymbol{\beta}_j^{(r)\prime} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)} \right\} \right] \\
&\quad \cdot \tau^{a_\tau - 1} \exp\left\{ -b_\tau \tau \right\} \mathrm{d}\tau \\
&\propto \prod_{r=1}^{R} \phi_r^{\alpha-1} \int_0^{+\infty} \left[\prod_{r=1}^{R} (\tau\phi_r)^{-I_0/2} \exp\left\{ -\frac{1}{2\tau\phi_r} \sum_{j=1}^{N} \boldsymbol{\beta}_j^{(r)\prime} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)} \right\} \right] \\
&\quad \cdot \tau^{a_\tau - 1} \exp\left\{ -b_\tau \tau \right\} \mathrm{d}\tau \,.
\end{aligned} \tag{E.79}$$

Define $C_r = \sum_{j=1}^{N} \boldsymbol{\beta}_j^{(r)\prime} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)}$, then group together the powers of τ and ϕ_r as follows:

$$\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{B}, \mathbf{W}) &\propto \prod_{r=1}^{R} \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau - 1 - \frac{RI_0}{2}} \exp\left\{ -b_\tau \tau \right\} \left[\prod_{r=1}^{R} \exp\left\{ -\frac{1}{2\tau\phi_r} C_r \right\} \right] \mathrm{d}\tau \\
&= \prod_{r=1}^{R} \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau - 1 - \frac{Rd_0}{2}} \exp\left\{ -b_\tau \tau - \sum_{r=1}^{R} \frac{C_r}{2\tau\phi_r} \right\} \mathrm{d}\tau \,.
\end{aligned} \tag{E.80}$$

Recall that the probability density function of a Generalized Inverse Gaussian in the parametrization with three parameters ($a > 0$, $b > 0$, $p \in \mathbb{R}$), with $x \in (0, +\infty)$, is given by:

$$x \sim GiG(a, b, p) \;\; \Rightarrow \;\; p(x|a, b, p) = \frac{\left(\frac{a}{b}\right)^{\frac{p}{2}}}{2K_p(\sqrt{ab})} x^{p-1} \exp\left\{ -\frac{1}{2} \left(ax + \frac{b}{x} \right) \right\} \,, \tag{E.81}$$

with $K_p(\cdot)$ a modified Bessel function of the second type. Our goal is to reconcile eq. (E.80) to the kernel of this distribution. Since by definition $\sum_{r=1}^{R} \phi_r = 1$, it holds that $\sum_{r=1}^{R} (b_\tau \tau \phi_r) = (b_\tau \tau) \sum_{r=1}^{R} \phi_r = b_\tau \tau$. This allows to rewrite the exponential as:

$$p(\boldsymbol{\phi}|\mathcal{B}, \mathbf{W}) \propto \prod_{r=1}^{R} \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{\left(a_\tau - \frac{RI_0}{2}\right)-1} \exp\left\{ -\sum_{r=1}^{R} \left(\frac{C_r}{2\tau\phi_r} + b_\tau \tau \phi_r \right) \right\} \mathrm{d}\tau$$

[9]We have used the property of the determinant: $\det(kA) = k^n \det(A)$, for A square matrix of size n and k scalar.

$$= \int_0^{+\infty} \left(\prod_{r=1}^{R} \phi_r^{\alpha - \frac{I_0}{2} - 1} \right) \tau^{\left(\alpha R - \frac{R I_0}{2}\right) - 1} \exp\left\{ - \sum_{r=1}^{R} \left(\frac{C_r}{2\tau\phi_r} + b_\tau \tau \phi_r \right) \right\} \mathrm{d}\tau \,, \quad \text{(E.82)}$$

where we expressed $a_\tau = \alpha R$. According to the results in Appendix A and Guhaniyogi et al. (2017), the function in the previous equation is the kernel of a generalized inverse Gaussian for $\psi_r = \tau\phi_r$, which yields the distribution of ϕ_r after normalization. Hence, for $r = 1, \ldots, R$, we first sample :

$$p(\psi_r | \mathcal{B}, \mathbf{W}, \tau, \alpha) \sim GiG \left(\alpha - \frac{I_0}{2}, 2b_\tau, 2C_r \right) \quad \text{(E.83)}$$

then, renormalizing, we obtain (see Kruijer et al. (2010)):

$$\phi_r = \frac{\psi_r}{\sum_{l=1}^{R} \psi_l} \,. \quad \text{(E.84)}$$

E.2 Full conditional distribution of τ

The posterior distribution of the global variance parameter, τ, is derived by simple application of Bayes' Theorem:

$$\begin{aligned} p(\tau | \mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) &\propto \pi(\tau) p(\mathcal{B} | \mathbf{W}, \boldsymbol{\phi}, \tau) \\ &\propto \tau^{a_\tau - 1} \exp\{-b_\tau \tau\} \left[\prod_{r=1}^{R} (\tau\phi_r)^{-\frac{I_0}{2}} \exp\left\{ -\frac{1}{2\tau\phi_r} \sum_{j=1}^{4} \boldsymbol{\beta}_j^{(r)\prime} (W_{j,r})^{-1} \boldsymbol{\beta}_j^{(r)} \right\} \right] \\ &\propto \tau^{a_\tau - \frac{R I_0}{2} - 1} \exp\left\{ -b_\tau \tau - \left(\sum_{r=1}^{R} \frac{C_r}{\phi_r} \frac{1}{\tau} \right) \right\} \,. \end{aligned} \quad \text{(E.85)}$$

This is the kernel of a generalized inverse Gaussian:

$$p(\tau | \mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) \sim GiG \left(a_\tau - \frac{R I_0}{2}, 2b_\tau, 2 \sum_{r=1}^{R} \frac{C_r}{\phi_r} \right) \,. \quad \text{(E.86)}$$

E.3 Full conditional distribution of $\lambda_{j,r}$

Start by observing that, for $j = 1, 2, 3, 4$ and $r = 1, \ldots, R$, the prior distribution on the vector $\boldsymbol{\beta}_j^{(r)}$ defined in eq. (29f) implies that each component follows a double exponential distribution:

$$\beta_{j,p}^{(r)} \sim DE \left(0, \frac{\lambda_{j,r}}{\sqrt{\tau\phi_r}} \right) \quad \text{(E.87)}$$

with probability density function, for $j = 1, 2, 3, 4$ and $r = 1, \ldots, R$, given by:

$$\pi(\beta_{j,p}^{(r)} | \lambda_{j,r}, \phi_r, \tau) = \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\left\{ -\frac{\left| \beta_{j,p}^{(r)} \right|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}} \right\} \,. \quad \text{(E.88)}$$

Then, exploiting the prior $\pi(\lambda_{j,r}) \sim \mathcal{G}a(a_\lambda, b_\lambda)$ and eq. (E.88):

$$p\left(\lambda_{j,r} | \boldsymbol{\beta}_j^{(r)}, \phi_r, \tau \right) \propto \pi(\lambda_{j,r}) p\left(\boldsymbol{\beta}_j^{(r)} | \lambda_{j,r}, \phi_r, \tau \right)$$

$$\propto \lambda_{j,r}^{a_\lambda - 1} \exp\left\{-b_\lambda \lambda_{j,r}\right\} \prod_{p=1}^{I_j} \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\left\{ -\frac{\left|\beta_{j,p}^{(r)}\right|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}} \right\}$$

$$= \lambda_{j,r}^{a_\lambda - 1} \left(\frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \right)^{I_j} \exp\left\{-b_\lambda \lambda_{j,r}\right\} \exp\left\{ -\frac{\sum_{p=1}^{I_j} \left|\beta_{j,p}^{(r)}\right|}{\sqrt{\tau\phi_r}/\lambda_{j,r}} \right\}$$

$$\propto \lambda_{j,r}^{(a_\lambda + I_j) - 1} \exp\left\{ -\left(b_\lambda + \frac{\left\|\boldsymbol{\beta}_j^{(r)}\right\|_1}{\sqrt{\tau\phi_r}} \right) \lambda_{j,r} \right\}. \tag{E.89}$$

Thus, for $j = 1, 2, 3, 4$, $r = 1, \dots, R$, the full conditional distribution of $\lambda_{j,r}$ is given by:

$$p(\lambda_{j,r}|\mathcal{B}, \phi_r, \tau) \sim \mathcal{G}a \left(a_\lambda + I_j, b_\lambda + \frac{\left\|\boldsymbol{\beta}_j^{(r)}\right\|_1}{\sqrt{\tau\phi_r}} \right). \tag{E.90}$$

E.4 Full conditional distribution of $w_{j,r,p}$

We sample independently each component $w_{j,r,p}$ of the matrix $W_{j,r} = \operatorname{diag}(\mathbf{w}_{j,r})$, for $p = 1, \dots, I_j$, $j = 1, 2, 3, 4$ and $r = \dots, R$, from the full conditional distribution:

$$\begin{aligned} p\left(w_{j,r,p}|\boldsymbol{\beta}_j^{(r)}, \lambda_{j,r}, \phi_r, \tau\right) &\propto p\left(\beta_{j,p}^{(r)}|w_{j,r,p}, \phi_r, \tau\right) \pi(w_{j,r,p}|\lambda_{j,r}) \\ &= (\tau\phi_r)^{-\frac{1}{2}} w_{j,r,p}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\tau\phi_r} \beta_{j,p}^{(r)^2} w_{j,r,p}^{-1} \right\} \frac{\lambda_{j,r}^2}{2} \exp\left\{ -\frac{\lambda_{j,r}^2}{2} w_{j,r,p} \right\} \\ &\propto w_{j,r,p}^{-\frac{1}{2}} \exp\left\{ -\frac{\lambda_{j,r}^2}{2} w_{j,r,p} - \frac{\beta_{j,p}^{(r)^2}}{2\tau\phi_r} w_{j,r,p}^{-1} \right\}, \end{aligned} \tag{E.91}$$

where the second row comes from the fact that $w_{j,r,p}$ influences only the p-th component of the vector $\boldsymbol{\beta}_j^{(r)}$. For $p = 1, \dots, I_j$, $j = 1, 2, 3, 4$ and $r = 1, \dots, R$ we get:

$$p\left(w_{j,r,p}|\boldsymbol{\beta}_j^{(r)}, \lambda_{j,r}, \phi_r, \tau\right) \sim \left(\frac{1}{2}, \lambda_{j,r}^2, \frac{\beta_{j,p}^{(r)^2}}{\tau\phi_r} \right). \tag{E.92}$$

E.5 Full conditional distributions of PARAFAC marginals $\boldsymbol{\beta}_j^{(r)}$, for $j = 1, 2, 3, 4$

Define $\boldsymbol{\alpha}_1 \in \mathbb{R}^I$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^J$ and $\boldsymbol{\alpha}_3 \in \mathbb{R}^K$ and let $\mathcal{A} = \operatorname{vec}(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3)$. Then it holds:

$$\begin{aligned} \operatorname{vec}(\mathcal{A}) &= \operatorname{vec}(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3) = \boldsymbol{\alpha}_3 \otimes \operatorname{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right) \\ &= \boldsymbol{\alpha}_3 \otimes (\boldsymbol{\alpha}_2 \otimes \mathbf{I}_I) \operatorname{vec}(\boldsymbol{\alpha}_1) = (\boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \mathbf{I}_I) \boldsymbol{\alpha}_1 & \text{(E.93)} \\ &= \boldsymbol{\alpha}_3 \otimes \left[(\mathbf{I}_J \otimes \boldsymbol{\alpha}_1) \operatorname{vec}\left(\boldsymbol{\alpha}_2'\right) \right] = (\boldsymbol{\alpha}_3 \otimes \mathbf{I}_J \otimes \boldsymbol{\alpha}_1) \boldsymbol{\alpha}_2 & \text{(E.94)} \\ &= \operatorname{vec}\left(\operatorname{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right) \boldsymbol{\alpha}_3' \right) = \left(\mathbf{I}_K \otimes \operatorname{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right) \right) \operatorname{vec}\left(\boldsymbol{\alpha}_3'\right) \\ &= \left(\mathbf{I}_K \otimes \operatorname{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right) \right) \boldsymbol{\alpha}_3 = (\mathbf{I}_K \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1) \boldsymbol{\alpha}_3 \,. & \text{(E.95)} \end{aligned}$$

Consider the model in eq. (26), it holds:

$$\begin{aligned}
\mathcal{Y}_t &= \mathcal{B} \times_4 \mathbf{x}_t + \mathcal{E}_t \\
\mathrm{vec}\left(\mathcal{Y}_t\right) &= \mathrm{vec}\left(\mathcal{B} \times_4 \mathbf{x}_t + \mathcal{E}_t\right) \\
&= \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right) + \mathrm{vec}\left(\mathcal{B}_r \times_4 \mathbf{x}_t\right) + \mathrm{vec}\left(\mathcal{E}_t\right) \\
&\propto \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} . \qquad \text{(E.96)}
\end{aligned}$$

It is then possible to make explicit the dependence on each PARAFAC marginal by exploiting the results in eq. (E.93)-(E.95), as follows:

$$\begin{aligned}
\mathrm{vec}\left(\mathcal{Y}_t\right) &\propto \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \mathbf{b}_4 \boldsymbol{\beta}_4^{(r)} & \text{(E.97)} \\
&\propto \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_I\right) \boldsymbol{\beta}_1^{(r)} = \mathbf{b}_1 \boldsymbol{\beta}_1^{(r)} & \text{(E.98)} \\
&\propto \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\beta}_1^{(r)}\right) \boldsymbol{\beta}_2^{(r)} = \mathbf{b}_2 \boldsymbol{\beta}_2^{(r)} & \text{(E.99)} \\
&\propto \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_K \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right) \boldsymbol{\beta}_3^{(r)} = \mathbf{b}_3 \boldsymbol{\beta}_3^{(r)} . & \text{(E.100)}
\end{aligned}$$

Given a sample of length T and assuming that the distribution at time $t = 0$ is known (as standard practice in time series analysis), the likelihood function is given by:

$$\begin{aligned}
L\left(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) &= \prod_{t=1}^{T} (2\pi)^{-\frac{k^2 q}{2}} \left|\Sigma_3\right|^{-\frac{k^2}{2}} \left|\Sigma_2\right|^{-\frac{kq}{2}} \left|\Sigma_1\right|^{-\frac{kq}{2}} \\
&\quad \cdot \exp\left\{-\frac{1}{2} (\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t) \times^{1\ldots3} \left(\circ_{j=1}^{3} \Sigma_j^{-1}\right) \times^{1\ldots3} (\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t)\right\} & \text{(E.101)} \\
&\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} \tilde{E}_t \times^{1\ldots3} (\Sigma_1^{-1} \circ \Sigma_2^{-1} \circ \Sigma_3^{-1}) \times^{1\ldots3} \tilde{E}_t\right\} , & \text{(E.102)}
\end{aligned}$$

with:

$$\tilde{E}_t = \left(\mathcal{Y}_t - \mathcal{B}_{-r} \times_4 \mathbf{x}_t - \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle\right) . \qquad \text{(E.103)}$$

Alternatively, by exploiting the relation between the tensor normal distribution and the multivariate normal distribution, we have:

$$\begin{aligned}
L\left(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) &= \prod_{t=1}^{T} (2\pi)^{-\frac{k^2 q}{2}} \left|\Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1\right|^{-\frac{1}{2}} \\
&\quad \cdot \exp\left\{-\frac{1}{2} \mathrm{vec}\left(\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t\right)' \left(\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}\right) \mathrm{vec}\left(\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^{T} \mathrm{vec}\left(\tilde{E}_t\right)' \left(\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}\right) \mathrm{vec}\left(\tilde{E}_t\right)\right\} , & \text{(E.104)}
\end{aligned}$$

where: with:

$$\begin{aligned}
\mathrm{vec}\left(\tilde{E}_t\right) &= \mathrm{vec}\left(\mathcal{Y}_t - \mathcal{B}_{-r} \times_4 \mathbf{x}_t - \left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle\right) \\
&= \mathrm{vec}\left(\mathcal{Y}_t\right) - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right) - \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle
\end{aligned}$$

$$\propto \mathrm{vec}\left(\mathcal{Y}_t\right) - \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle . \tag{E.105}$$

Thus, defining $\mathbf{y}_t = \mathrm{vec}\left(\mathcal{Y}_t\right)$ and $\boldsymbol{\Sigma}^{-1} = \Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}$, one gets:

$$\begin{aligned}
L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) &\propto \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \mathrm{vec}\left(\tilde{E}_t\right)' \left(\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}\right) \mathrm{vec}\left(\tilde{E}_t\right) \right\} \\
&\propto \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} \left[\mathrm{vec}\left(\mathcal{Y}_t\right) - \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right]' \boldsymbol{\Sigma}^{-1} \\
&\cdot \left[\mathrm{vec}\left(\mathcal{Y}_t\right) - \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right] \Bigg\} \\
&= \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} \mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t - \mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \\
&- \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \mathbf{y}_t \\
&+ \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \Bigg\} \\
&\propto \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \\
&+ \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \Bigg\} .
\end{aligned} \tag{E.106}$$

Now, we focus on a specific $j = 1, 2, 3, 4$ and derive proportionality results whihc will be necessary to obtain the posterior full conditional distributions of the PARAFAC marginals of the tensor $\mathcal{B}$. Consider the case $j = 1$. By exploiting eq. (E.98) we get:

$$\begin{aligned}
L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) &\propto \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} \\
&+ \left(\mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right)' \boldsymbol{\Sigma}^{-1} \left(\mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right) \Bigg\} \\
&= \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \\
&+ \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \right]' \boldsymbol{\Sigma}^{-1} \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \right] \Bigg\} \\
&= \exp\Bigg\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)\prime} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \\
&- 2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \Bigg\} \boldsymbol{\beta}_1^{(r)} \\
&= \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)\prime} \mathbf{S}_1^L(t) \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L(t) \boldsymbol{\beta}_1^{(r)} \right\} ,
\end{aligned} \tag{E.107}$$

with:

$$\mathbf{S}_1^L(t) = \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\boldsymbol{\beta}_3^{(r)\prime} \otimes \boldsymbol{\beta}_2^{(r)\prime} \otimes \mathbf{I}_{I_1} \right) \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1} \right) \tag{E.108}$$

$$\mathbf{m}_1^L(t) = \mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1} \right) . \tag{E.109}$$

Consider the case $j = 2$. From eq. (E.99) we get:

$$\begin{aligned}
L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) \propto \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} \\
&+ \left(\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right)' \boldsymbol{\Sigma}^{-1} \left(\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right) \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \circ \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_2^{(r)} \\
&+ \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_2^{(r)} \right]' \boldsymbol{\Sigma}^{-1} \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_2^{(r)} \right] \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_2^{(r)\prime} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_2^{(r)} \\
&- 2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \Bigg\} \boldsymbol{\beta}_2^{(r)} \\
= \exp &\left\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_2^{(r)\prime} \mathbf{S}_2^L(t) \boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L(t) \boldsymbol{\beta}_2^{(r)} \right\} ,
\end{aligned} \tag{E.110}$$

with:

$$\mathbf{S}_2^L(t) = \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\boldsymbol{\beta}_3^{(r)\prime} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)\prime} \right) \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) \tag{E.111}$$

$$\mathbf{m}_2^L(t) = \mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)} \right) . \tag{E.112}$$

Consider the case $j = 3$, by exploiting eq. (E.100) we get:

$$\begin{aligned}
L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) \propto \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} \\
&+ \left(\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right)' \boldsymbol{\Sigma}^{-1} \left(\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} \right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right) \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} -2\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_3^{(r)} \\
&+ \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_3^{(r)} \right]' \boldsymbol{\Sigma}^{-1} \left[\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_3^{(r)} \right] \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)\prime} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\Sigma}^{-1} \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)} \right) \boldsymbol{\beta}_3^{(r)}
\end{aligned}$$

$$- 2\mathbf{y}_t'\mathbf{\Sigma}^{-1}\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right)\Bigg\}\boldsymbol{\beta}_3^{(r)}$$

$$= \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)'}\mathbf{S}_3^L(t)\boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L(t)\boldsymbol{\beta}_3^{(r)}\right\}, \tag{E.113}$$

with:

$$\mathbf{S}_3^L(t) = \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle^2 \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)'} \otimes \boldsymbol{\beta}_1^{(r)'}\right) \mathbf{\Sigma}^{-1} \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right) \tag{E.114}$$

$$\mathbf{m}_3^L(t) = \mathbf{y}_t'\mathbf{\Sigma}^{-1}\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle \left(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right). \tag{E.115}$$

Finally, in the case $j = 4$. From eq. (E.106) we get:

$$L\left(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3\right) \propto \exp\Bigg\{-\frac{1}{2}\sum_{t=1}^{T} -2\mathbf{y}_t'\mathbf{\Sigma}^{-1}\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)}$$
$$+ \boldsymbol{\beta}_4^{(r)'}\mathbf{x}_t \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \mathbf{\Sigma}^{-1}\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)}\Bigg\} \tag{E.116}$$

$$= \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_4^{(r)'}\mathbf{S}_4^L(t)\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L(t)\boldsymbol{\beta}_4^{(r)}\right\}, \tag{E.117}$$

with:

$$\mathbf{S}_4^L(t) = \mathbf{x}_t \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \mathbf{\Sigma}^{-1}\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\mathbf{x}_t' \tag{E.118}$$

$$\mathbf{m}_4^L(t) = \mathbf{y}_t'\mathbf{\Sigma}^{-1}\operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\mathbf{x}_t'. \tag{E.119}$$

It is now possible to derive the full conditional distributions for the PARAFAC marginals $\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}, \boldsymbol{\beta}_4^{(r)}$, for $r = 1, \ldots, R$, as shown in the following.

E.5.1 Full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$ is obtained by combining the prior distribution in eq. (29f) and the likelihood in eq. (E.107) as follows:

$$p(\boldsymbol{\beta}_1^{(r)}|\boldsymbol{\beta}_{-1}^{(r)}, \mathcal{B}_{-r}, W_{1,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\boldsymbol{\beta}_1^{(r)}|W_{1,r}, \phi_r, \tau)$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)'}\mathbf{S}_1^L(t)\boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L(t)\boldsymbol{\beta}_1^{(r)}\right\} \cdot \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_1^{(r)'}(W_{1,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_1^{(r)}\right\}$$
$$= \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)'}\mathbf{S}_1^L(t)\boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L(t)\boldsymbol{\beta}_1^{(r)} + \boldsymbol{\beta}_1^{(r)'}(W_{1,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_1^{(r)}\right]\right\}$$
$$= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_1^{(r)'}\left(\sum_{t=1}^{T}\mathbf{S}_1^L(t) + (W_{1,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_1^{(r)} - 2\left(\sum_{t=1}^{T}\mathbf{m}_1^L(t)\right)\boldsymbol{\beta}_1^{(r)}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_1^{(r)\prime}\bar{\Sigma}_{\boldsymbol{\beta}_1^r}^{-1}\boldsymbol{\beta}_1^{(r)} - 2\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1^r}\boldsymbol{\beta}_1^{(r)}\right]\right\},$$

where:

$$\bar{\Sigma}_{\boldsymbol{\beta}_1^r} = \left[(W_{1,r}\phi_r\tau)^{-1} + \sum_{t=1}^{T}\mathbf{S}_1^L(t)\right]^{-1}$$

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1^r} = \bar{\Sigma}_{\boldsymbol{\beta}_1^r}\left[\sum_{t=1}^{T}\mathbf{m}_1^L(t)\right]'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$, for $r = 1, \dots, R$, is given by:

$$p(\boldsymbol{\beta}_1^{(r)}|\boldsymbol{\beta}_{-1}^{(r)}, \mathcal{B}_{-r}, W_{1,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \sim \mathcal{N}_{I_1}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_1^r}, \bar{\Sigma}_{\boldsymbol{\beta}_1^r}). \tag{E.120}$$

E.5.2 Full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$ is obtained by combining the prior distribution in eq. (29f) and the likelihood in eq. (E.110) as follows:

$$\begin{aligned}
&p(\boldsymbol{\beta}_2^{(r)}|\boldsymbol{\beta}_{-2}^{(r)}, \mathcal{B}_{-r}, W_{2,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\boldsymbol{\beta}_2^{(r)}|W_{2,r}, \phi_r, \tau) \\
&\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\beta}_2^{(r)\prime}\mathbf{S}_2^L(t)\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L(t)\boldsymbol{\beta}_2^{(r)}\right\}\cdot\exp\left\{-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}(W_{2,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_2^{(r)}\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}\boldsymbol{\beta}_2^{(r)\prime}\mathbf{S}_2^L(t)\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L(t)\boldsymbol{\beta}_2^{(r)} + \boldsymbol{\beta}_2^{(r)\prime}(W_{2,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_2^{(r)}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_2^{(r)\prime}\left(\sum_{t=1}^{T}\mathbf{S}_2^L(t) + (W_{2,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_2^{(r)} - 2\left(\sum_{t=1}^{T}\mathbf{m}_2^L(t)\right)\boldsymbol{\beta}_2^{(r)}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_2^{(r)\prime}\bar{\Sigma}_{\boldsymbol{\beta}_2^r}^{-1}\boldsymbol{\beta}_2^{(r)} - 2\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2^r}\boldsymbol{\beta}_2^{(r)}\right]\right\},
\end{aligned}$$

where:

$$\bar{\Sigma}_{\boldsymbol{\beta}_2^r} = \left[(W_{2,r}\phi_r\tau)^{-1} + \sum_{t=1}^{T}\mathbf{S}_2^L(t)\right]^{-1}$$

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2^r} = \bar{\Sigma}_{\boldsymbol{\beta}_2^r}\left[\sum_{t=1}^{T}\mathbf{m}_2^L(t)\right]'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$, for $r = 1, \dots, R$, is given by:

$$p(\boldsymbol{\beta}_2^{(r)}|\boldsymbol{\beta}_{-2}^{(r)}, \mathcal{B}_{-r}, W_{2,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \sim \mathcal{N}_{I_2}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_2^r}, \bar{\Sigma}_{\boldsymbol{\beta}_2^r}). \tag{E.121}$$

E.5.3 Full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$ is obtained by combining the prior distribution in eq. (29f) and the likelihood in eq. (E.113) as follows:

$$
\begin{aligned}
&p(\boldsymbol{\beta}_3^{(r)}|\boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, W_{3,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\boldsymbol{\beta}_3^{(r)}|W_{3,r}, \phi_r, \tau) \\
&\propto \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)\prime} \mathbf{S}_3^L(t) \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L(t)\boldsymbol{\beta}_3^{(r)} \right\} \cdot \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}_3^{(r)\prime} (W_{3,r}\phi_r\tau)^{-1} \boldsymbol{\beta}_3^{(r)} \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[\sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)\prime} \mathbf{S}_3^L(t) \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L(t)\boldsymbol{\beta}_3^{(r)} + \boldsymbol{\beta}_3^{(r)\prime} (W_{3,r}\phi_r\tau)^{-1} \boldsymbol{\beta}_3^{(r)} \right] \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_3^{(r)\prime} \left(\sum_{t=1}^{T} \mathbf{S}_3^L(t) + (W_{3,r}\phi_r\tau)^{-1} \right) \boldsymbol{\beta}_3^{(r)} - 2 \left(\sum_{t=1}^{T} \mathbf{m}_3^L(t) \right) \boldsymbol{\beta}_3^{(r)} \right] \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_3^{(r)\prime} \bar{\Sigma}_{\boldsymbol{\beta}_3^r}^{-1} \boldsymbol{\beta}_3^{(r)} - 2\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3^r} \boldsymbol{\beta}_3^{(r)} \right] \right\} ,
\end{aligned}
$$

where:

$$
\begin{aligned}
\bar{\Sigma}_{\boldsymbol{\beta}_3^r} &= \left[(W_{3,r}\phi_r\tau)^{-1} + \sum_{t=1}^{T} \mathbf{S}_3^L(t) \right]^{-1} \\
\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3^r} &= \bar{\Sigma}_{\boldsymbol{\beta}_3^r} \left[\sum_{t=1}^{T} \mathbf{m}_3^L(t) \right]' .
\end{aligned}
$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$, for $r = 1, \ldots, R$, is given by:

$$
p(\boldsymbol{\beta}_3^{(r)}|\boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, W_{3,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \sim \mathcal{N}_{I_3}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_3^r}, \bar{\Sigma}_{\boldsymbol{\beta}_3^r}) . \tag{E.122}
$$

E.5.4 Full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$ is obtained by combining the prior distribution in eq. (29f) and the likelihood in eq. (E.117) as follows:

$$
\begin{aligned}
&p(\boldsymbol{\beta}_4^{(r)}|\boldsymbol{\beta}_{-4}^{(r)}, \mathcal{B}_{-r}, W_{4,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\boldsymbol{\beta}_4^{(r)}|W_{4,r}, \phi_r, \tau) \\
&\propto \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_4^{(r)\prime} \mathbf{S}_4^L(t) \boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L(t)\boldsymbol{\beta}_4^{(r)} \right\} \cdot \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}_4^{(r)\prime} (W_{4,r}\phi_r\tau)^{-1} \boldsymbol{\beta}_4^{(r)} \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[\sum_{t=1}^{T} \boldsymbol{\beta}_4^{(r)\prime} \mathbf{S}_4^L(t) \boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L(t)\boldsymbol{\beta}_4^{(r)} + \boldsymbol{\beta}_4^{(r)\prime} (W_{4,r}\phi_r\tau)^{-1} \boldsymbol{\beta}_4^{(r)} \right] \right\} \\
&= \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_4^{(r)\prime} \left(\sum_{t=1}^{T} \mathbf{S}_4^L(t) + (W_{4,r}\phi_r\tau)^{-1} \right) \boldsymbol{\beta}_4^{(r)} - 2 \left(\sum_{t=1}^{T} \mathbf{m}_4^L(t) \right) \boldsymbol{\beta}_4^{(r)} \right] \right\}
\end{aligned}
$$

$$= \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_4^{(r)\prime} \bar{\Sigma}_{\boldsymbol{\beta}_4^r}^{-1} \boldsymbol{\beta}_4^{(r)} - 2\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_4^r} \boldsymbol{\beta}_4^{(r)} \right] \right\} ,$$

where:

$$\bar{\Sigma}_{\boldsymbol{\beta}_4^r} = \left[(W_{4,r}\phi_r\tau)^{-1} + \sum_{t=1}^{T} \mathbf{S}_4^L(t) \right]^{-1}$$

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_4^r} = \bar{\Sigma}_{\boldsymbol{\beta}_4^r} \left[\sum_{t=1}^{T} \mathbf{m}_4^L(t) \right]' .$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$, for $r = 1, \dots, R$, is given by:

$$p(\boldsymbol{\beta}_4^{(r)} | \boldsymbol{\beta}_{-4}^{(r)}, \mathcal{B}_{-r}, W_{4,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \sim \mathcal{N}_{I_1 I_2 I_3}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_4^r}, \bar{\Sigma}_{\boldsymbol{\beta}_4^r}) . \tag{E.123}$$

E.6 Full conditional distribution of Σ_1

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_1 is conjugate. For ease of notation, define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t}$ the mode-1 matricization of $\tilde{\mathcal{E}}_t$ and $\mathbf{Z}_1 = \Sigma_3^{-1} \otimes \Sigma_2^{-1}$. By exploiting the relation between the tensor normal distribution and the multivariate normal distribution and the properties of the vectorization and trace operators, we obtain:

$$\begin{aligned}
& p(\Sigma_1 | \mathcal{B}, \mathbf{Y}, \Sigma_2, \Sigma_3, \gamma) \propto L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) \pi(\Sigma_1 | \gamma) \\
& \propto |\Sigma_1|^{-\frac{T I_2 I_3}{2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \mathrm{vec}\,(\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t)' \, (\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \, \mathrm{vec}\,(\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t) \right\} \\
& \cdot |\Sigma_1|^{-\frac{\nu_1 + I_1 + 1}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \sum_{t=1}^{T} \mathrm{vec}\left(\tilde{\mathcal{E}}_t \right)' (\mathbf{Z}_1 \otimes \Sigma_1^{-1}) \, \mathrm{vec}\left(\tilde{\mathcal{E}}_t \right) \right] \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \sum_{t=1}^{T} \mathrm{vec}\left(\tilde{\mathbf{E}}_{(1),t} \right)' (\mathbf{Z}_1 \otimes \Sigma_1^{-1}) \, \mathrm{vec}\left(\tilde{\mathbf{E}}_{(1),t} \right) \right] \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \sum_{t=1}^{T} \mathrm{tr}\left(\mathrm{vec}\left(\tilde{\mathbf{E}}_{(1),t} \right)' \mathrm{vec}\left(\Sigma_1^{-1} \tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_1 \right) \right) \right] \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \sum_{t=1}^{T} \mathrm{tr}\left(\tilde{\mathbf{E}}_{(1),t}' \Sigma_1^{-1} \tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_1 \right) \right] \right\} \\
& \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \sum_{t=1}^{T} \mathrm{tr}\left(\tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_1 \tilde{\mathbf{E}}_{(1),t}' \Sigma_1^{-1} \right) \right] \right\} .
\end{aligned} \tag{E.124}$$

For ease of notation, define $S_1 = \sum_{t=1}^{T} \tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_1 \tilde{\mathbf{E}}_{(1),t}'$. Then:

$$p(\Sigma_1 | \mathcal{B}, \mathbf{Y}, \Sigma_2, \Sigma_3) \propto |\Sigma_1|^{-\frac{\nu_1 + I_1 + T I_2 I_3 + 1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}\left(\gamma \Psi_1 \Sigma_1^{-1} \right) + \mathrm{tr}\left(S_1 \Sigma_1^{-1} \right) \right] \right\}$$

$$\propto |\Sigma_1|^{-\frac{(\nu_1+TI_2I_3)+I_1+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\left(\gamma\Psi_1+S_1\right)\Sigma_1^{-1}\right)\right\}, \tag{E.125}$$

Therefore, the posterior full conditional distribution of Σ_1 is given by:

$$p(\Sigma_1|\mathcal{B},\mathbf{Y},\Sigma_2,\Sigma_3,\gamma) \sim \mathcal{IW}_{I_1}\left(\nu_1+TI_2I_3,\gamma\Psi_1+S_1\right). \tag{E.126}$$

E.7 Full conditional distribution of Σ_2

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_2 is conjugate. For ease of notation, define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B}\times_4\mathbf{x}_t$ and $\tilde{\mathbf{E}}_{(2),t}$ the mode-2 matricization of $\tilde{\mathcal{E}}_t$. By exploiting the relation between the tensor normal distribution and the matrix normal distribution and the properties of the Kronecker product and of the vectorization and trace operators we obtain:

$$\begin{aligned}
&p(\Sigma_2|\mathcal{B},\mathbf{Y},\Sigma_1,\Sigma_3,\gamma) \propto L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2,\Sigma_3)\pi(\Sigma_2|\gamma)\\
&\propto |\Sigma_2|^{-\frac{TI_1I_3}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\mathcal{Y}_t-\mathcal{B}\times_4\mathbf{x}_t)\times^{1\ldots3}(\Sigma_1^{-1}\circ\Sigma_2^{-1}\circ\Sigma_3^{-1})\times^{1\ldots3}(\mathcal{Y}_t-\mathcal{B}\times_4\mathbf{x}_t)\right\}\\
&\cdot|\Sigma_2|^{-\frac{\nu_2+I_2+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Psi_2\Sigma_2^{-1}\right)\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\sum_{t=1}^{T}\tilde{\mathcal{E}}_t\times^{1\ldots3}(\Sigma_1^{-1}\circ\Sigma_2^{-1}\circ\Sigma_3^{-1})\times^{1\ldots3}\tilde{\mathcal{E}}_t\right]\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\sum_{t=1}^{T}\operatorname{tr}\left(\tilde{\mathbf{E}}'_{(2),t}(\Sigma_3^{-1}\otimes\Sigma_1^{-1}\otimes\Sigma_2^{-1})\tilde{\mathbf{E}}_{(2),t}\right)\right]\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\sum_{t=1}^{T}\operatorname{tr}\left((\Sigma_3^{-1}\otimes\Sigma_1^{-1})\tilde{\mathbf{E}}'_{(2),t}\Sigma_2^{-1}\tilde{\mathbf{E}}_{(2),t}\right)\right]\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right)+\operatorname{tr}\left(\sum_{t=1}^{T}\tilde{\mathbf{E}}_{(2),t}(\Sigma_3^{-1}\otimes\Sigma_1^{-1})\tilde{\mathbf{E}}'_{(2),t}\Sigma_2^{-1}\right)\right]\right\}\\
&\propto |\Sigma_2|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_2\Sigma_2^{-1}+S_2\Sigma_2^{-1}\right)\right\},
\end{aligned}$$

where for ease of notation we defined $S_2=\sum_{t=1}^{T}\tilde{\mathbf{E}}_{(2),t}(\Sigma_3^{-1}\otimes\Sigma_1^{-1})\tilde{\mathbf{E}}'_{(2),t}$. Therefore, the posterior full conditional distribution of Σ_2 is given by:

$$p(\Sigma_2|\mathcal{B},\mathbf{Y},\Sigma_1,\Sigma_3) \sim \mathcal{IW}_{I_2}\left(\nu_2+TI_1I_3,\gamma\Psi_2+S_2\right). \tag{E.127}$$

E.8 Full conditional distribution of Σ_3

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_3 is conjugate. For ease of notation, define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B}\times_4\mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t}$ the mode-1 matricization of $\tilde{\mathcal{E}}_t$ and $\mathbf{Z}_3 = \Sigma_2^{-1}\otimes\Sigma_1^{-1}$. By exploiting the relation between the tensor normal distribution and the multivariate normal distribution and the properties of the vectorization and trace operators, we obtain:

$$p(\Sigma_3|\mathcal{B},\mathbf{Y},\Sigma_1,\Sigma_2,\gamma) \propto L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2,\Sigma_3)\pi(\Sigma_3|\gamma)$$

$$
\begin{aligned}
&\propto |\Sigma_3|^{-\frac{TI_1I_2}{2}} \exp\left\{ -\frac{1}{2}\sum_{t=1}^{T} \operatorname{vec}\left(\mathcal{Y}_t - \mathcal{B}\times_4 \mathbf{x}_t\right)' \left(\Sigma_3^{-1}\otimes\Sigma_2^{-1}\otimes\Sigma_1^{-1}\right) \operatorname{vec}\left(\mathcal{Y}_t - \mathcal{B}\times_4 \mathbf{x}_t\right) \right\} \\
&\cdot |\Sigma_3|^{-\frac{\nu_3+I_3+1}{2}} \exp\left\{ -\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) \right\} \\
&\propto |\Sigma_3|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\left\{ -\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^{T} \operatorname{vec}\left(\tilde{\mathcal{E}}_t\right)' \left(\Sigma_3^{-1}\otimes\mathbf{Z}_3\right) \operatorname{vec}\left(\tilde{\mathcal{E}}_t\right) \right] \right\} \\
&\propto |\Sigma_3|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\left\{ -\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^{T} \operatorname{vec}\left(\tilde{\mathbf{E}}_{(1),t}\right)' \left(\Sigma_3^{-1}\otimes\mathbf{Z}_3\right) \operatorname{vec}\left(\tilde{\mathbf{E}}_{(1),t}\right) \right] \right\} \\
&\propto |\Sigma_3|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\left\{ -\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^{T} \operatorname{tr}\left(\operatorname{vec}\left(\tilde{\mathbf{E}}_{(1),t}\right)' \operatorname{vec}\left(\mathbf{Z}_3\tilde{\mathbf{E}}_{(1),t}\Sigma_3^{-1}\right) \right) \right] \right\} \\
&\propto |\Sigma_3|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\left\{ -\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^{T} \operatorname{tr}\left(\tilde{\mathbf{E}}_{(1),t}'\mathbf{Z}_3\tilde{\mathbf{E}}_{(1),t}\Sigma_3^{-1} \right) \right] \right\} .
\end{aligned}
\tag{E.128}
$$

For ease of notation, define $S_3 = \sum_{t=1}^{T} \tilde{\mathbf{E}}_{(1),t}\mathbf{Z}_3\tilde{\mathbf{E}}_{(1),t}'$. Then:

$$
\begin{aligned}
p(\Sigma_3|\mathcal{B},\mathbf{Y},\Sigma_1,\Sigma_2) &\propto |\Sigma_3|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\left\{ -\frac{1}{2}\left[\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \operatorname{tr}\left(S_3\Sigma_3^{-1}\right) \right] \right\} \\
&\propto |\Sigma_3|^{-\frac{(\nu_3+TI_1I_2)+I_3+1}{2}} \exp\left\{ -\frac{1}{2}\operatorname{tr}\left(\left(\gamma\Psi_3 + S_3\right)\Sigma_3^{-1} \right) \right\} ,
\end{aligned}
\tag{E.129}
$$

Therefore, the posterior full conditional distribution of Σ_3 is given by:

$$
p(\Sigma_3|\mathcal{B},\mathbf{Y},\Sigma_1,\Sigma_2) \sim \mathcal{IW}_{I_3}\left(\nu_3 + TI_1I_2, \gamma\Psi_3 + S_3\right) . \tag{E.130}
$$

E.9 Full conditional distribution of γ

Using a gamma prior distribution we have:

$$
\begin{aligned}
p(\gamma|\Sigma_1,\Sigma_2,\Sigma_3) &\propto p(\Sigma_1,\Sigma_2,\Sigma_3|\gamma)\pi(\gamma) \\
&\propto \prod_{i=1}^{3} |\gamma\Psi_i|^{-\frac{\nu_i}{2}} \exp\left\{ -\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_i\Sigma_i^{-1}\right) \right\} \gamma^{a_\gamma-1}\exp\{-b_\gamma\gamma\} \\
&\propto \gamma^{a_\gamma - \frac{\sum_{i=1}^{3}\nu_iI_i}{2} - 1} \exp\left\{ -\frac{1}{2}\operatorname{tr}\left(\sum_{i=1}^{3}\Psi_i\Sigma_i^{-1} \right) - b_\gamma\gamma \right\} ,
\end{aligned}
\tag{E.131}
$$

thus:

$$
p(\gamma|\Sigma_1,\Sigma_2,\Sigma_3) \sim \mathcal{G}a\left(a_\gamma + \frac{1}{2}\sum_{i=1}^{3}\nu_iI_i, b_\gamma + \frac{1}{2}\operatorname{tr}\left(\sum_{i=1}^{3}\Psi_i\Sigma_i^{-1} \right) \right) . \tag{E.132}
$$

F Additional simulations' output

F.1 Simulation 10x10

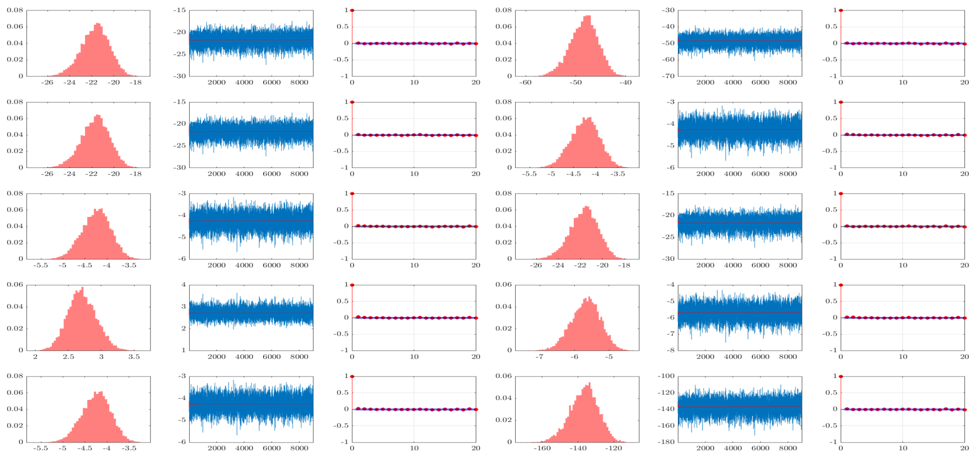

Figure 16: Posterior distribution (*first, fourth* columns), MCMC output (*second, fifth* columns) and autocorrelation function (*third, sixth* columns) of some entries of the estimated covariance matrix Σ_1.

F.2 Simulation 20x20

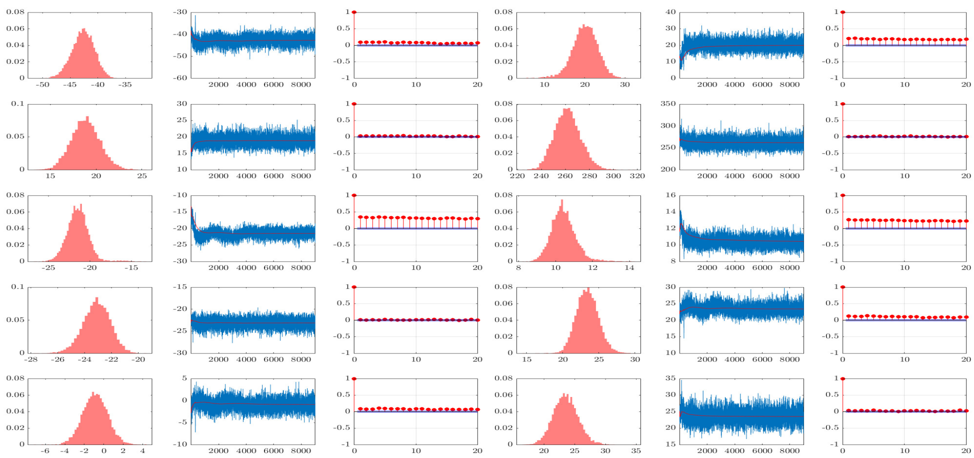

Figure 17: Posterior distribution (*first, fourth* columns), MCMC output (*second, fifth* columns) and autocorrelation function (*third, sixth* columns) of some entries of the estimated covariance matrix Σ_1.

G Additional application's output

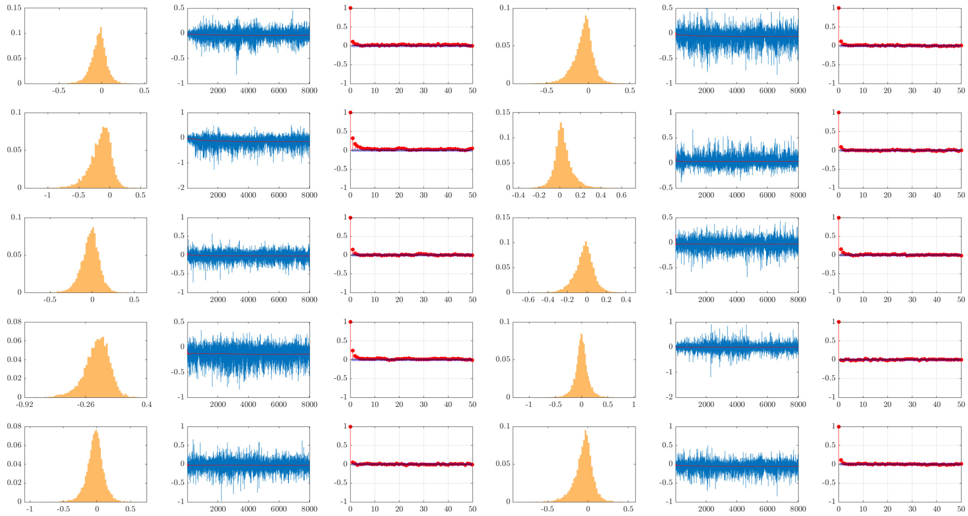

Figure 18: Posterior distribution (*first, fourth* columns), MCMC output (*second, fifth* columns) and autocorrelation function (*third, sixth* columns) of some entries of the estimated coefficient tensor.

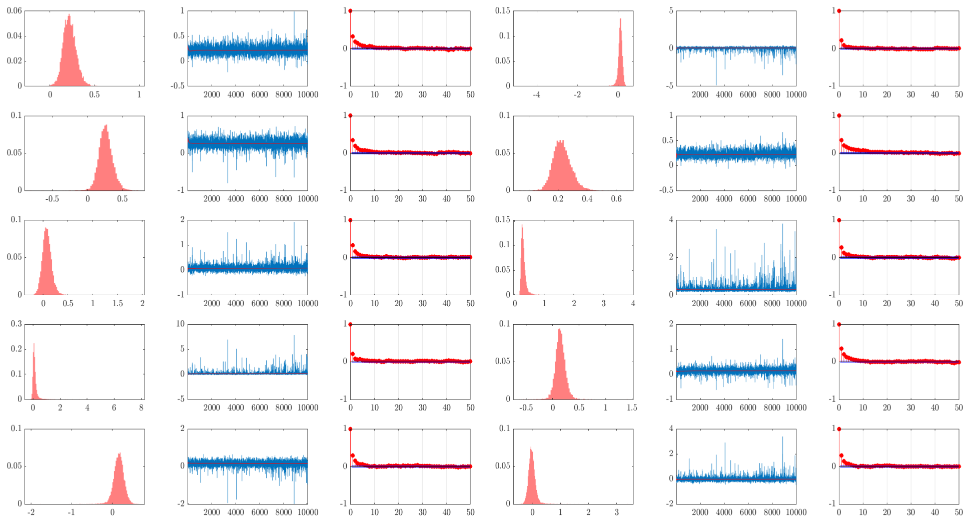

Figure 19: Posterior distribution (*first, fourth* columns), MCMC output (*second, fifth* columns) and autocorrelation function (*third, sixth* columns) of some entries of the estimated error covariance matrix Σ_1.

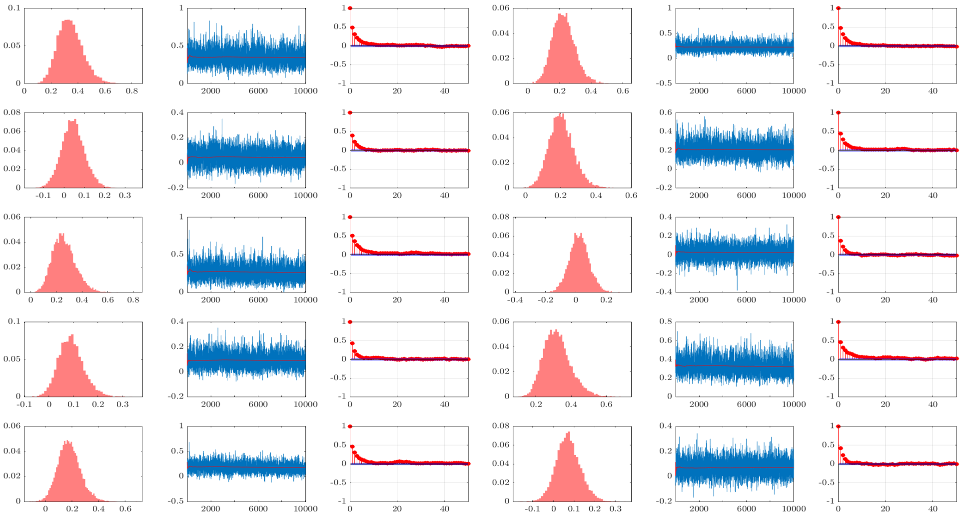

Figure 20: Posterior distribution (*first, fourth* columns), MCMC output (*second, fifth* columns) and autocorrelation function (*third, sixth* columns) of some entries of the estimated error covariance matrix Σ_2.