Bayesian Dynamic Tensor Regression

Journal:	*Journal of the Royal Statistical Society*
Manuscript ID	JRSS-SB-Jan-20-0040
Manuscript Type:	Series B
Date Submitted by the Author:	28-Jan-2020
Complete List of Authors:	Billio, Monica ; Ca'Foscari University of Venice Casarin, Roberto; Università Ca' Foscari Iacopini, Matteo; Scuola Normale Superiore Kaufmann, Sylvia; Study Center Gerzensee
Keywords:	Bayesian inference, dynamic networks, forecasting, multidimensional autoregression, tensor models

Bayesian Dynamic Tensor Regression†

Monica Billio
Ca' Foscari University of Venice, Italy
Roberto Casarin‡
Ca' Foscari University of Venice, Italy
E-mail: r.casarin@unive.it
Matteo Iacopini
Scuola Normale Superiore of Pisa, Italy
Sylvia Kaufmann
Study Center Gerzensee, Foundation of the Swiss National Bank, Switzerland

Abstract. High- and multi-dimensional array data are becoming increasingly available. They admit a natural representation as tensors and call for appropriate statistical tools. We propose a new linear autoregressive tensor process (ART) for tensor-valued data, that encompasses some well-known time series models as special cases. We study its properties and derive the associated impulse response function. We exploit the PARAFAC low rank decomposition for providing a parsimonious parametrization and develop a Bayesian inference allowing for shrinking effects. We apply the ART model to time series of multilayer networks and study the propagation of shocks across nodes, layers and time.

†We are grateful to Federico Bassetti, Sylvia Frühwirth-Schnatter, Christian Gouriéroux, Søren Johansen, Siem Jan Koopman, Gary Koop, André Lucas, Alain Monfort, Peter Phillips, Christian Robert, Mike West, for their comments and suggestions. Also, we thank the seminar participants at: CREST, University of Southampton, Vrije University of Amsterdam, London School of Economics, Maastricht University, Polytechnic University of Milan. Moreover, we thank the conference and workshop participants at: "ICEEE 2019" in Lecce, 2019, "CFENetwork 2018" in Pisa, 2018, "29th EC2 conference" in Rome, 2018, "12th RCEA Annual meeting" in Rimini, 2018, "8th MAF" in Madrid, 2018, "CFENetwork 2017" in London, 2017, "ICEEE 2017" in Messina, 2017, "3rd Vienna Workshop on High-dimensional Time Series in Macroeconomics and Finance" in Wien, 2017, "BISP10" in Milan, 2017, "ESOBE" in Venice, 2016, "CFENetwork" in Seville, 2016, and the "SIS Intermediate Meeting" of the Italian Statistical Society in Florence, 2016. This research used the SCSCF multiprocessor cluster system and is part of the project Venice Center for Risk Analytics (VERA) at Ca' Foscari University of Venice. Iacopini's research is supported by the "Bando Vinci 2016" grant by the Université Franco-Italienne.

‡*Address for correspondence:* Roberto Casarin, Department of Economics, Ca' Foscari University of Venice, Venice, 30123, Italy

Keywords: Bayesian inference; dynamic networks; forecasting; multidimensional autoregression; tensor models

1. Introduction

The increasing availability of long time series of complex-structured data, such as multidimensional tables (Balazsi et al. (2015), Carvalho and West (2007)), multidimensional panel data (Bayer et al. (2016), Baltagi et al. (2015), Davis (2002), Shin et al. (2019)), multilayer networks (Aldasoro and Alves (2018), Poledna et al. (2015)), EEG (Li and Zhang (2017)), neuroimaging (Zhou et al. (2013)) has put forward some limitations of the existing multivariate econometric models. Tensors, i.e. multidimensional arrays, are the natural class where this kind of complex data belongs.

A naïve approach to model tensors relies on reshaping them into lower-dimensional objects (e.g., vectors and matrices) which can then be easily handled using standard multivariate statistical tools. However, mathematical representations of tensor-valued data in terms of vectors have non-negligible drawbacks, such as the difficulty of accounting for the intrinsic structure of the data (e.g., cells of a matrix representing a geographical map or pairwise relations, contiguous pixels in an image). Neglecting this information in the modelling might lead to inefficient estimation and misleading results. Tensor-valued data entries are highly likely to depend on contiguous cells (within and between modes) and collapsing the data into a vector destroys this information. Thus, statistical approaches based on vectorization are unsuited for modelling tensor-valued data.

Tensors have been recently introduced in statistics and machine learning (e.g., Hackbusch (2012), Kroonenberg (2008)) and provide a fundamental background for efficient algorithms in *Big Data* handling (e.g., Cichocki (2014)). However, a compelling statistical approach extending results for scalar random variables to multidimensional random objects beyond dimension 2 (i.e., matrix-valued random variables, see Gupta and Nagar (1999)) is lacking and constitutes a promising field of research.

The development of novel statistical methods able to deal directly with tensor-valued data (i.e., without relying on vectorization) is currently an open field of research in statistics and econometrics, where such kind of data is becoming increasingly available. The main purpose of this article is to contribute to this growing literature by proposing an extension of standard multivariate econometric regression models to tensor-valued response and covariates.

Matrix-valued statistical models have been widely employed in time series econometrics over the past decades, especially for state space representations (Harrison and West (1999)), dynamic linear models (Carvalho and West (2007), Wang and West (2009)), Gaussian graphical models (Carvalho et al.

(2007)), stochastic volatility (Uhlig (1997), Gouriéroux et al. (2009), Golosnoy et al. (2012)), classification of longitudinal datasets (Viroli (2011)), models for network data (Durante and Dunson (2014), Zhu et al. (2017), Zhu et al. (2019)) and factor models (Chen et al. (2019)).

Ding and Cook (2018) proposed a bilinear multiplicative matrix regression model, which in vector form becomes a VAR(1) with restrictions on the covariance matrix. The main shortcoming in using bilinear models is the difficulty in introducing shrinking effects. Shrinking a subset of the reduced form coefficients towards zero implies similar effects on the structural coefficients.

Recent papers dealing with tensor-valued data include Zhou et al. (2013) and Xu et al. (2013), who proposed a generalized linear model to predict a scalar real or binary outcome by exploiting the tensor-valued covariate. Instead, Zhao et al. (2013), Zhao et al. (2014) and Imaizumi and Hayashi (2016) followed a Bayesian nonparametric approach for regressing a scalar on tensor-valued covariate. Another stream of the literature considers regression models with tensor-valued response and covariates. In this framework, Li and Zhang (2017) proposed a model for cross-sectional data where response and covariates are tensors, and performed sparse estimation by means of the envelope method and iterative maximum likelihood. Hoff (2015) employed the Tucker product to define a tensor-on-tensor regression extending the standard bilinear model to a multilinear one. Instead, we use the contracted product that generalises the Cayley matrix multiplication to tensors and allows us to define a linear regression model for tensors.

We introduce a new autoregressive tensor model (ART) which admits as special cases the standard linear models for time series, such as VARs (Lütkepohl (2005)). Taking advantage of the properties of the contracted product recently proved in Behera et al. (2019), Ji and Wei (2018), Wang et al. (2020), we derive new results on tensor algebra and study the main properties of the ART process. In addition, we derive the impulse response function and the forecast error variance decomposition for the ART, which are important tools to make prediction and to analyse the shock propagation.

We use the PARAFAC representation (Hackbusch (2012)) of the coefficient tensor to provide a parsimonious parametrization of the ART and to deal with the high dimensionality issue. We address further the overfitting by choosing a Bayesian perspective and using a global-local prior to promote coefficient shrinkage.

One of the areas where our model can find application is network data modelling. Most statistical models for network data are static (De Paula (2017)), whereas dynamic models maybe more adequate for many applications (e.g., banking) where network data are collected over time. Despite increasing data availability, few attempts have been made to model time-varying networks (e.g., Hoff (2015), Anacleto and Queen (2017)), and most of the contributions have focused on providing a representation and a description of temporally evolving graphs (e.g., Holme and Saramäki (2012), Kostakos (2009)). We provide an original study of time-varying economic and financial networks and show that our model can be successfully used to carry out impulse response analysis in this multidimensional setting.

The remainder of this paper is organized as follows. Section 2 provides an introduction to tensor algebra and presents the new modelling framework. Section 3 discusses parametrization strategies and a Bayesian inference procedure. Section 4 provides an empirical application and Section 5 gives some concluding remarks. Further details and results are provided in the supplementary material.

2. A Dynamic Tensor Model

In this section, we present a dynamic tensor regression model and discuss some of its properties and special cases. We review some notions of multilinear algebra which will be used in this paper, and refer the reader to Appendix A and the supplement for novel results on tensor algebra and further details.

2.1. Tensor Calculus and Decompositions

The use of tensors is well established in physics and mechanics (e.g., see Aris (2012) and Abraham et al. (2012)), but few contributions have been made beyond these disciplines. For a general introduction to the algebraic properties of tensor spaces, see Hackbusch (2012). Noteworthy introductions to operations on tensors and tensor decompositions are Lee and Cichocki (2018) and Kolda and Bader (2009), respectively.

A N-order real-valued tensor is a N-dimensional array $\mathcal{X} = (\mathcal{X}_{i_1,\ldots,i_N}) \in \mathbb{R}^{I_1\times\ldots\times I_N}$ with entries $\mathcal{X}_{i_1,\ldots,i_N}$ with $i_n = 1,\ldots,I_n$ and $n = 1,\ldots,N$. The *order* is the number of dimensions (also called modes). Vectors and matrices are examples of 1- and 2-order tensors, respectively. 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. We use the symbol ":" to indicate selection of all elements of a given mode of a tensor. The mode-k *fiber* is the vector obtained by fixing all but the k-th index of the tensor, i.e. the equivalent of rows and columns in a matrix. Tensor *slices* and their generalizations, are obtained by keeping fixed all but two or more dimensions of the tensor.

It can be shown that the set of N-order tensors $\mathbb{R}^{I_1\times\ldots\times I_N}$ endowed with the standard addition $\mathcal{A}+\mathcal{B} = (\mathcal{A}_{i_1,\ldots,i_N} + \mathcal{B}_{i_1,\ldots,i_N})$ and scalar multiplication $\alpha\mathcal{A} = (\alpha\mathcal{A}_{i_1,\ldots,i_N})$, with $\alpha \in \mathbb{R}$, is a vector space. We now introduce some operators on the set of real tensors, starting with the *contracted product*, which generalizes the matrix product to tensors. The contracted product between $\mathcal{X} \in \mathbb{R}^{I_1\times\ldots\times I_M}$ and $\mathcal{Y} \in \mathbb{R}^{J_1\times\ldots\times J_N}$ with $I_M = J_1$, is denoted by $\mathcal{X} \times_M \mathcal{Y}$ and yields a $(M+N-2)$-order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times \ldots \times I_{M-1} \times J_1 \times \ldots \times J_{N-1}}$, with entries

$$\mathcal{Z}_{i_1,\ldots,i_{M-1},j_2,\ldots,j_N} = (\mathcal{X} \times_M \mathcal{Y})_{i_1,\ldots,i_{M-1},j_2,\ldots,j_N} = \sum_{i_M=1}^{I_M} \mathcal{X}_{i_1,\ldots,i_{M-1},i_M} \mathcal{Y}_{i_M,j_2,\ldots,j_N}.$$

When $\mathcal{Y} = \mathbf{y}$ is a vector, the contracted product is also called *mode-M product*. We define with $\mathcal{X}\bar{\times}_N\mathcal{Y}$ a sequence of contracted products between the $(K+N)$-order tensor $\mathcal{X} \in \mathbb{R}^{J_1 \times \ldots \times J_K \times I_1 \times \ldots \times I_N}$ and the $(N+M)$-order tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times H_1 \times \ldots \times H_M}$. Entry-wise, it is defined as

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

Note that the contracted product is not commutative. The *outer product* $\circ$ between a M-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_M}$ and a N-order tensor $\mathcal{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ is a $(M+N)$-order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_N}$ with entries $\mathcal{Z}_{i_1,\ldots,i_M,j_1,\ldots,j_N} = (\mathcal{X} \circ \mathcal{Y})_{i_1,\ldots,i_M,j_1,\ldots,j_N} = \mathcal{X}_{i_1,\ldots,i_M} \mathcal{Y}_{j_1,\ldots,j_N}$.

Tensor decompositions allow to represent a tensor as a function of lower dimensional variables, such as matrices of vectors, linked by suitable multidimensional operations. In this paper, we use the low-rank parallel factor (PARAFAC) decomposition, which allows to represent a N-order tensor in terms of a collection of vectors (called marginals). A N-order tensor is of rank 1 when it is the outer product of N vectors. Let R be the rank of the tensor $\mathcal{X}$, that is minimum number of rank-1 tensors whose linear combination yields $\mathcal{X}$. The PARAFAC(R) decomposition is rank-R decomposition which represents a N-order tensor $\mathcal{B}$ as a finite sum of R rank-1 tensors $\mathcal{B}_r$ defined by the outer products of N vectors (called marginals) $\boldsymbol{\beta}_j^{(r)} \in \mathbb{R}^{I_j}$

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

The *mode-n matricization* (or unfolding), denoted by $\mathbf{X}_{(n)} = \mathrm{mat}_n(\mathcal{X})$, 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 $I_n \times I^*_{(-n)}$ with $I^*_{(-n)} = \prod_{i \neq n} I_i$. 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 $j = 1 + \sum_{m \neq n}(i_m - 1) \prod_{p \neq n}^{m-1} I_p$. 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 3-order tensor, its mode-1 matricization $\mathbf{X}_{(1)}$ is a $I_1 \times I_2 I_3$ matrix obtained by horizontally stacking the mode-$(1,2)$ slices of the tensor. The *vectorization* operator stacks all the elements in direct lexicographic order, forming a vector of length $I^* = \prod_i I_i$.

Other orderings are possible, 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 the relationship $\mathrm{vec}\left(\mathcal{X}\right) = \mathrm{vec}\left(\mathbf{X}_{(1)}\right)$, where $\mathrm{vec}\left(\mathbf{X}_{(1)}\right)$ stacks vertically into a vector the columns of the matrix $\mathbf{X}_{(1)}$. Many product operations have been defined for tensors (e.g., see Lee and Cichocki (2018)), but here we constrain ourselves to the operators used in this work. For the ease of notation, we will use the multiple-index summation for indicating the sum over all the corresponding indices.

Remark 2.1. *Consider a N-order tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ with a PARAFAC(R) decomposition (with marginals $\boldsymbol{\beta}_j^{(r)}$), 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*

$$\mathcal{Y} = \mathcal{B} \times_N \mathbf{x} \iff \mathrm{vec}\left(\mathcal{Y}\right) = \mathbf{B}_{(N)}' \mathbf{x} \iff \mathrm{vec}\left(\mathcal{Y}\right)' = \mathbf{x}' \mathbf{B}_{(N)}$$

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

2.2. A General Dynamic Tensor Model

Let $\mathcal{Y}_t$ be a $(I_1 \times \ldots \times I_N)$-dimensional tensor of endogenous variables, $\mathcal{X}_t$ a $(J_1 \times \ldots \times J_M)$-dimensional tensor of covariates, and $S_y = \times_{j=1}^{N} \{1, \ldots, I_j\} \subset \mathbb{N}^N$ and $S_x = \times_{j=1}^{M} \{1, \ldots, J_j\} \subset \mathbb{N}^M$ sets of n-tuples of integers. We define the autoregressive tensor model of order p, ART(p), as the system of equations

$$\mathcal{Y}_{\mathbf{i},t} = \mathcal{A}_{\mathbf{i},0} + \sum_{j=1}^{p} \sum_{\mathbf{k} \in S_y} \mathcal{A}_{\mathbf{i},\mathbf{k},j} \mathcal{Y}_{\mathbf{k},t-j} + \sum_{\mathbf{m} \in S_x} \mathcal{B}_{\mathbf{i},\mathbf{m}} \mathcal{X}_{\mathbf{m},t} + \mathcal{E}_{\mathbf{i},t}, \quad \mathcal{E}_{\mathbf{i},t} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{\mathbf{i}}^2), \quad (2)$$

$t = 1, 2, \ldots,$ with given initial conditions $\mathcal{Y}_{-p+1}, \ldots, \mathcal{Y}_0 \in \mathbb{R}^{I_1 \times \ldots \times I_N}$, where $\mathbf{i} = (i_1, \ldots, i_N) \in S_y$ and $\mathcal{Y}_{\mathbf{i},t}$ is the $\mathbf{i}$-th entry of $\mathcal{Y}_t$. The general model in eq. (2) allows for measuring the effect of all the cells of $\mathcal{X}_t$ and of the lagged values of $\mathcal{Y}_t$ on each endogenous variable.

We give two equivalent compact representations of the multilinear system (2). The first one is used for studying the stability property of the process and is obtained through the contracted product that provides a natural setting for multilinear forms, decompositions and inversions. From (2) one gets

$$\mathcal{Y}_t = \mathcal{A}_0 + \sum_{j=1}^{p} \widetilde{\mathcal{A}}_j \bar{\times}_N \mathcal{Y}_{t-j} + \widetilde{\mathcal{B}} \bar{\times}_M \mathcal{X}_t + \mathcal{E}_t, \quad \mathcal{E}_t \stackrel{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{O}, \Sigma_1, \ldots, \Sigma_N), \quad (3)$$

where $\bar{\times}_{a,b}$ is a shorthand notation for the contracted product $\times_{a+1 \ldots a+b}^{1 \ldots a}$ and $\bar{\times}_a$ is equivalent to $\bar{\times}_{a,0}$, $\widetilde{\mathcal{A}}_0$ is a N-order tensor of the same size as $\mathcal{Y}_t$, $\widetilde{\mathcal{A}}_j$, $j = 1, \ldots, p$, are $2N$-order tensors of size $(I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N)$ and $\mathcal{B}$ is a $(N+M)$-order tensor of size $(I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M)$. The error term $\mathcal{E}_t$ follows a N-order tensor normal distribution (Ohlson et al. (2013)) with probability density function

$$f_{\mathcal{E}}(\mathcal{E}) = \frac{\exp\Big(-\frac{1}{2}(\mathcal{E}-\mathcal{M})\bar{\times}_N\big(\circ_{j=1}^N \Sigma_j^{-1}\big)\bar{\times}_N(\mathcal{E}-\mathcal{M})\Big)}{(2\pi)^{I^*/2}\prod_{j=1}^N \left|\Sigma_j\right|^{I^*_{-j}/2}}, \tag{4}$$

where $I^* = \prod_i I_i$ and $I^*_{-i} = \prod_{j\neq i} I_j$, $\mathcal{E}$ and $\mathcal{M}$ are N-order tensors of size $I_1 \times \ldots \times I_N$. Each covariance matrix $\Sigma_j \in \mathbb{R}^{I_j \times I_j}$, $j = 1, \ldots, N$, accounts for the dependence along the corresponding mode of $\mathcal{E}$.

The second representation of the ART(p) in eq. (2) is used for developing inference. Let $\mathcal{K}_m$ be the $(I_1 \times \ldots \times I_N \times m)$-dimensional commutation tensor such that $\mathcal{K}_m^\sigma \bar{\times}_{N,0}\mathcal{K}_m = \mathbf{I}_m$, where $\mathcal{K}_m^\sigma$ is the tensor obtained by flipping the modes of $\mathcal{K}_m$. Define the $(I_1 \times \ldots \times I_N \times I^*)$-dimensional tensor $\mathcal{A}_j = \widetilde{\mathcal{A}}_j \bar{\times}_N \mathcal{K}_{I^*}$ and the $(I_1 \times \ldots \times I_N \times J^*)$-dimensional tensor $\mathcal{B} = \widetilde{\mathcal{B}}\bar{\times}_N \mathcal{K}_{J^*}$, with $J^* = \prod_j J_j$. We obtain $\mathcal{A}_j \times_{N+1} \operatorname{vec}\big(\mathcal{Y}_{t-j}\big) = \widetilde{\mathcal{A}}_j \bar{\times}_N \mathcal{Y}_{t-j}$ and the compact representation

$$\begin{aligned} &\mathcal{Y}_t = \mathcal{A}_0 + \sum_{j=1}^p \mathcal{A}_j \times_{N+1} \operatorname{vec}\big(\mathcal{Y}_{t-j}\big) + \mathcal{B} \times_{N+1} \operatorname{vec}\big(\mathcal{X}_t\big) + \mathcal{E}_t, \\ &\mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{O}, \Sigma_1, \ldots, \Sigma_N). \end{aligned} \tag{5}$$

Let $\mathbb{T} = (\mathbb{R}^{I_1\times\ldots\times I_N \times I_1 \times\ldots\times I_N}, \bar{\times}_N)$ be the space of $(I_1\times\ldots\times I_N\times I_1\times\ldots\times I_N)$-dimensional tensors endowed with the contracted product $\bar{\times}_N$. We define the identity tensor $\mathcal{I} \in \mathbb{T}$ to be the neutral element of $\bar{\times}_N$, that is the tensor whose entries are $\mathcal{I}_{i_1,\ldots,i_N,i_{N+1},\ldots,i_{2N}} = 1$ if $i_k = i_{k+N}$ for all $k = 1, \ldots, N$ and 0 otherwise. The inverse of a tensor $\mathcal{A} \in \mathbb{T}$ is the tensor $\mathcal{A}^{-1} \in \mathbb{T}$ satisfying $\mathcal{A}^{-1}\bar{\times}_N\mathcal{A} = \mathcal{A}\bar{\times}_N\mathcal{A}^{-1} = \mathcal{I}$. A complex number $\lambda \in \mathbb{C}$ and a nonzero tensor $\mathcal{X} \in \mathbb{R}^{I_1\times\ldots\times I_N}$ are called eigenvalue and eigenvector of the tensor $\mathcal{A} \in \mathbb{T}$ if they satisfy the multilinear equation $\mathcal{A}\bar{\times}_N\mathcal{X} = \lambda\mathcal{X}$. We define the spectral radius $\rho(\mathcal{A})$ of $\mathcal{A}$ to be the largest modulus of the eigenvalues of $\mathcal{A}$. We define a stochastic process to be weakly stationary if the first and second moment of its finite dimensional distributions are finite and constant in t. Finally, note that it is always possible to rewrite an ART(p) process as a ART(1) process on an augmented state space, by stacking the endogenous tensors along the first mode. Thus, without loss of generality, we focus on the case $p = 1$. We use the definition of inverse tensor, spectral radius and the convergence of power series of tensors to prove the following result.

LEMMA 2.1. *Every $(I_1\times I_2\times\ldots\times I_N\times I_1\times I_2\times\ldots\times I_N)$-dimensional ART(p) process $\mathcal{Y}_t = \sum_{k=1}^p \mathcal{A}_k\bar{\times}_N\mathcal{Y}_{t-j} + \mathcal{E}_t$ can be rewritten as a $(pI_1 \times I_2 \times \ldots \times I_N \times pI_1 \times I_2 \times \ldots \times I_N)$-dimensional ART(1) process $\underline{\mathcal{Y}}_t = \underline{\mathcal{A}}\bar{\times}_N\underline{\mathcal{Y}}_{t-1} + \underline{\mathcal{E}}_t$.*

Proposition 2.1 (Stationarity). *If $\rho(\widetilde{\mathcal{A}}_1) < 1$ and the process $\mathcal{X}_t$ is weakly stationary, then the ART process in eq. (3), with $p = 1$, is weakly stationary and admits the representation*

$$\mathcal{Y}_t = (\mathcal{I} - \widetilde{\mathcal{A}}_1)^{-1} \bar{\times}_N \widetilde{\mathcal{A}}_0 + \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N \widetilde{\mathcal{B}} \bar{\times}_M \mathcal{X}_{t-k} + \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N \mathcal{E}_{t-k}.$$

Proposition 2.2. *The VAR(p) in eq. (16) is weakly stationary if and only if the ART(p) in eq. (3) is weakly stationary.*

2.3. Parametrization

The unrestricted model in eq. (5) 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. The models in eqq. (5)-(3) are equivalent but the assuming a PARAFAC decomposition for the coefficient tensors leads to different degrees of parsimony, as shown in the following remark.

Remark 2.2 (Parametrization via contracted product). *The two models (5) and (3) combined with the PARAFAC decomposition for the tensor coefficients allow for different degree of parsimony. To show this, without loss of generality, focus on the coefficient tensor $\widetilde{\mathcal{A}}_1$ (similar argument holds for $\widetilde{\mathcal{A}}_j$, $j = 2, \ldots, p$ and $\widetilde{\mathcal{B}}$). By assuming a PARAFAC(R) decomposition for $\widetilde{\mathcal{A}}_1$ in (3) and for $\mathcal{A}_1$ in (5), we get, respectively*

$$\widetilde{\mathcal{A}}_1 = \sum_{r=1}^{R} \widetilde{\boldsymbol{\alpha}}_1^{(r)} \circ \ldots \circ \widetilde{\boldsymbol{\alpha}}_N^{(r)} \circ \widetilde{\boldsymbol{\alpha}}_{N+1}^{(r)} \circ \ldots \circ \widetilde{\boldsymbol{\alpha}}_{2N}^{(r)}, \quad \mathcal{A}_1 = \sum_{r=1}^{R} \boldsymbol{\alpha}_1^{(r)} \circ \ldots \circ \boldsymbol{\alpha}_N^{(r)} \circ \boldsymbol{\alpha}_{N+1}^{(r)},$$

The length of the vectors $\boldsymbol{\alpha}_j^{(r)}$ and $\widetilde{\boldsymbol{\alpha}}_j^{(r)}$ coincide for each $j = 1, \ldots, N$. However, $\boldsymbol{\alpha}_{N+1}^{(r)}$ has length I^* while $\widetilde{\boldsymbol{\alpha}}_{N+1}^{(r)}, \ldots, \widetilde{\boldsymbol{\alpha}}_{2N}^{(r)}$ have length $I_1, \ldots, I_N$, respectively. Therefore, the number of free parameters in the coefficient tensor $\mathcal{A}_1$ is $R(I_1 + \ldots + I_N + \prod_{j=1}^{N} I_j)$, while it is $2R(I_1 + \ldots + I_N)$ for $\widetilde{\mathcal{A}}_1$. This highlights the greater parsimony granted by the use of the PARAFAC(R) decomposition in model (3) as compared to model (5).

Remark 2.3 (Vectorization). *There is a relation between the $(I_1 \times \ldots \times I_N)$-dimensional ART(p) and a $(I_1 \cdot \ldots \cdot I_N)$-dimensional VAR(p) model. The vector form of (5) is*

$$\mathrm{vec}\left(\mathcal{Y}_t\right) = \mathrm{vec}\left(\mathcal{A}_0\right) + \sum_{j=1}^{p} \mathrm{mat}_{N+1}(\mathcal{A}_j)\, \mathrm{vec}\left(\mathcal{Y}_{t-j}\right) + \mathrm{mat}_{N+1}(\mathcal{B})\, \mathrm{vec}\left(\mathcal{X}_t\right) + \mathrm{vec}\left(\mathcal{E}_t\right)$$

(a) vectorized (b) contracted product $\bar{\times}_N$ as (5) (c) contracted product $\bar{\times}_N$ as (3)

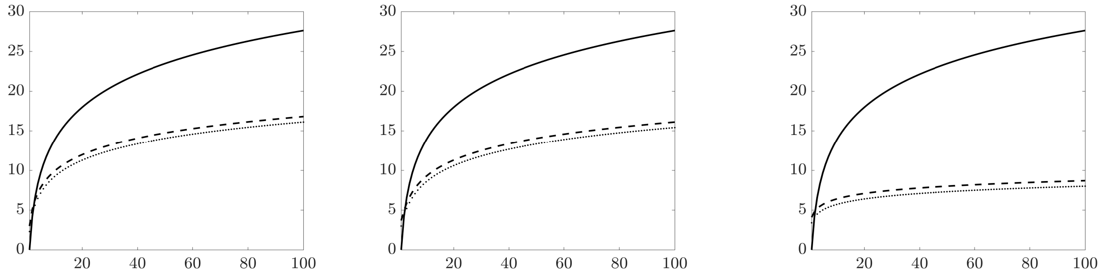

Figure 1: Number of parameters in $\mathcal{A}_0$, in log-scale (*vertical axis*) as function of the size d of the $(d \times d \times d)$-dimensional tensor $\mathcal{Y}_t$ (*horizontal axis*) in a ART(1) model. In all plots: unconstrained model (*solid* line), PARAFAC(R) parametrization with $R = 10$ (*dashed* line) and $R = 5$ (*dotted* line). Parametrizations: vectorized model (panel a), mode-n product of (5) (panel b) and contracted product of (3) (panel c).

$$\mathbf{y}_t = \boldsymbol{\alpha}_0 + \sum_{j=1}^{p} \mathbf{A}'_{(N+1),j}\mathbf{y}_{t-j} + \mathbf{B}'_{(N+1)}\mathbf{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}_{I^*}(\mathbf{0}, \Sigma_N \otimes \ldots \otimes \Sigma_1), \tag{6}$$

where the constraint on the covariance matrix stems from the one-to-one relation between the tensor normal distribution for $\mathcal{X}$ and the distribution of its vectorization (Ohlson et al. (2013)) given by $\mathcal{X} \sim \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_N)$ if and only if $\mathrm{vec}\left(\mathcal{X}\right) \sim \mathcal{N}_{I^*}(\mathrm{vec}\left(\mathcal{M}\right), \Sigma_N \otimes \ldots \otimes \Sigma_1)$. The restriction on the covariance structure for the vectorized tensor provides a parsimonious parametrization of the multivariate normal distribution, while allowing both within and between mode dependence. Alternative parametrizations for the covariance lead to generalizations of standard models. For example, assuming an additive covariance structure results in the tensor ANOVA. This is an active field for further research.

Example 2.1. *For the sake of exposition, consider the model in eq. (5), where $p = 1$, the response is a 3-order tensor $\mathcal{Y}_t \in \mathbb{R}^{d \times d \times d}$ and the covariates include only a constant coefficient tensor $\mathcal{A}_0$. Define by $k_{\mathcal{E}}$ the number of parameters of the noise distribution. The total number of parameters to estimate in the unrestricted case is $(d^{2N}) + k_{\mathcal{E}} = O(d^{2N})$, with $N = 3$ in this example. Instead, in a ART model defined via the mode-n product in eq. (5), assuming a PARAFAC(R) decomposition on $\mathcal{A}_0$ the total number of parameters is $\sum_{r=1}^{R}(d^N + d^N) + k_{\mathcal{E}} = O(d^N)$. Finally, in the ART model defined by the contracted product in eq. (3) with a PARAFAC(R) decomposition on $\widetilde{\mathcal{A}}_0$ the number of parameters is $\sum_{r=1}^{R} Nd + k_{\mathcal{E}} = O(d)$. A comparison of the different parsimony granted by the PARAFAC decomposition in all models is illustrated in Fig. 1.*

The structure of the PARAFAC decomposition poses an identification problem for the marginals $\boldsymbol{\beta}_j^{(r)}$, which may arise from three sources:

(i) *scale* identification, since $\lambda_{jr}\boldsymbol{\beta}_j^{(r)} \circ \lambda_{kr}\boldsymbol{\beta}_k^{(r)} = \boldsymbol{\beta}_j^{(r)} \circ \boldsymbol{\beta}_k^{(r)}$ for any collection $\{\lambda_{jr}\}_{j,r}$ such that $\prod_{j=1}^{J} \lambda_{jr} = 1$;

(ii) *permutation* identification, since $\boldsymbol{\beta}_j^{(\pi(r))} \circ \boldsymbol{\beta}_k^{(\pi(r))} = \boldsymbol{\beta}_j^{(r)} \circ \boldsymbol{\beta}_k^{(r)}$ for any permutation π of the indices $\{1, \ldots, R\}$;

(iii) *orthogonal transformation* identification, since $\boldsymbol{\beta}_j^{(r)} Q \circ \boldsymbol{\beta}_k^{(r)} Q = \boldsymbol{\beta}_j^{(r)} Q (\boldsymbol{\beta}_k^{(r)} Q)' = \boldsymbol{\beta}_j^{(r)} \circ \boldsymbol{\beta}_k^{(r)}$ for any orthonormal matrix Q.

Note that in our framework these issues do not hamper the inference, since our object of interest is the coefficient tensor $\mathcal{B}$, which is exactly identified. The marginals $\boldsymbol{\beta}_j^{(r)}$ have no interpretation, as the PARAFAC decomposition is assumed on the coefficient tensor for the sake of providing a parsimonious parametrization.

2.4. Important Special Cases

The model in eq. (5) is a generalization of several well-known linear econometric models, such as univariate regression, VARX, SUR, panel VAR, VECM and matrix autoregressive models (MAR). See Sections S.3-S.4 for further details. Dropping the covariates $\mathcal{X}_t$ from eq. (5), we obtain an autoregressive tensor model of order p (or ART(p))

$$\mathcal{Y}_t = \mathcal{A}_0 + \sum_{j=1}^{p} \mathcal{A}_j \times_{N+1} \operatorname{vec}\left(\mathcal{Y}_{t-j}\right) + \mathcal{E}_t, \qquad \mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathbf{0}, \Sigma_1, \ldots, \Sigma_N). \quad (7)$$

2.5. Impulse Response Analysis

In this section we derive two impulse response functions (IRF) for ART models, the block Cholesky IRF and the block generalised IRF, exploiting the relationship between ART and VAR models. Without loss of generality, we focus on the ART(p) model in eq. (7), with $p = 1$ and $\mathcal{A}_0 = \mathbf{0}$, and introduce the following notation. Let $\mathbf{y}_t = \operatorname{vec}\left(\mathcal{Y}_t\right)$ and $\boldsymbol{\epsilon}_t = \operatorname{vec}\left(\mathcal{E}_t\right) \sim \mathcal{N}_{I^*}(\mathbf{0}, \Sigma)$ be the $(I^* \times 1)$ tensor response and noise term in vector form, respectively, where $\Sigma = \Sigma_N \otimes \ldots \otimes \Sigma_1$ is the $(I^* \times I^*)$ covariance of the model in vector form and $I^* = \prod_{k=1}^{N} I_k$. Partition Σ in blocks as

$$\Sigma = \left(\begin{array}{c|c} A & B \\ \hline B' & C \end{array} \right), \quad (8)$$

where A is $n \times n$, B is $n \times (I^* - n)$ and C is $(I^* - n) \times (I^* - n)$. Then, denoting by $S = C - B'A^{-1}B$ the Schur complement of A, the LDU decomposition of Σ is

$$\Sigma = \left(\begin{array}{c|c} I_n & \mathrm{O}_{n,I^*-n} \\ \hline B'A^{-1} & I_{I^*-n} \end{array} \right) \left(\begin{array}{c|c} A & \mathrm{O}_{n,I^*-n} \\ \hline \mathrm{O}'_{n,I^*-n} & S \end{array} \right) \left(\begin{array}{c|c} I_n & A^{-1}B \\ \hline \mathrm{O}'_{n,I^*-n} & I_{I^*-n} \end{array} \right) = LDL'.$$

Hence Σ can be block-diagonalised

$$D = L^{-1}\Sigma(L')^{-1} = \left(\begin{array}{c|c} A & \mathrm{O}_{n,I^*-n} \\ \hline \mathrm{O}'_{n,I^*-n} & S \end{array} \right). \tag{9}$$

From the Cholesky decomposition of D one obtains a block Cholesky decomposition

$$\Sigma = \left(\begin{array}{c|c} L_A & \mathrm{O}_{n,I^*-n} \\ \hline B'(L_A^{-1})' & L_S \end{array} \right) \left(\begin{array}{c|c} L'_A & L_A^{-1}B \\ \hline \mathrm{O}'_{n,I^*-n} & L'_S \end{array} \right) = PP',$$

where L_A, L_S are the Cholesky factors of A and S, respectively. Assume the vectorised ART process admits an infinite MA representation, with $\Psi_0 = I_{I^*}$ and $\Psi_i = \mathrm{mat}_{(4)}(\mathcal{B})'\Psi_{i-1}$, then using the previous results we get:

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \Psi_i \boldsymbol{\epsilon}_{t-i} = \sum_{i=0}^{\infty} (\Psi_i L)(L^{-1}\boldsymbol{\epsilon}_{t-i}) = \sum_{i=0}^{\infty} (\Psi_i L)\boldsymbol{\eta}_{t-i} \quad \boldsymbol{\eta}_t \sim \mathcal{N}_{I^*}(\mathbf{0}, D), \tag{10}$$

where $\boldsymbol{\eta}_t = L^{-1}\boldsymbol{\epsilon}_t$ are the block-orthogonalised shocks and D is the block-diagonal matrix in eq. (9). Denote with E_n the $I^* \times n$ matrix that selects n columns from a pre-multiplying matrix, i.e. DE_n is a matrix containing n columns of D. Denote with $\boldsymbol{\delta}^*$ a n-dimensional vector of shocks. Using the property of the multivariate Normal distribution, and recalling that the top-left block of size n of D is A, we extend the generalised IRF of Koop et al. (1996) and Pesaran and Shin (1998) by defining the block generalised IRF

$$\begin{aligned} \boldsymbol{\psi}^G(h; n) &= \mathbb{E}\big(\mathrm{vec}\,(\mathcal{Y}_{t+h}) |\, \mathrm{vec}\,(\mathcal{E}_t)' = (\boldsymbol{\delta}^{*\prime}, \mathbf{0}'_{I^*-n}), \mathcal{F}_{t-1} \big) - \mathbb{E}\big(\mathrm{vec}\,(\mathcal{Y}_{t+h}) | \mathcal{F}_{t-1} \big) \\ &= (\Psi_h L) D E_n A^{-1} \boldsymbol{\delta}^*, \end{aligned} \tag{11}$$

where $\mathcal{F}_t$ is the natural filtration associated to the stochastic process. Starting from eq. (10) we derive the block Cholesky IRF (OIRF) as

$$\begin{aligned} \boldsymbol{\psi}^O(h; n) &= \mathbb{E}\big(\mathrm{vec}\,(\mathcal{Y}_{t+h}) |\, \mathrm{vec}\,(\mathcal{E}_t)' = (\boldsymbol{\delta}^{*\prime}, \mathbf{0}'_{I^*-n}), \mathcal{F}_{t-1} \big) \\ &\quad - \mathbb{E}\big(\mathrm{vec}\,(\mathcal{Y}_{t+h}) |\, \mathrm{vec}\,(\mathcal{E}_t)' = \mathbf{0}'_{I^*}, \mathcal{F}_{t-1} \big) \\ &= (\Psi_h L) P E_n \boldsymbol{\delta}^*. \end{aligned} \tag{12}$$

Define with $\mathbf{e}_j$ the j-th column of the I^*-dimensional identity matrix. The impact of a shock δ^* to the j-th variable on all I^* variables is given below in eq. (13), whereas the impact of a shock to the j-th variable on the i-th variable is given in eq. (14).

$$\boldsymbol{\psi}_j^G(h;n) = \Psi_h L D \mathbf{e}_j D_{jj}^{-1}\delta^*, \qquad \boldsymbol{\psi}_j^O(h;n) = \Psi_h L P \mathbf{e}_j \delta^* \tag{13}$$

$$\psi_{ij}^G(h;n) = \mathbf{e}_i'\Psi_h L D \mathbf{e}_j D_{jj}^{-1}\delta^*, \qquad \psi_{ij}^O(h;n) = \mathbf{e}_i'\Psi_h L P \mathbf{e}_j \delta^*. \tag{14}$$

Finally, denoting $\boldsymbol{\delta}_j = \mathbf{e}_j\delta^*$, we have the compact notation

$$\boldsymbol{\psi}_j^G(h;n) = \Psi_h L D D_{jj}^{-1}\boldsymbol{\delta}_j, \qquad \boldsymbol{\psi}_j^O(h;n) = \Psi_h L P \boldsymbol{\delta}_j$$
$$\psi_{ij}^G(h;n) = \mathbf{e}_i'\Psi_h L D D_{jj}^{-1}\boldsymbol{\delta}_j, \qquad \psi_{ij}^O(h;n) = \mathbf{e}_i'\Psi_h L P \boldsymbol{\delta}_j.$$

3. Bayesian Inference

In this section, without loss of generality, we present the inference procedure for a special case of the model in eq. (5), given by

$$\mathcal{Y}_t = \mathcal{B} \times_4 \operatorname{vec}\left(\mathcal{Y}_{t-1}\right) + \mathcal{E}_t, \qquad \mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,I_2,I_3}(\mathbf{0}, \Sigma_1, \Sigma_2, \Sigma_3). \tag{15}$$

Here $\mathcal{Y}_t$ is a 3-order tensor response of size $I_1 \times I_2 \times I_3$, $\mathcal{X}_t = \mathcal{Y}_{t-1}$ and $\mathcal{B}$ is thus a 4-order coefficient tensor of size $I_1 \times I_2 \times I_3 \times I_4$, with $I_4 = I_1 I_2 I_3$. This is a 3-order *tensor autoregressive model* of lag-order 1, or ART(1), coinciding with eq. (7) for $p = 1$ and $\mathcal{A}_0 = \mathbf{0}$. The noise term $\mathcal{E}_t$ has as tensor normal distribution, with zero mean and covariance matrices $\Sigma_1, \Sigma_2, \Sigma_3$ of sizes $I_1 \times I_1$, $I_2 \times I_2$ and $I_3 \times I_3$, respectively, accounting for the covariance along each of the three dimensions of $\mathcal{Y}_t$. The specification of a tensor model with a tensor normal noise instead of a vector model (like a Gaussian VAR) has the advantage of being more parsimonious. By vectorising (15), we get the equivalent VAR

$$\operatorname{vec}\left(\mathcal{Y}_t\right) = \mathbf{B}_{(4)}' \operatorname{vec}\left(\mathcal{Y}_{t-1}\right) + \operatorname{vec}\left(\mathcal{E}_t\right), \quad \operatorname{vec}\left(\mathcal{E}_t\right) \overset{iid}{\sim} \mathcal{N}_{I^*}(\mathbf{0}, \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1), \tag{16}$$

whose covariance has a Knocker structure, which contains $(I_1(I_1+1) + I_2(I_2+1) + I_3(I_3+1))/2$ parameters (as opposed to $(I^*(I^*+1))/2$ of an unrestricted VAR) and allows for heteroskedasticity.

The choice the Bayesian approach for inference is motivated by the fact that the large number of parameters may lead to an overfitting 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 distributions. In the unrestricted model (15) it would be necessary to define a prior distribution on the 4-order tensor $\mathcal{B}$. The literature on tensor-valued distributions is limited to the elliptical family (e.g., Ohlson et al. (2013)), which includes the tensor normal and tensor t. Both distributions do not easily allow for the specification of restrictions on a subset of the entries of the tensor, hampering the use of standard regularization prior distributions (such as shrinkage priors).

The PARAFAC(R) decomposition of the coefficient tensor provides a way to circumvent this issue. This decomposition allows to represent a tensor through a collection of vectors (the marginals), for which many flexible shrinkage prior distributions are available. Indirectly, this introduces *a priori* shrinkage to zero of the coefficient tensor.

3.1. Prior Specification

The choice of the prior distribution on the PARAFAC marginals is crucial for shrinking towards zero some elements of the coefficient tensor and for increasing the efficiency of the inference. Global-local prior distributions are based on scale mixtures of normal distributions, where the different components of the covariance matrix govern the amount of prior shrinkage. Compared to spike-and-slab distributions (e.g., Mitchell and Beauchamp (1988), George and McCulloch (1997), Ishwaran and Rao (2005)) which become infeasible as the parameter space grows, global-local priors have better scalability properties in high-dimensional settings. They do not provide automatic variable selection, which can nonetheless be obtained by post-estimation thresholding (Park and Casella (2008)).

Motivated by these arguments, we define a global-local shrinkage prior for the marginals $\boldsymbol{\beta}_j^{(r)}$ of the coefficient tensor $\mathcal{B}$ following the hierarchical prior specification of Guhaniyogi et al. (2017) (see also Bhattacharya et al. (2015), Zhou et al. (2015)). For each $\boldsymbol{\beta}_j^{(r)}$, we define a prior distributions as a scale mixture of normals centred in zero, with three components for the covariance. The global parameter τ governs the overall variance, the middle parameter ϕ_r defines the common shrinkage for the marginals in r-th component of the PARAFAC, and the local parameter $W_{j,r} = \mathrm{diag}(\mathbf{w}_{j,r})$ drives the shrinkage of each entry of each marginal. Summarizing, for $p = 1, \dots, I_j$, $j = 1, \dots, J$ ($J = 4$ in eq. (15)) and $r = 1, \dots, R$, the hierarchical prior structure (we use the shape-rate formulation for the gamma distribution) for each vector of the PARAFAC(R) decomposition in eq. (1) is

$$
\begin{gathered}
\pi(\boldsymbol{\phi}) \sim \mathcal{D}ir(\alpha \mathbf{1}_R) \quad \pi(\tau) \sim \mathcal{G}a(a_\tau, b_\tau) \quad \pi(\lambda_{j,r}) \sim \mathcal{G}a(a_\lambda, b_\lambda) \\
\pi(w_{j,r,p} | \lambda_{j,r}) \sim \mathcal{E}xp(\lambda_{j,r}^2/2) \\
\pi\big(\boldsymbol{\beta}_j^{(r)} \big| W_{j,r}, \boldsymbol{\phi}, \tau\big) \sim \mathcal{N}_{I_j}(\mathbf{0}, \tau \phi_r W_{j,r}),
\end{gathered}
\tag{17}
$$

where $\mathbf{1}_R$ is the vector of ones of length R and we assume $a_\tau = \alpha R$ and $b_\tau = \alpha R^{1/J}$. The conditional prior distribution of a generic entry $b_{i_1,\dots,i_J}$ of $\mathcal{B}$ is the law of a sum of product Normals (a product Normal is the distribution of the product of n independent centred Normal random variables): it is symmetric around zero, with fatter tails than both a standard Gaussian or a standard Laplace distribution (see Section S.5 for further details). The peak at zero of the product Normal prior promotes shrinking effects. The following result characterises the conditional prior distribution of an entry of the coefficient tensor $\mathcal{B}$ induced by the hierarchical prior in eq. (17). See Section S.5 for the proof.

LEMMA 3.1. *Let $b_{ijkp} = \sum_{r=1}^{R} \beta_r$, where $\beta_r = \beta_{1,i}^{(r)} \beta_{2,j}^{(r)} \beta_{3,k}^{(r)} \beta_{4,p}^{(r)}$, and let $m_1 = i$, $m_2 = j$, $m_3 = k$ and $m_4 = p$. Under the prior specification in (17), the generic entry b_{ijkp} of the coefficient tensor $\mathcal{B}$ has the conditional prior distribution*

$$\pi(b_{ijkp}|\tau, \boldsymbol{\phi}, \mathbf{W}) = p\left(\sum_{r=1}^{R} \beta_r \Big| -\right) = p(\beta_1|-) * \ldots * p(\beta_R|-),$$

where $*$ denotes convolution and

$$p(\beta_r|-) = K_r \cdot G_{4,0}^{4,0}\left(\beta_r^2 \prod_{h=1}^{4} (2\tau\phi_r w_{h,r,m_h})^{-1} \Big| \mathbf{0}\right),$$

with $G_{p,q}^{m,n}(x| \begin{smallmatrix}\mathbf{a}\\ \mathbf{b}\end{smallmatrix})$ a Meijer G-function and

$$G_{4,0}^{4,0}\left(\beta_r^2 \prod_{h=1}^{4} (2\tau\phi_r w_{h,r,m_h})^{-1} \Big| \mathbf{0}\right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\beta_r^2 \prod_{h=1}^{4} (2\tau\phi_r w_{h,r,m_h})^{-1}\right)^{-s} \mathrm{d}s$$

$$K_r = (2\pi)^{-4/2} \prod_{h=1}^{4} (2\tau\phi_r w_{h,r,m_h})^{-1}.$$

The use of Meijer G- and Fox H-functions is not new in econometrics (e.g., Abadir and Paruolo (1997)), and they have been recently used for defining prior distributions in Bayesian statistics (Andrade and Rathie (2015), Andrade and Rathie (2017)).

From eq. (4), we have that the covariance matrices Σ_j enter the likelihood in a multiplicative way, therefore separate identification of their scales requires further restrictions. Wang and West (2009) and Dobra (2015) adopt independent hyper-inverse Wishart prior distributions (Dawid and Lauritzen (1993)) for each Σ_j, then impose the identification restriction $\Sigma_{j,11} = 1$ for $j = 2, \ldots, J-1$. The hard constraint $\Sigma_j = \mathbf{I}_{I_j}$ (where $\mathbf{I}_j$ is the identity matrix of size j), for all but one n, implicitly imposes that the dependence structure within different modes is the same, but there is no dependence between modes. We follow Hoff (2011), who suggests to introduce dependence between the Inverse Wishart prior distribution of each Σ_j via a hyper-parameter γ affecting their prior scale. To account for marginal dependence, we add a level of hierarchy, thus obtaining

$$\pi(\gamma) \sim \mathcal{G}a(a_\gamma, b_\gamma) \qquad \pi(\Sigma_j|\gamma) \sim \mathcal{IW}_{I_j}(\nu_j, \gamma\Psi_j). \tag{18}$$

Define $\Lambda = \{\lambda_{j,r} : j = 1, \ldots, J, r = 1, \ldots, R\}$ and $\mathbf{W} = \{W_{j,r} : j = 1, \ldots, J, r = 1, \ldots, R\}$, and let $\boldsymbol{\theta}$ denote the collection of all parameters. The directed acyclic graph (DAG) of the prior structure is given in Fig. 2.

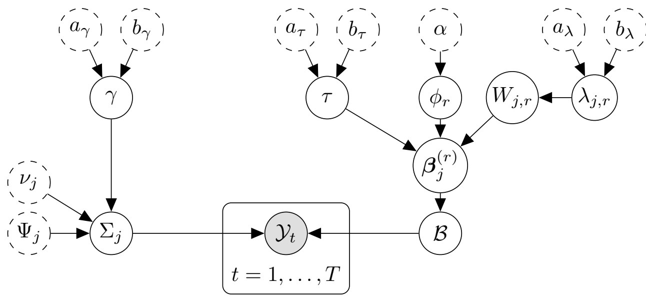

Figure 2: Directed acyclic graph of the model in eq. (15) and prior structure in eqq. (17)-(18). Gray circles denote observable variables, white solid circles indicate parameters, white dashed circles indicate fixed hyperparameters. Directed edges represent the conditional independence relationships.

Note that our prior specification is flexible enough to include Minnesota-type restrictions or hierarchical structures as in Canova and Ciccarelli (2004).

3.2. Posterior Computation

Define $\mathbf{Y} = \{\mathcal{Y}_t\}_{t=1}^T$, $I_0 = \sum_{j=1}^J I_j$, $\boldsymbol{\beta}_{-j}^{(r)} = \{\boldsymbol{\beta}_i^{(r)} : i \neq j\}$ and $\mathcal{B}_{-r} = \{B_i : i \neq r\}$, with $B_r = \boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_4^{(r)}$. The likelihood function of model (15) is

$$L(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} (2\pi)^{-\frac{I_4}{2}} \prod_{j=1}^{3} \left|\Sigma_j\right|^{-\frac{I_{-j}}{2}} \tag{19}$$
$$\cdot \exp\Big(-\frac{1}{2} \Sigma_2^{-1} (\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{y}_{t-1}) \times_{1\ldots3}^{1\ldots3} \big(\circ_{j=1}^{3} \Sigma_j^{-1} \big) \times_{1\ldots3}^{1\ldots3} (\mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{y}_{t-1}) \Big),$$

where $\mathbf{y}_{t-1} = \mathrm{vec}\left(\mathcal{Y}_{t-1}\right)$. Since the posterior distribution is not tractable in closed form, we adopt an MCMC procedure based on Gibbs sampling. The technical details of the derivation of the posterior distributions are given in Appendix B. We articulate the sampler in three main blocks:

(I) sample the global and middle variance hyper-parameters of the marginals, from

$$p(\psi_r|\mathcal{B}, \mathbf{W}, \alpha) \propto \mathrm{GiG}\big(\alpha - I_0/2,\, 2b_\tau,\, 2C_r\big) \tag{20}$$

$$p(\tau|\mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) \propto \mathrm{GiG}\Big(a_\tau - RI_0/2,\, 2b_\tau,\, 2\sum_{r=1}^{R} C_r/\phi_r\Big), \tag{21}$$

where $C_r = \sum_{j=1}^J \boldsymbol{\beta}_j^{(r)\prime} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)}$, then set $\phi_r = \psi_r / \sum_{l=1}^R \psi_l$. To improve the mixing, we sample τ with a Hamiltonian Monte Carlo (HMC) step (Neal (2011)).

(II) sample the local variance hyper-parameters of the marginals and the marginals themselves, from

$$p\big(\lambda_{j,r}|\boldsymbol{\beta}_j^{(r)},\phi_r,\tau\big) \propto \mathcal{G}a\big(a_\lambda + I_j,\, b_\lambda + \big\|\boldsymbol{\beta}_j^{(r)}\big\|_1(\tau\phi_r)^{-1/2}\big) \tag{22}$$

$$p\big(w_{j,r,p}|\lambda_{j,r},\phi_r,\tau,\boldsymbol{\beta}_j^{(r)}\big) \propto \mathrm{GiG}\big(1/2,\, \lambda_{j,r}^2,\, (\beta_{j,p}^{(r)})^2/(\tau\phi_r)\big) \tag{23}$$

$$p\big(\boldsymbol{\beta}_j^{(r)}|\boldsymbol{\beta}_{-j}^{(r)},\mathcal{B}_{-r},W_{j,r},\phi_r,\tau,\mathbf{Y},\Sigma_1,\ldots,\Sigma_3\big) \propto \mathcal{N}_{I_j}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_j},\, \bar{\Sigma}_{\boldsymbol{\beta}_j}). \tag{24}$$

(III) sample the covariance matrices and the latent scale, from

$$p(\Sigma_j|\mathcal{B},\mathbf{Y},\Sigma_{-j},\gamma) \propto \mathcal{IW}_{I_j}(\nu_j + I_j,\, \gamma\Psi_j + S_j) \tag{25}$$

$$p(\gamma|\Sigma_1,\ldots,\Sigma_3) \propto \mathcal{G}a\Big(a_\gamma + \sum_{j=1}^{3}\nu_j I_j,\, b_\gamma + \sum_{j=1}^{3}\mathrm{tr}(\Psi_j\Sigma_j^{-1})\Big). \tag{26}$$

4. Application to Multilayer Dynamic Networks

We apply the proposed methodology to study jointly the dynamics of international trade and credit networks. The international trade network has been previously studied by several authors (e.g., Fieler (2011), Eaton and Kortum (2002)), but to the best of our knowledge, this is the first attempt to model the dynamics of two networks jointly. Moreover, the impulse response analysis in this setting can be used for predicting possible trade creation and diversion effects (e.g., see Bikker (2010)).

The bilateral trade data come from the COMTRADE database, whereas the data on bilateral outstanding credit come from the Bank of International Settlements database. Our sample of yearly observations for 10 countries runs from 2003 to 2016. At at each time t, the 3-order tensor $\mathcal{Y}_t$ has size $(10, 10, 2)$ and represents a 2-layer node-aligned network (or multiplex) with 10 vertices (countries), where each edge is given by a bilateral trade flow or financial stock. See Section S.9 for data description.

We estimate the tensor autoregressive model in eq. (15), using the prior structure described in Section 3, running the Gibbs sampler for $N = 100,000$ iterations after $30,000$ burn-in iterations. We retain every second draw for posterior inference.

The mode-4 matricization of the estimated coefficient tensor, $\hat{B}_{(4)}$, is shown in the left panel of Fig. 3. The (i,j)-th entry of the matrix $\hat{B}_{(4)}$ reports the impact of the edge j on edge i in vectorised form (e.g., $j = 21$ and $i = 4$ corresponds to the coefficient of entry $\mathcal{Y}_{1,3,1,t-1}$ on $\mathcal{Y}_{4,1,1,t}$). The first 100 rows/columns correspond to the edges in the first layer. Hence, two rows of the matricized coefficient tensor are similar when two edges are affected by all the edges of the (lagged) network in a similar way, whereas two similar columns identify the situation where two edges impact the (next period) network in a similar way. The overall distribution

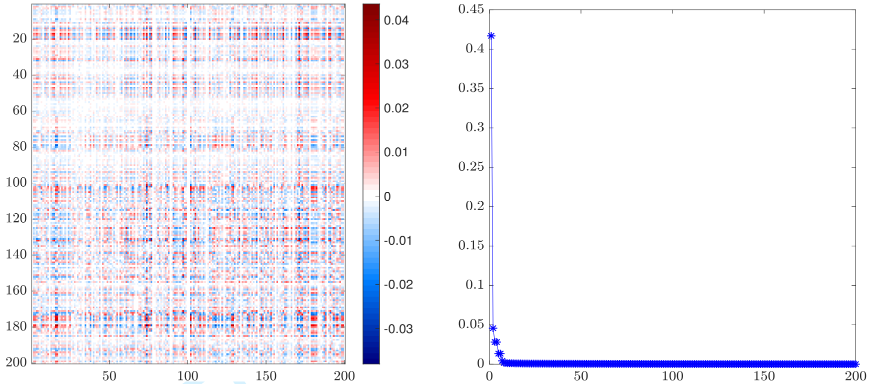

Figure 3: *Left:* mode-4 matricization of estimated coefficient tensor $\hat{B}_{(4)}$. *Right:* log-spectrum of $\hat{B}_{(4)}$, decreasing order.

of the estimated entries of $\hat{B}_{(4)}$ is symmetric around zero and leptokurtic, as a consequence of the shrinkage to zero of the estimated coefficients. The right panel of Fig. 3 shows the log-spectrum of $\hat{B}_{(4)}$. As all eigenvalues of $\hat{B}_{(4)}$ have modulus smaller than one, we conclude that the estimated ART(1) model is stationary. In fact, it can be shown that the stationarity of the mode-4 matricised coefficient tensor implies stationarity of the ART(1) process. Additional estimation results are provided in Section S.10.

After estimating the ART(1) model (15), we may investigate shock propagation across the network computing generalised and orthogonalised impulse response functions presented in equations (11) and (12), respectively. Impulse responses allow us to analyse the propagation of shocks both across the network, within and across layers, and over time. For illustration, we study the responses to a shock in all edges of country, by applying block Cholesky factorisation to Σ, in such a way that the shocked country contemporaneously affects all others and not vice-versa (we do not report generalised IRFs, which are very similar). Thus, the matrices A and C in eq. (8) reflect contemporaneous correlations across transactions of the shock-originating country and with transactions of all other countries, respectively. For expositional convenience, we report only statistically significant responses.

In this analysis we consider a negative 1% shock to US trade imports (i.e., we allocate the shock across import originating countries to match import shares as in the last period of the sample). The results of the block Cholesky IRF at horizon 1 are given in Fig. 4. We report the impact on the whole network (panel (a)) and, for illustrative purposes, the impact on Germany's transactions. The main findings follows.

Global effect on the network. The negative shock to US imports has an effect on both layers (trade and financial) of the network. There is evidence of heterogeneous responses across countries and country-specific transactions.

On average, trade flows exhibit a slight expansion in response to the shock. Switzerland is the most positively affected, both in terms of exports and imports, and trade imports of the US show (on average) a reverted positive response one period after the shock. This reflects an oscillating impulse response. The overall average effect on the financial layer is negative, similar in magnitude to the effect on the trade layer. More specifically, we observe that Denmark's and Sweden's exports to Switzerland, Germany and France show a contraction, whereas the effect on US's, Japan's and Ireland's exports to these countries is positive. We may interpret these effects as substitution effects: The decreasing share of Denmark's and Sweden's exports to Switzerland, Germany and France is offset by an increase of US, Japanese and Irish exports. In conclusion, model (15) permits to forecast trade creation and diversion effects (Bikker (2010)).

Local effect on Germany. In panel (b) of Fig. 4 we report the response of Germany's transactions to the negative shock in US imports. The effects on imports are mixed: while Germany's imports from most other EU countries increase, imports from Sweden and Denmark decrease. Likewise, Germany's exports show heterogeneous responses, whereby exports to Switzerland react strongest (positively). The shock of US imports does not have a significant impact on Germany's outstanding credit against most countries (except Switzerland and Japan). On the other hand, the reactions of Germany's outstanding debt reflect those on trade imports.

Local effect on other countries. We observe that the most affected trade transactions are those of Denmark, Japan, Ireland, Sweden and US (as exporters) vis-à-vis Switzerland and France (as importers). The financial layer mirrors these effects with opposite sign, while the magnitudes are comparable. Outstanding credit of Ireland and Japan to Switzerland, Germany and France decrease at horizon 1. By contrast, Denmark's outstanding credit to these countries increases. Note that outstanding debt of US vis-à-vis almost all countries decreases after the shock. Overall, responses to a shock on US imports at horizon 1 are heterogeneous in sign but rather low in magnitude, whereas at horizon 2 (plot not reported) the propagation of the shock has vanished. We interpret this as a sign of fast (and monotone) decay of the IRF.

In addition, Section S.10 in the Supplement shows the results of additional impulse response analyses consisting in (i) a negative 1% shock to GB's outstanding debt and (ii) a 1% negative shock to GB's outstanding debt coupled with a 1% positive shock to GB's outstanding credit.

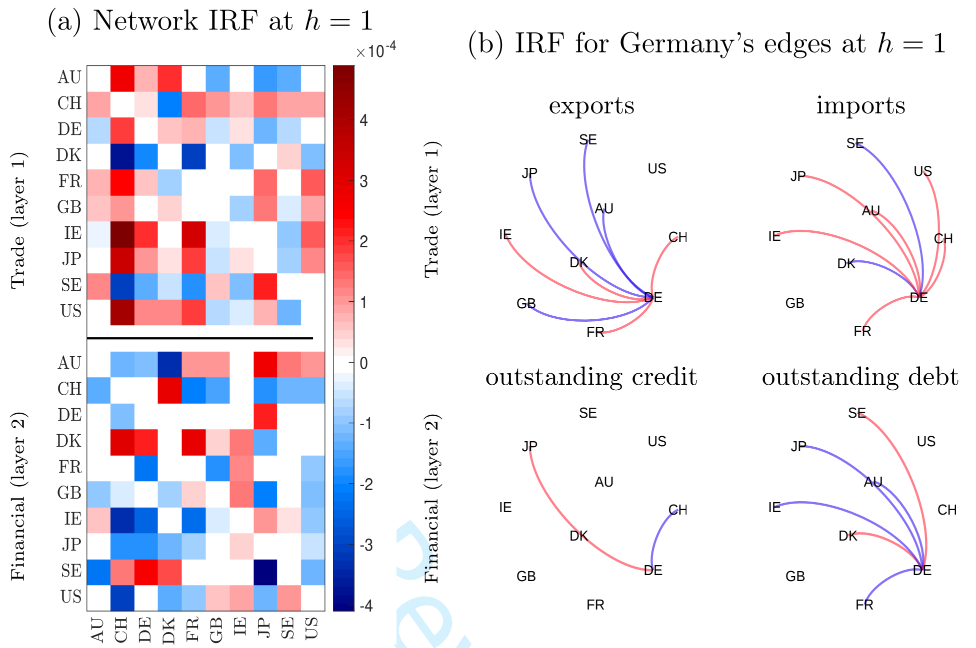

Figure 4: Shock to US trade imports by -1%. IRF at horizon $h = 1$ for all (*panel a*) and Germany (*panel b*) financial and trade transactions. In each plot negative coefficients are in blue and positive in red.

5. Conclusions

We defined a new statistical framework for dynamic tensor regression, customarily called ART. It encompasses a large number of models frequently used in time series analysis as special cases, such as VAR, panel VAR, SUR and MAR models. We exploited a low rank decomposition of the regression coefficient to reduce the dimension of the parameter space and specified a global-local shrinkage prior to address further the overfitting. Taking advantage of the properties of the contracted product, we studied the main properties of the ART process and derived the impulse response function and the forecast error variance decomposition, which are important tools for making predictions.

The proposed methodology has been applied to a time series of international trade and financial multilayer network. We are able to provide evidence of stationarity of the network process, heterogeneity in the shock propagation across countries and over time.

References

Abadir, K. M. and P. Paruolo (1997). Two mixed normal densities from cointegration analysis. *Econometrica*, 671–680.

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

Aldasoro, I. and I. Alves (2018). Multiplex interbank networks and systemic importance: an application to European data. *Journal of Financial Stability 35*, 17–37.

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

Andrade, J. and P. Rathie (2017). Exact posterior computation in non-conjugate gaussian location-scale parameters models. *Communications in Nonlinear Science and Numerical Simulation 53*, 111–129.

Andrade, J. and P. N. Rathie (2015). On exact posterior distributions using h-functions. *Journal of Computational and Applied Mathematics 290*, 459–475.

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

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

Baltagi, B. H., G. Bresson, and J.-M. Etienne (2015). Hedonic housing prices in paris: An unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics 30*(3), 509–528.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2016). A dynamic model of demand for houses and neighborhoods. *Econometrica 84*(3), 893–942.

Behera, R., A. K. Nandi, and J. K. Sahoo (2019). Further results on the drazin inverse of even order tensors. *arXiv preprint arXiv:1904.10783*.

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

Bikker, J. A. (2010). *The gravity model in international trade: advances and applications*, Chapter An extended gravity model with substitution applied to international trade, pp. 135–164. Cambridge University Press.

Brazell, M., N. Li, C. Navasca, and C. Tamon (2013). Solving multilinear systems via tensor inversion. *SIAM Journal on Matrix Analysis and Applications 34*(2), 542–570.

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

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

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

446 Chen, E. Y., R. S. Tsay, and R. Chen (2019). Constrained factor models for high-
447 dimensional matrix-variate time series. *Journal of the American Statistical*
448 *Association* ((forthcoming)).

449 Cichocki, A. (2014). Era of Big data processing: a new approach via tensor
450 networks and tensor decompositions. In *Proceedings of the International*
451 *Workshop on Smart Info-Media Systems in Asia (SISA2013).*

452 Davis, P. (2002). Estimating multi-way error components models with
453 unbalanced data structures. *Journal of Econometrics 106*(1), 67–95.

454 Dawid, A. P. and S. L. Lauritzen (1993). Hyper Markov laws in the statistical
455 analysis of decomposable graphical models. *The Annals of Statistics 21*(3),
456 1272–1317.

457 De Paula, A. (2017). Econometrics of network models. In *Advances in Economics*
458 *and Econometrics: Theory and Applications, Eleventh World Congress*, pp.
459 268–323. Cambridge University Press Cambridge.

460 Ding, S. and R. D. Cook (2018). Matrix-variate regressions and envelope
461 models. *Journal of the Royal Statistical Society: Series B (Statistical*
462 *Methodology) 80*(2), 387–408.

463 Dobra, A. (2015). *Handbook of Spatial Epidemiology* (First ed.)., Chapter
464 Graphical Modeling of Spatial Health Data. Chapman & Hall /CRC.

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

467 Eaton, J. and S. Kortum (2002). Technology, geography, and trade.
468 *Econometrica 70*(5), 1741–1779.

469 Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a
470 quantitative explanation. *Econometrica 79*(4), 1069–1101.

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

473 Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012). The conditional
474 autoregressive wishart model for multivariate stock market volatility. *Journal*
475 *of Econometrics 167*(1), 211–223.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181.

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

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

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

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

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

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

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

Ji, J. and Y. Wei (2018). The Drazin inverse of an even-order tensor and its application to singular tensor equations. *Computers & Mathematics with Applications 75*(9), 3402–3413.

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

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1), 119–147.

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.

Lee, N. and A. Cichocki (2018). Fundamental tensor operations for large-scale data analysis in tensor train formats. *Multidimensional Systems and Signal Processing 29*(3), 921–960.

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

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mitchell, T. J. and J. J. Beauchamp (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 S. Brooks, A. Gelman, J. L. Galin, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Chapter 5. Chapman & Hall /CRC.

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

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

Pesaran, H. H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters 58*(1), 17–29.

Poledna, S., J. L. Molina-Borboa, S. Martínez-Jaramillo, M. Van Der Leij, and S. Thurner (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.

Shin, Y., L. Serlenga, and G. Kapetanios (2019). Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure. *Journal of Econometrics (forthcoming)*.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica 65*, 59–73.

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

Wang, B., H. Du, and H. Ma (2020). Perturbation bounds for dmp and cmp inverses of tensors via einstein product. *Computational and Applied Mathematics 39*(1), 28.

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

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

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

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

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

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

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics 45*(3), 1096–1123.

Zhu, X., W. Wang, H. Wang, and W. K. Härdle (2019). Network quantile autoregression. *Journal of Econometrics 212*(1), 345–358.

A. Background Material on Tensor Calculus

This appendix provides the main tools used in the paper. See the supplement for further results and details. 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.

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 $\mathcal{X}_{(S_1, \ldots, S_M)} = (\Pi_{S_1} \mathcal{X}) \otimes \ldots \otimes (\Pi_{S_M} \mathcal{X}) = U_1 \otimes \ldots \otimes U_M$. The mapping is an isomorphism between $V_1 \otimes \ldots \otimes V_N$ and $U_1 \otimes \ldots \otimes U_M$.

The matricization is a particular case of reshaping a N-order tensor into a 2-order tensor, by choosing a mapping between the tensor modes and the rows and columns of the resulting matrix, then permuting the tensor and reshaping it, accordingly.

DEFINITION A.2 (MATRICIZATION). *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 $\mathbf{N} = \{1, \ldots, N\}$. The matricized tensor is defined by*

$$\mathrm{mat}_{\mathscr{R},\mathscr{C}}(\mathcal{X}) = \mathbf{X}_{(\mathscr{R},\mathscr{C})} \in \mathbb{R}^{J\times K}, \qquad J = \prod_{n\in\mathscr{R}} I_n, \quad K = \prod_{n\in\mathscr{C}} I_n .$$

Indices of $\mathscr{R}, \mathscr{C}$ are mapped to the rows and the columns, respectively, and

$$\left(\mathbf{X}_{(\mathscr{R}\times\mathscr{C})}\right)_{j,k} = \mathcal{X}_{i_1,i_2,\ldots,i_N}, \quad j = 1+\sum_{l=1}^{L}\left((i_{r_l}-1)\prod_{l'=1}^{l-1} I_{r'_l}\right), \quad k = 1+\sum_{m=1}^{M}\left((i_{c_m}-1)\prod_{m'=1}^{m-1} I_{c'_m}\right).$$

The *inner product* between two $(I_1\times\ldots\times I_N)$-dimensional tensors $\mathcal{X}, \mathcal{Y}$ is defined as

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

The PARAFAC(R) decomposition (e.g., see Kolda and Bader (2009)), is rank-R 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 (called marginals) $\boldsymbol{\beta}_j^{(r)} \in \mathbb{R}^{I_j}$

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

LEMMA A.1 (CONTRACTED PRODUCT – SOME PROPERTIES). *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 $(\mathscr{S}_1, \mathscr{S}_2)$ be a partition of $\{1, \ldots, N+P\}$, where $\mathscr{S}_1 = \{1, \ldots, N\}$, $\mathscr{S}_2 = \{N+1, \ldots, N+P\}$. It holds:*

(i) if $P = 0$ and $I_n = J_n$, $n = 1, \ldots, N$, then $\mathcal{X}\bar{\times}_N\mathcal{Y} = \langle\mathcal{X}, \mathcal{Y}\rangle = \mathrm{vec}\left(\mathcal{X}\right)'\cdot\mathrm{vec}\left(\mathcal{Y}\right)$.

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

$$\mathcal{X}\bar{\times}_N\mathcal{Y} = \mathrm{vec}\left(\mathcal{X}\right)\times_1\mathcal{Y}_{(\mathscr{S}_1,\mathscr{S}_2)} \quad \in \mathbb{R}^{j_1\times\ldots\times j_P}$$
$$\mathcal{Y}\bar{\times}_N\mathcal{X} = \mathcal{Y}_{(\mathscr{S}_1,\mathscr{S}_2)}\times_1\mathrm{vec}\left(\mathcal{X}\right) \quad \in \mathbb{R}^{j_1\times\ldots\times j_P}.$$

(iii) let $\mathscr{R} = \{1, \ldots, N\}$ and $\mathscr{C} = \{N+1, \ldots, 2N\}$. If $P = N$ and $I_n = J_n = J_{N+n}$, $n = 1, \ldots, N$, then

$$\mathcal{X}\bar{\times}_N\mathcal{Y}\bar{\times}_N\mathcal{X} = \mathrm{vec}\left(\mathcal{X}\right)'\mathbf{Y}_{(\mathscr{R},\mathscr{C})}\,\mathrm{vec}\left(\mathcal{X}\right).$$

(iv) let $M = N+P$, then $\mathcal{X}\circ\mathcal{Y} = \underline{\mathcal{X}}\bar{\times}_1\underline{\mathcal{Y}}^T$, where $\underline{\mathcal{X}},\underline{\mathcal{Y}}$ are $(I_1\times\ldots\times I_N\times 1)$- and $(J_1\times\ldots\times J_M\times 1)$-dimensional tensors, respectively, given by $\underline{\mathcal{X}}_{:,\ldots,:,1} = \mathcal{X}$, $\underline{\mathcal{Y}}_{:,\ldots,:,1} = \mathcal{Y}$ and $\underline{\mathcal{Y}}^T_{j_1,\ldots,j_M,j_{M+1}} = \underline{\mathcal{Y}}_{j_{M+1},j_M,\ldots,j_1}$.

Proof. Case (i). By definition of contracted product and tensor scalar product

$$\mathcal{X}\bar{\times}_N\mathcal{Y} = \sum_{i_1=1}^{I_1}\ldots\sum_{i_N=1}^{I_N}\mathcal{X}_{i_1,\ldots,i_N}\mathcal{Y}_{i_1,\ldots,i_N} = \langle\mathcal{X},\mathcal{Y}\rangle = \operatorname{vec}(\mathcal{X})'\cdot\operatorname{vec}(\mathcal{Y}).$$

Case (ii). 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

$$\mathcal{X}\bar{\times}_N\mathcal{Y} = \sum_{i_1=1}^{I_1}\ldots\sum_{i_N=1}^{I_N}\mathcal{X}_{i_1,\ldots,i_N}\mathcal{Y}_{i_1,\ldots,i_N,j_{N+1},\ldots,j_{N+P}} = \sum_{k=1}^{I^*}\mathcal{X}_k\mathcal{Y}_{k,j_{N+1},\ldots,j_{N+P}}.$$

Note that the one-to-one correspondence established by the mapping between k and $(i_1,\ldots,i_N)$ corresponds to that of the vectorization of a $(I_1\times\ldots\times I_N)$-dimensional tensor. It also corresponds to the mapping established by the tensor reshaping of a $(N+P)$-order tensor with dimensions $I_1,\ldots,I_N,J_{N+1},\ldots,J_{N+P}$ into a $(P+1)$-order tensor with dimensions $I^*,J_{N+1},\ldots,J_{N+P}$. Let $\mathscr{S}_1 = \{1,\ldots,N\}$, then

$$\mathcal{X}\bar{\times}_N\mathcal{Y} = \sum_{i_1=1}^{I_1}\ldots\sum_{i_N=1}^{I_N}\mathcal{X}_{i_1,\ldots,i_N}\mathcal{Y}_{i_1,\ldots,i_N,:,\ldots,:} = \sum_{s_1=1}^{|\mathscr{S}_1|}\mathbf{x}_{s_1}\bar{\mathcal{Y}}_{s_1,:,\ldots,:}$$

where $\bar{\mathcal{Y}} = \operatorname{reshape}_{(\mathscr{S}_1,N+1,\ldots,N+P)}(\mathcal{Y})$. Following the same approach, and defining $\mathscr{S}_2 = \{N+1,\ldots,N+P\}$, we obtain the second part of the result.
Case (iii). We follow the same strategy adopted in case b). Let $\mathbf{x} = \operatorname{vec}(\mathcal{X})$, $S_1 = \{1,\ldots,N\}$ and $S_2 = \{N+1,\ldots,N+P\}$, such that $(S-1,S_2)$ is a partition of $\{1,\ldots,N+P\}$. Let k,k' be defined as in case b). Then

$$\mathcal{X}\bar{\times}_N\mathcal{Y}\bar{\times}_N\mathcal{X} = \sum_{i_1=1}^{I_1}\ldots\sum_{i_N=1}^{I_N}\sum_{i'_1=1}^{I_1}\ldots\sum_{i'_N=1}^{I_N}\mathcal{X}_{i_1,\ldots,i_N}\mathcal{Y}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N}\mathcal{X}_{i'_1,\ldots,i'_N}$$

$$= \sum_{k=1}^{I^*}\sum_{i'_1=1}^{I_1}\ldots\sum_{i'_N=1}^{I_N}\mathbf{x}_k\mathcal{Y}_{k,i'_1,\ldots,i'_N}\mathcal{X}_{i'_1,\ldots,i'_N} = \sum_{k=1}^{I^*}\sum_{k'=1}^{I^*}\mathbf{x}_k\mathcal{Y}_{k,k'}\mathbf{x}_{k'} = \operatorname{vec}(\mathcal{X})'\mathcal{Y}_{(S_1,S_2)}\operatorname{vec}(\mathcal{X}).$$

Case (iv). Let $\mathbf{i} = (i_1,\ldots,i_N)$ and $\mathbf{j} = (j_1,\ldots,j_M)$ be two multi-indexes. By the definition of outer and contracted product we get $(\mathcal{X}\circ\mathcal{Y})_{\mathbf{i},\mathbf{j}} = \underline{\mathcal{X}}_{\mathbf{i},1}\underline{\mathcal{Y}}_{1,\mathbf{j}} = (\underline{\mathcal{X}}\bar{\times}_1\underline{\mathcal{Y}}^T)_{\mathbf{i},\mathbf{j}}$. Therefore, with a slight abuse of notation, we use $\underline{\mathcal{Y}} = \mathcal{Y}$ and write $\mathcal{Y}\circ\mathcal{Y} = \mathcal{Y}\bar{\times}_1\mathcal{Y}^T$, when the meaning of the products is clear form the context.

Lemma A.2 (Kronecker - matricization). *Let X_n be a $I_n \times I_n$ matrix, for $n = 1, \ldots, N$, and let $\mathcal{X} = X_1 \circ \ldots \circ X_N$ be the $(I_1 \times \ldots \times I_N \times I_1 \times \ldots \times I_N)$-dimensional tensor obtained as the outer product of the matrices $X_1, \ldots, X_N$. Let $(\mathscr{S}_1, \mathscr{S}_2)$ be a partition of $I_{\mathbf{N}} = \{1, \ldots, 2N\}$, where $\mathscr{S}_1 = \{1, \ldots, N\}$ and $\mathscr{S}_2 = \{N+1, \ldots, N\}$. Then $\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)} = \mathbf{X}_{(\mathscr{R},\mathscr{C})} = (X_N \otimes \ldots \otimes X_1)$.*

Proof. Use the pair of indices (i_n, i'_n) for the entries of the matrix X_n, $n = 1, \ldots, N$. By definition of outer product $(X_1 \circ \ldots \circ X_N)_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} = (X_1)_{i_1,i'_1} \cdot \ldots \cdot (X_N)_{i_N,i'_N}$. By definition of matricization, $\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)} = \mathbf{X}_{(\mathscr{R},\mathscr{C})}$. Moreover $(\mathcal{X}_{(\mathscr{S}_1,\mathscr{S}_2)})_{h,k} = \mathcal{X}_{i_1,\ldots,i_{2N}}$ with $h = \sum_{p=1}^{N} (i_{S_{1,p}} - 1) \prod_{q=1}^{p-1} J_{S_{1,p}}$ and $k = \sum_{p=1}^{N} (i_{S_{2,p}} - 1) \prod_{q=1}^{p-1} J_{S_{2,p}}$. By definition of the Kronecker product, the entry (h', k') of $(X_N \otimes \ldots \otimes X_1)$ is $(X_N \otimes \ldots \otimes X_1)_{h',k'} = (X_N)_{i'_N,i'_N} \cdot \ldots \cdot (X_1)_{i_1,i'_1}$, where $h' = \sum_{p=1}^{N} (i_{S_{1,p}} - 1) \prod_{q=1}^{p-1} J_{S_{1,p}}$ and $k' = \sum_{p=1}^{N} (i_{S_{2,p}} - 1) \prod_{q=1}^{p-1} J_{S_{2,p}}$. 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.

Lemma A.3 (Outer product and vectorization). *Let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ be vectors such that $\boldsymbol{\alpha}_i$ has length d_i, for $i = 1, \ldots, n$. Then, for each $j = 1, \ldots, n$, it holds*

$$\operatorname{vec}\Big(\mathop{\circ}_{i=1}^{n} \boldsymbol{\alpha}_i \Big) = \bigotimes_{i=1}^{n} \boldsymbol{\alpha}_{n-i+1} = \big(\boldsymbol{\alpha}_n \otimes \ldots \otimes \boldsymbol{\alpha}_{j+1} \otimes \mathbf{I}_{d_j} \otimes \boldsymbol{\alpha}_{j-1} \otimes \ldots \otimes \boldsymbol{\alpha}_1\big)\boldsymbol{\alpha}_j.$$

Proof. The result follows from the definitions of vectorisation operator and outer product. For $n = 2$, the result follows directly from

$$\operatorname{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2\big) = \operatorname{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}'_2\big) = \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1 = (\boldsymbol{\alpha}_2 \otimes \mathbf{I}_{d_1})\boldsymbol{\alpha}_1 = (\mathbf{I}_{d_2} \otimes \boldsymbol{\alpha}_1)\boldsymbol{\alpha}_2.$$

For $n > 2$ consider, without loss of generality, $n = 3$ (an analogous proof holds for $n > 3$). Then, from the definitions of outer product and Kronecker product we have

$$\begin{aligned}
&\operatorname{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3\big) = \\
&= (\boldsymbol{\alpha}'_1 \cdot \alpha_{2,1}\alpha_{3,1}, \ldots, \boldsymbol{\alpha}'_1 \cdot \alpha_{2,d_2}\alpha_{3,1}, \boldsymbol{\alpha}'_1 \cdot \alpha_{2,1}\alpha_{3,2}, \ldots, \boldsymbol{\alpha}'_1 \cdot \alpha_{2,d_2}\alpha_{3,2}, \ldots, \boldsymbol{\alpha}'_1 \cdot \alpha_{2,d_2}\alpha_{3,d_3})' \\
&= \boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1 = (\boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \mathbf{I}_{d_1})\boldsymbol{\alpha}_1 = (\boldsymbol{\alpha}_3 \otimes \mathbf{I}_{d_2} \otimes \boldsymbol{\alpha}_1)\boldsymbol{\alpha}_2 = (\mathbf{I}_{d_3} \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1)\boldsymbol{\alpha}_3.
\end{aligned}$$

Proof (of Proposition 2.1). Denote with L the lag operator, s.t. $L\mathcal{Y}_t = \mathcal{Y}_{t-1}$, by properties of the contracted product in Lemma A.1, case (iv), we get $(\mathcal{I} - \widetilde{\mathcal{A}}_1 L)\bar{\times}_N \mathcal{Y}_t = \widetilde{\mathcal{A}}_0 + \widetilde{\mathcal{B}}\bar{\times}_M \mathcal{X}_t + \mathcal{E}_t$. We apply to both sides the operator $(\mathcal{I} + \widetilde{\mathcal{A}}_1 L + \widetilde{\mathcal{A}}_1^2 L^2 + \ldots + \widetilde{\mathcal{A}}_1^{t-1} L^{t-1})$, take $t \to \infty$, and get

$$\lim_{t\to\infty} (\mathcal{I} - \widetilde{\mathcal{A}}_1^t L^t)\bar{\times}_N \mathcal{Y}_t = \Big(\sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k L^k \Big) \bar{\times}_N (\widetilde{\mathcal{A}}_0 + \widetilde{\mathcal{B}}\bar{\times}_M \mathcal{X}_t + \mathcal{E}_t).$$

From Behera et al. (2019), if $\rho(\widetilde{\mathcal{A}}_1) < 1$ and $\mathcal{Y}_0$ is finite a.s., then $\lim_{t\to\infty} \widetilde{\mathcal{A}}_1^t \bar{\times}_N \mathcal{Y}_0 = \mathcal{O}$ and the operator $\sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k L^k$ applied to a sequence $\mathcal{Y}_t$ s.t. $|\mathcal{Y}_{\mathbf{i},t}| < c$ a.s. $\forall \mathbf{i}$ converges to the inverse operator $(\mathcal{I} - \widetilde{\mathcal{A}}_1 L)^{-1}$. By the properties of the contracted product we get

$$
\begin{aligned}
\mathcal{Y}_t &= \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N (L^k \widetilde{\mathcal{A}}_0) + \sum_{k=0}^{\infty} (\widetilde{\mathcal{A}}_1^k \bar{\times}_N \widetilde{\mathcal{B}}) \bar{\times}_M (L^k \mathcal{X}_t) + \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N (L^k \mathcal{E}_t) \\
&= (\mathcal{I} - \widetilde{\mathcal{A}}_1 L)^{-1} \bar{\times}_N \widetilde{\mathcal{A}}_0 + \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N \widetilde{\mathcal{B}} \bar{\times}_M \mathcal{X}_{t-k} + \sum_{k=0}^{\infty} \widetilde{\mathcal{A}}_1^k \bar{\times}_N \mathcal{E}_{t-k} \,.
\end{aligned}
$$

From the assumption $\mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{O}, \Sigma_1, \ldots, \Sigma_N)$, we know that $\mathbb{E}(\mathcal{Y}_t) = \mathcal{Y}_0$, which is finite. Consider the auto-covariance at lag $h \geq 1$. From Lemma A.1, we have $\mathbb{E}\big((\mathcal{Y}_t - \mathbb{E}(\mathcal{Y}_t)) \circ (\mathcal{Y}_{t-h} - \mathbb{E}(\mathcal{Y}_{t-h}))\big) = \mathbb{E}\big(\mathcal{Y}_t \circ \mathcal{Y}_{t-h}\big) = \mathbb{E}\big(\mathcal{Y}_t \bar{\times}_1 \mathcal{Y}_{t-h}^T\big)$. Using the infinite moving average representation for $\mathcal{Y}_t$, we get

$$
\begin{aligned}
\mathbb{E}\Big(\mathcal{Y}_t \bar{\times}_1 \mathcal{Y}_{t-h}^T\Big) &= \mathbb{E}\Big(\Big(\sum_{k=0}^{h-1} \mathcal{A}^k \bar{\times}_N \mathcal{E}_{t-k} + \sum_{k=0}^{\infty} \mathcal{A}^{k+h} \bar{\times}_N \mathcal{E}_{t-k-h}\Big) \bar{\times}_1 \Big(\sum_{k=0}^{\infty} \mathcal{A}^k \bar{\times}_N \mathcal{E}_{t-k-h}\Big)^T\Big) \\
&= \mathbb{E}\Big(\Big(\sum_{k=0}^{\infty} \mathcal{A}^{k+h} \bar{\times}_N \mathcal{E}_{t-k-h}\Big) \bar{\times}_1 \Big(\sum_{k=0}^{\infty} \mathcal{E}_{t-k-h}^T \bar{\times}_N (\mathcal{A}^T)^k\Big)\Big),
\end{aligned}
$$

where we used the assumption of independence of $\mathcal{E}_t, \mathcal{E}_{t-h}$, for any $h \geq 0$, and the fact that $(\mathcal{X} \bar{\times}_N \mathcal{Y})^T = (\mathcal{Y}^T \bar{\times}_N \mathcal{X}^T)$. Using $\mathbb{E}(\mathcal{E}_t) = \mathcal{O}$ and linearity of expectation and of the contracted product we get

$$
\begin{aligned}
\mathbb{E}\Big(\mathcal{Y}_t \bar{\times}_1 \mathcal{Y}_{t-h}^T\Big) &= \sum_{k=0}^{\infty} \mathcal{A}^{k+h} \bar{\times}_N \mathbb{E}\Big(\mathcal{E}_{t-k-h} \bar{\times}_1 \mathcal{E}_{t-k-h}^T\Big) \bar{\times}_N (\mathcal{A}^T)^k \\
&= \sum_{k=0}^{\infty} \mathcal{A}^{k+h} \bar{\times}_N \mathbf{\Sigma} \bar{\times}_N (\mathcal{A}^T)^k = \mathcal{A}^h \bar{\times}_N (\mathcal{I} - \mathcal{A} \bar{\times}_N \mathbf{\Sigma} \bar{\times}_N \mathcal{A}^T)^{-1},
\end{aligned}
$$

where $\mathbb{E}(\mathcal{E}_{t-k-h} \bar{\times}_1 \mathcal{E}_{t-k-h}^T) = \mathbb{E}(\mathcal{E}_{t-k-h} \circ \mathcal{E}_{t-k-h}) = \mathbf{\Sigma} = \Sigma_1 \circ \ldots \circ \Sigma_N$. From the assumption $\rho(\mathcal{A}) < 1$ it follows that the above series converges to a finite limit, which is independent from t, thus proving that the process is weakly stationary.

Proof (of Proposition 2.2). From Brazell et al. (2013, Theorem 3.2, Corollary 3.3), we know that $\mathbb{T}$ is a group (called tensor group) and that the matricization operator $\mathrm{mat}_{1:N,1:N}$ is an isomorphism between $\mathbb{T}$ and the linear group of square matrices of size $I^* = \prod_{n=1}^{N} I_n$. Therefore, there exists a one-to-one relationship between the two eigenvalue problems $\mathcal{A} \bar{\times}_N \mathcal{X} = \lambda \mathcal{X}$ and $A\mathbf{x} = \tilde{\lambda}\mathbf{x}$, where $A = \mathrm{mat}_{1:N,1:N}(\mathcal{A})$. In particular, $\lambda = \tilde{\lambda}$ and $\mathbf{x} = \mathrm{vec}\big(\mathcal{X}\big)$. Consequently, $\rho(A) = \rho(\mathcal{A})$ and the result follows for $p = 1$ from the fact that

$\rho(A) < 1$ is a sufficient condition for the VAR(1) stationarity Lütkepohl (2005, Proposition 2.1). Since any VAR(p) and ART(p) processes can be rewritten as VAR(1) and ART(1), respectively, on an augmented state space, the result follows for any $p \geq 1$.

PROOF (OF LEMMA 2.1). Consider a ART(p) process with $\mathcal{Y}_t \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and $p \geq 1$. We define the $(pI_1 \times I_2 \times \ldots \times I_N)$-dimensional tensors $\underline{\mathcal{Y}}_t$ and $\underline{\mathcal{E}}_t$ as $\underline{\mathcal{Y}}_{(k-1)I_1+1:kI_1,:,\ldots,:,t} = \mathcal{Y}_{t-k}$ and $\underline{\mathcal{E}}_{(k-1)I_1+1:kI_1,:,\ldots,:,t} = \mathcal{E}_{t-k}$, for $k = 0, \ldots, p$, respectively. Define the $(pI_1 \times I_2 \times \ldots \times I_N \times pI_1 \times I_2 \times \ldots \times I_N)$-dimensional tensor $\underline{\mathcal{A}}$ as $\underline{\mathcal{A}}_{(1:I_1,:,\ldots,:,(k-1)I_1+1:kI_1,:,\ldots,:} = \mathcal{A}_k$, for $k = 1, \ldots, p$, $\underline{\mathcal{A}}_{(kI_1+1:(k+1)I_1,:,\ldots,:,(k-1)I_1+1:kI_1,:,\ldots,:} = \mathcal{I}$, for $k = 1, \ldots, p-1$ and 0 elsewhere. Using this notation, we can rewrite the $(I_1 \times I_2 \times \ldots \times I_N)$-dimensional ART($p$) process $\mathcal{Y}_t = \sum_{k=1}^{p} \mathcal{A}_k \bar{\times}_N \mathcal{Y}_{t-j} + \mathcal{E}_t$ as the $(pI_1 \times I_2 \times \ldots \times I_N)$-dimensional ART(1) process $\underline{\mathcal{Y}}_t = \underline{\mathcal{A}} \bar{\times}_N \underline{\mathcal{Y}}_{t-1} + \underline{\mathcal{E}}_t$.

B. Computational Details

This appendix shows the derivation of the results. See the supplement for details.

B.1. Full conditional distribution of ϕ_r

Define $C_r = \sum_{j=1}^{J} \boldsymbol{\beta}_j^{(r)'} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)}$ and note that, since $\sum_{r=1}^{R} \phi_r = 1$, it holds $\sum_{r=1}^{R} b_\tau \tau \phi_r = b_\tau \tau$. The posterior full conditional distribution of $\boldsymbol{\phi}$, integrating out τ, is

$$
\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{B}, \mathbf{W}) &\propto \pi(\boldsymbol{\phi}) \int_0^{+\infty} p(\mathcal{B}|\mathbf{W}, \boldsymbol{\phi}, \tau)\pi(\tau)\mathrm{d}\tau \\
&\propto \prod_{r=1}^{R} \phi_r^{\alpha-1} \int_0^{+\infty} \Big(\prod_{r=1}^{R}\prod_{j=1}^{J} (\tau\phi_r)^{-I_j/2} \exp\Big(-\frac{1}{2\tau\phi_r} \boldsymbol{\beta}_j^{(r)'} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)} \Big) \Big) \tau^{a_\tau - 1} e^{-b_\tau \tau} \mathrm{d}\tau \\
&\propto \int_0^{+\infty} \Big(\prod_{r=1}^{R} \phi_r^{\alpha - \frac{I_0}{2} - 1} \Big) \tau^{\left(\alpha R - \frac{R I_0}{2}\right) - 1} \exp\Big(-\sum_{r=1}^{R} \Big(\frac{C_r}{2\tau\phi_r} + b_\tau \tau \phi_r \Big) \Big) \mathrm{d}\tau
\end{aligned}
$$

where the integrand is the kernel of the GiG for $\psi_r = \tau\phi_r$ in eq. (20). Then, by renormalizing, $\phi_r = \psi_r / \sum_{l=1}^{R} \psi_l$.

B.2. Full conditional distribution of τ

The posterior full conditional distribution of τ is

$$
p(\tau|\mathcal{B}, \mathbf{W}, \boldsymbol{\phi}) \propto \tau^{a_\tau - 1} e^{-b_\tau \tau} \Big(\prod_{r=1}^{R} (\tau\phi_r)^{-\frac{I_0}{2}} \exp\Big(-\frac{1}{2\tau\phi_r} \sum_{j=1}^{4} \boldsymbol{\beta}_j^{(r)'} (W_{j,r})^{-1} \boldsymbol{\beta}_j^{(r)} \Big) \Big)
$$

$$\propto \tau^{a_\tau - \frac{RI_0}{2} - 1} \exp\Big(- b_\tau \tau - \tau^{-1} \sum_{r=1}^{R} \frac{C_r}{\phi_r} \Big),$$

which is the kernel of the GiG in eq. (21).

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

The full conditional distribution of $\lambda_{j,r}$, integrating out $W_{j,r}$, is

$$\begin{aligned} p(\lambda_{j,r}|\boldsymbol{\beta}_j^{(r)}, \phi_r, \tau) &\propto \lambda_{j,r}^{a_\lambda - 1} e^{-b_\lambda \lambda_{j,r}} \prod_{p=1}^{I_j} \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\Big(- \frac{|\beta_{j,p}^{(r)}|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}} \Big) \\ &\propto \lambda_{j,r}^{(a_\lambda + I_j) - 1} \exp\Big(- \Big(b_\lambda + \frac{\|\boldsymbol{\beta}_j^{(r)}\|_1}{\sqrt{\tau\phi_r}} \Big) \lambda_{j,r} \Big), \end{aligned}$$

which is the kernel of the Gamma in eq. (23).

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

The posterior full conditional distribution of $w_{j,r,p}$ is

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

which is the kernel of the GiG in eq. (23).

B.5. Full conditional distributions of PARAFAC marginals

Consider the model in eq. (15), it holds

$$\mathrm{vec}\left(\mathcal{Y}_t\right) = \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right) + \mathrm{vec}\left(\mathcal{B}_r \times_4 \mathbf{x}_t\right) + \mathrm{vec}\left(\mathcal{E}_t\right),$$

with $\mathrm{vec}\left(\mathcal{B}_r \times_4 \mathbf{x}_t\right) = \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)}$. From Lemma A.3, we have

$$\mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \mathbf{b}_4 \boldsymbol{\beta}_4^{(r)} \tag{B.1}$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_I\right) \boldsymbol{\beta}_1^{(r)} = \mathbf{b}_1 \boldsymbol{\beta}_1^{(r)} \tag{B.2}$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\beta}_1^{(r)}\right) \boldsymbol{\beta}_2^{(r)} = \mathbf{b}_2 \boldsymbol{\beta}_2^{(r)} \tag{B.3}$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_K \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right) \boldsymbol{\beta}_3^{(r)} = \mathbf{b}_3 \boldsymbol{\beta}_3^{(r)}. \tag{B.4}$$

Define with $\mathbf{y}_t = \mathrm{vec}\left(\mathcal{Y}_t\right)$ and $\boldsymbol{\Sigma}^{-1} = \Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}$, we obtain

$$L(\mathbf{Y}|\boldsymbol{\theta}) \propto \exp\Big(-\frac{1}{2}\sum_{t=1}^{T} \mathrm{vec}\left(\tilde{\mathcal{E}}_t\right)'\left(\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}\right) \mathrm{vec}\left(\tilde{\mathcal{E}}_t\right)\Big)$$

$$\propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right)'\big)\boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle$$

$$+ \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)'\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\Bigg). \tag{B.5}$$

Consider the case $j = 1$. By exploiting eq. (B.2) we get

$$L(\mathbf{Y}|\boldsymbol{\theta}) \propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)\prime}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle^2\big(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\big)'\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\big)$$

$$\cdot \boldsymbol{\beta}_1^{(r)} - 2\big(\mathbf{y}_t' - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right)'\big)\boldsymbol{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\big)\boldsymbol{\beta}_1^{(r)}\Bigg)$$

$$= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_1^{(r)\prime}\mathbf{S}_1^L\boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L\boldsymbol{\beta}_1^{(r)}\Big). \tag{B.6}$$

Consider the case $j = 2$. From eq. (B.3) we get

$$L(\mathbf{Y}|\boldsymbol{\theta}) \propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_2^{(r)\prime}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle^2\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)$$

$$\cdot \boldsymbol{\beta}_2^{(r)} - 2\big(\mathbf{y}_t' - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right)'\big)\boldsymbol{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)}\Bigg)$$

$$= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}\mathbf{S}_2^L\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L\boldsymbol{\beta}_2^{(r)}\Big). \tag{B.7}$$

Consider the case $j = 3$, by exploiting eq. (B.4) we get

$$L(\mathbf{Y}|\boldsymbol{\theta}) \propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)\prime}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle^2\big(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\Sigma}^{-1}\big(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\big)$$

$$\cdot \boldsymbol{\beta}_3^{(r)} - 2\big(\mathbf{y}_t' - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right)'\big)\boldsymbol{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)}\Bigg)$$

$$= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_3^{(r)\prime}\mathbf{S}_3^L\boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L\boldsymbol{\beta}_3^{(r)}\Big). \tag{B.8}$$

Finally, in the case $j = 4$. From eq. (B.5) we get

$$L(\mathbf{Y}|\boldsymbol{\theta}) \propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \mathrm{vec}\left(\mathcal{B}_{-r} \times_4 \mathbf{x}_t\right)'\big)\boldsymbol{\Sigma}^{-1} \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)$$

$$\cdot \mathbf{x}_t'\boldsymbol{\beta}_4^{(r)} + \boldsymbol{\beta}_4^{(r)\prime}\mathbf{x}_t \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)'\boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)}\Bigg)$$

$$= \exp\left(-\frac{1}{2}\boldsymbol{\beta}_4^{(r)\prime}\mathbf{S}_4^L\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)}\right). \tag{B.9}$$

637 *B.5.1. Full conditional distribution of $\beta_1^{(r)}$*

From eq. (17)-(B.6), the posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$ is

$$p(\boldsymbol{\beta}_1^{(r)}|-) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}_1^{(r)\prime}\mathbf{S}_1^L\boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L\boldsymbol{\beta}_1^{(r)}\right) \cdot \exp\left(-\frac{1}{2}\boldsymbol{\beta}_1^{(r)\prime}(W_{1,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_1^{(r)}\right)$$
$$= \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_1^{(r)\prime}\left(\mathbf{S}_1^L + (W_{1,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L\boldsymbol{\beta}_1^{(r)}\right)\right),$$

638 which is the kernel of the Normal in eq. (24).

639 *B.5.2. Full conditional distribution of $\beta_2^{(r)}$*

From eq. (17)-(B.7), the posterior full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$ is

$$p(\boldsymbol{\beta}_2^{(r)}|-) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}\mathbf{S}_2^L\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L\boldsymbol{\beta}_2^{(r)}\right) \cdot \exp\left(-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}(W_{2,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_2^{(r)}\right)$$
$$= \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_2^{(r)\prime}\left(\mathbf{S}_2^L + (W_{2,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L\boldsymbol{\beta}_2^{(r)}\right)\right),$$

640 which is the kernel of the Normal in eq. (24).

641 *B.5.3. Full conditional distribution of $\beta_3^{(r)}$*

From eq. (17)-(B.8), the posterior full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$ is

$$p(\boldsymbol{\beta}_3^{(r)}|-) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}_3^{(r)\prime}\mathbf{S}_3^L\boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L\boldsymbol{\beta}_3^{(r)}\right) \cdot \exp\left(-\frac{1}{2}\boldsymbol{\beta}_3^{(r)\prime}(W_{3,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_3^{(r)}\right)$$
$$= \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_3^{(r)\prime}\left(\mathbf{S}_3^L + (W_{3,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L\boldsymbol{\beta}_3^{(r)}\right)\right),$$

642 which is the kernel of the Normal in eq. (24).

643 *B.5.4. Full conditional distribution of $\beta_4^{(r)}$*

From eq. (17)-(B.9), the posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$ is

$$p(\boldsymbol{\beta}_4^{(r)}|-) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}_4^{(r)\prime}\mathbf{S}_4^L\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)}\right) \cdot \exp\left(-\frac{1}{2}\boldsymbol{\beta}_4^{(r)\prime}(W_{4,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_4^{(r)}\right)$$
$$= \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_4^{(r)\prime}\left(\mathbf{S}_4^L + (W_{4,r}\phi_r\tau)^{-1}\right)\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)}\right)\right),$$

644 which is the kernel of the Normal in eq. (24).

B.6. Full conditional distribution of Σ_1

Define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t} = \mathrm{mat}_{(3)}(\tilde{\mathcal{E}}_t)$, $\mathbf{Z}_1 = \Sigma_3^{-1} \otimes \Sigma_2^{-1}$ and $S_1 = \sum_{t=1}^T \tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_1 \tilde{\mathbf{E}}'_{(1),t}$. The posterior full conditional distribution of Σ_1 is

$$\begin{aligned} p(\Sigma_1|-) &\propto \frac{\exp\Big(-\frac{1}{2}\big(\mathrm{tr}\left(\gamma\Psi_1\Sigma_1^{-1}\right) + \sum_{t=1}^T \mathrm{tr}\left(\tilde{\mathbf{E}}_{(1),t}\mathbf{Z}_1\tilde{\mathbf{E}}'_{(1),t}\Sigma_1^{-1}\right)\big)\Big)}{\left|\Sigma_1\right|^{\frac{\nu_1+I_1+TI_2I_3+1}{2}}} \\ &\propto \left|\Sigma_1\right|^{-\frac{(\nu_1+TI_2I_3)+I_1+1}{2}} \exp\Big(-\frac{1}{2}\,\mathrm{tr}\left((\gamma\Psi_1 + S_1)\Sigma_1^{-1}\right)\Big), \end{aligned}$$

which is the kernel of the Inverse Wishart in eq. (25).

B.7. Full conditional distribution of Σ_2

Define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t$, $\tilde{\mathbf{E}}_{(2),t} = \mathrm{mat}_{(2)}(\tilde{\mathcal{E}}_t)$ and $S_2 = \sum_{t=1}^T \tilde{\mathbf{E}}_{(2),t}(\Sigma_3^{-1} \otimes \Sigma_1^{-1})\tilde{\mathbf{E}}'_{(2),t}$. The posterior full conditional distribution of Σ_2 is

$$\begin{aligned} p(\Sigma_2|-) &\propto \frac{\exp\Big(-\frac{1}{2}\big(\mathrm{tr}\left(\gamma\Psi_2\Sigma_2^{-1}\right) + \mathrm{tr}\left(\sum_{t=1}^T \tilde{\mathbf{E}}_{(2),t}(\Sigma_3^{-1}\otimes\Sigma_1^{-1})\tilde{\mathbf{E}}'_{(2),t}\Sigma_2^{-1}\right)\big)\Big)}{\left|\Sigma_2\right|^{\frac{\nu_2+I_2+TI_1I_3+1}{2}}} \\ &\propto \left|\Sigma_2\right|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\Big(-\frac{1}{2}\,\mathrm{tr}\left(\gamma\Psi_2\Sigma_2^{-1} + S_2\Sigma_2^{-1}\right)\Big), \end{aligned}$$

which is the kernel of the Inverse Wishart in eq. (25).

B.8. Full conditional distribution of Σ_3

Define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \times_4 \mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t} = \mathrm{mat}_{(1)}(\tilde{\mathcal{E}}_t)$, $\mathbf{Z}_3 = \Sigma_2^{-1} \otimes \Sigma_1^{-1}$ and $S_3 = \sum_{t=1}^T \tilde{\mathbf{E}}_{(1),t} \mathbf{Z}_3 \tilde{\mathbf{E}}'_{(1),t}$. The posterior full conditional distribution of Σ_3 is

$$\begin{aligned} p(\Sigma_3|-) &\propto \frac{\exp\Big(-\frac{1}{2}\big(\mathrm{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^T \mathrm{vec}\left(\tilde{\mathcal{E}}_t\right)'(\Sigma_3^{-1}\otimes\mathbf{Z}_3)\,\mathrm{vec}\left(\tilde{\mathcal{E}}_t\right)\big)\Big)}{\left|\Sigma_3\right|^{\frac{\nu_3+I_3+TI_1I_2+1}{2}}} \\ &\propto \left|\Sigma_3\right|^{-\frac{(\nu_3+TI_1I_2)+I_3+1}{2}} \exp\Big(-\frac{1}{2}\,\mathrm{tr}\left((\gamma\Psi_3 + S_3)\Sigma_3^{-1}\right)\Big), \end{aligned}$$

which is the kernel of the Inverse Wishart in eq. (25).

B.9. Full conditional distribution of γ

The posterior full conditional distribution is

$$p(\gamma|\Sigma_1,\Sigma_2,\Sigma_3) \propto \prod_{i=1}^3 \left|\gamma\Psi_i\right|^{-\frac{\nu_i}{2}} \exp\Big(-\frac{1}{2}\,\mathrm{tr}\left(\gamma\Psi_i\Sigma_i^{-1}\right)\Big)\gamma^{a_\gamma-1}e^{-b_\gamma\gamma}$$

$$\propto \gamma^{a_\gamma - \frac{\sum_{i=1}^{3} \nu_i I_i}{2} - 1} \exp\Big(-\frac{1}{2} \operatorname{tr}\big(\sum_{i=1}^{3} \Psi_i \Sigma_i^{-1} \big) - b_\gamma \gamma \Big)$$

which is the kernel of the Gamma in eq. (26).