Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Bayesian Dynamic Tensor Regression

Monica Billio, Roberto Casarin, Matteo Iacopini & Sylvia Kaufmann

To cite this article: Monica Billio, Roberto Casarin, Matteo Iacopini & Sylvia Kaufmann (2023) Bayesian Dynamic Tensor Regression, Journal of Business & Economic Statistics, 41:2, 429-439, DOI: 10.1080/07350015.2022.2032721

To link to this article: https://doi.org/10.1080/07350015.2022.2032721

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 2, 429–439
https://doi.org/10.1080/07350015.2022.2032721

Bayesian Dynamic Tensor Regression

Monica Billio[a], Roberto Casarin[a], Matteo Iacopini[b,c], and Sylvia Kaufmann[d]

[a]Department of Economics, Ca' Foscari University of Venice, Venezia, Italy; [b]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [c]Tinbergen Institute, Amsterdam, Netherlands; [d]Study Center Gerzensee, Foundation of the Swiss National Bank, Gerzensee, 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 parameterization 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.

ARTICLE HISTORY
Received September 2020
Accepted January 2022

KEYWORDS
Bayesian inference; Dynamic networks; Forecasting; Multidimensional autoregression; Tensor models

1. Introduction

Many modern datasets in applied science have a complex and multidimensional structure which is naturally represented by multidimensional arrays, or tensors (e.g., Hackbusch 2012). In statistics and machine learning, tensor algebra provides a fundamental background for effective modeling and efficient algorithm design in big data handling (e.g., Cichocki 2014). The increasing availability of long temporal sequences of tensor-valued data, such as multidimensional tables (Balazsi, Matyas, and Wansbeek 2015), multidimensional panel data (Kapetanios, Serlenga, and Shin 2021), multilayer networks (Aldasoro and Alves 2018), electroencephalogram (a.k.a. EEG, Li and Zhang 2017), neuroimaging (Zhou, Li, and Zhu 2013) has put forward some limitations of the existing multivariate time series models. A naïve approach to model tensors ignores the intrinsic structure of the data and fits a multivariate regression on the vectorized tensor data. However, this might result in inefficient estimation and misleading results (Yuan and Zhang 2016), thus, making such representations unsuited for tensor-valued data.

Tensor modeling in statistics is in its infancy and most of the research in this field has focused on the analysis of cross-sectional data, as applied in neuroimaging (e.g., functional magnetic resonance image, a.k.a. fMRI, EEG) and signal processing, whereas the literature on tensors in time series analysis is scarce. Most often, a tensor-valued covariate is used to predict a scalar outcome (e.g., see Zhou, Li, and Zhu 2013; Xu et al. 2013; Guhaniyogi, Qamar, and Dunson 2017), and only a few articles analyze tensor-on-tensor regression models (e.g., see Lock 2018). Estimation of tensor regressions requires parameter regularization or dimension reduction since the number of entries of the coefficient tensor is larger than the sample size.

In contrast to the existing literature, this article introduces dynamics in tensor regression models by defining a new framework for linear time series regression with tensor-valued response and covariates. We study the properties of the stochastic process, such as stationarity, and derive impulse response functions. Standard multivariate regression models are obtained as special cases. To address the dimensionality challenges of dynamic tensor models, we propose a low-rank representation of the coefficient tensor and impose parameter regularization based on the shrinkage prior distribution of Guhaniyogi, Qamar, and Dunson (2017).

Guhaniyogi, Qamar, and Dunson (2017) design a predictive model in a cross-sectional setting to investigate the relationship between a scalar medical index and matrix-valued brain images. Instead, we propose a new framework for dynamic tensor-on-tensor regression, and use it to investigate multilayer international economic networks.

Recent articles on tensor regression exploit tensor-valued covariates to predict a scalar outcome in a generalized linear model (Zhou, Li, and Zhu 2013; Xu et al. 2013), whereas Li et al. (2018) use the Tucker decomposition to propose low-rank approximations to the coefficient tensor. On the other hand, the tensor-on-vector regression is an alternative approach used to assess the impact of a vector of factors on a tensor-valued observable. Rabusseau and Kadri (2016) consider a higher-order low-rank regression, which is a tensor-on-vector linear model with a low-rank constraint on the coefficient tensor. They

CONTACT Matteo Iacopini *m.iacopini@vu.nl* Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/UBES*.

propose an algorithm to obtain an approximate solution to the restricted least-squares problem. In a related contribution, Guha and Rodriguez (2020) develop a Bayesian linear model for assessing the impact of vector covariates on matrix-valued MRIs for several patients. They adopt a symmetric parallel factor (PARAFAC) decomposition to identify the tensor nodes and cells related to each predictor. To study the impact of one or more external stimuli or predictors on the human brain, Guhaniyogi and Spencer (2021) have developed a regression framework with a tensor response and scalar covariates, coupled with a novel multiway stick-breaking shrinkage prior distribution on the coefficient tensor. The method has been extended by Spencer, Guhaniyogi, and Prado (2020) to an additive mixed regression model with a tensor response, with region-specific random effects to capture the connectivity between the measurements on a set of prespecified groups of brain voxels. In the presence of structured tensor-response variables, such as maps of neural connections in the brain, Guha and Guhaniyogi (2021) have proposed a Bayesian generalized linear model with a symmetric tensor response and scalar predictors. A brief review of the most recent contributions on tensor regression models is presented in Guhaniyogi (2020).

Another stream of the literature considers regression models with tensor-valued responses and covariates. Hoff (2015) employs the Tucker product to define a tensor-on-tensor regression, generalizing the standard bilinear to a multilinear model. Tensors have also been used in the analysis of large multivariate categorical response vectors (Zhou et al. 2015) and in high-dimensional classification problems (Yang and Dunson 2016). Extending all these approaches, we consider a novel linear autoregressive model for real-valued tensor response and covariates, and we apply it in a time series framework to investigate dynamic multilayer networks.

We exploit the contracted product, an operator that generalizes the Cayley matrix multiplication to tensors (Ji and Wei 2018; Behera, Nandi, and Sahoo 2020; Wang, Du, and Ma 2020), to introduce a new autoregressive tensor model (ART) which generalizes the existing tensor regression frameworks along two lines. First, the ART model introduces dynamics in linear tensor regression and provides the tools for analyzing shock propagation in multidimensional dynamical systems. Second, we allow for both tensor-valued outcomes and covariates, a more general framework encompassing existing tensor as well as multivariate linear models (e.g., vector autoregressions, or VARs). Taking advantage of the properties of the contracted product, we derive new results on tensor algebra and study the main properties of the ART process. Besides, we derive the impulse response function and the forecast error variance decomposition for making predictions and analyzing shock propagation in the system.

Besides handling multidimensional data, tensor regression models are usually characterized by a high dimensional parameter space, which calls for the use of dimension reduction or shrinkage estimation techniques. Li and Zhang (2017) define a tensor-response linear regression on a vector covariate for studying the relationship between brain activity and individual control variables, using cross-sectional data. They use the envelope method for estimation, which assumes that part of the response variables (a set of linear combinations of them) is irrelevant to the regression. Moreover, their optimization framework depends on tuning parameters (e.g., the envelope dimensions), the choice of which depends on the tensor dimensions and the signal-to-noise ratio (i.e., the degree of sparsity). Here, we propose to use a PARAFAC representation (Hackbusch 2012) of the coefficient tensor to obtain a parsimonious parameterization of the ART.

Parameter regularization and sparse estimation in high-dimensional models can be achieved through alternative approaches, such as the Lasso (Zhou, Li, and Zhu 2013), the spike-and-slab (Guha and Rodriguez 2020), and the envelope method (Li and Zhang 2017). Alternative approaches induce element-wise sparsity or assume reduced-rank coefficient tensors. In neuroimaging, Sun and Li (2017) propose a regression framework for a tensor response and a vector predictor, where the coefficient tensor embeds both types of sparse structures. Raskutti, Yuan, and Chen (2019) derive general risk bounds of the estimated coefficient in high-dimensional tensor regression problems with several regularizers, such as Lasso penalty and reduced-rank. Goldsmith, Huang, and Crainiceanu (2014) develop scalar-on-3D-image regression that includes a latent binary indicator to discriminate between image locations with predictive and nonpredictive power. Here, we adopt the more flexible regularization approach based on the global-local shrinkage prior developed in Guhaniyogi, Qamar, and Dunson (2017). In particular, we impose this prior on the marginal vectors of the PARAFAC representation of the coefficient tensor and we show that, for rank-1 coefficient tensor, the conditional prior on the entries is a Meijer-G prior with heavier tails than the Normal distribution (see, e.g., Zhang et al. 2020).

The literature on network data modeling has rapidly increased after the recent financial crisis, both in theoretical and empirical analyses. Dynamic tensor models are a natural framework for the analysis of multilayer network data in finance, biology, and sociology. An example of a time series of network data consists of a collection of yearly snapshots of interbank or international trade networks. However, despite dynamic models may be more adequate for studying network data collected over time, most statistical models for network data remained static so far (De Paula 2017). Few attempts have been made to model time-varying networks (e.g., Hoff 2015; Anacleto and Queen 2017), and most of the existing approaches focus on providing a representation and a description of temporally evolving graphs (e.g., Kostakos 2009; Holme and Saramäki 2012). We contribute to this literature by providing an original study of time-varying economic and financial networks and show that our dynamic tensor model can be used successfully to carry out impulse response analysis in a multidimensional setting.

The remainder of the article is organized as follows. Section 2 provides an introduction to tensor algebra and presents the new modeling framework. Section 3 discusses parameterization 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 materials.

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 article, and refer the reader to the supplementary materials 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., Aris 2012; Abraham, Marsden, and Ratiu 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 $\mathcal{X} = (\mathcal{X}_{i_1,\dots,i_N}) \in \mathbb{R}^{I_1\times\cdots\times I_N}$ with entries $\mathcal{X}_{i_1,\dots,i_N}$ with $i_n = 1,\dots,I_n$ and $n = 1,\dots,N$. 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 article 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 kth index of the tensor, that is, 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 $\mathbb{R}^{I_1\times\dots\times I_N}$ endowed with the standard addition $\mathcal{A}+\mathcal{B} = (\mathcal{A}_{i_1,\dots,i_N} + \mathcal{B}_{i_1,\dots,i_N})$ and scalar multiplication $\alpha\mathcal{A} = (\alpha\mathcal{A}_{i_1,\dots,i_N})$, 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 $\mathcal{X}\in\mathbb{R}^{I_1\times\cdots\times I_M}$ and $\mathcal{Y}\in\mathbb{R}^{J_1\times\cdots\times J_N}$ with $I_M = J_1$, is denoted by $\mathcal{X}\times_M\mathcal{Y}$ and yields a $(M+N-2)$-order tensor $\mathcal{Z}\in\mathbb{R}^{I_1\times\cdots\times I_{M-1}\times J_1\times\cdots\times J_{N-1}}$, with entries

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

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 $\mathcal{X}\in\mathbb{R}^{J_1\times\cdots\times J_K\times I_1\times\cdots\times I_N}$ and the $(N+M)$-order tensor $\mathcal{Y}\in\mathbb{R}^{I_1\times\cdots\times I_N\times H_1\times\cdots\times H_M}$. Entry-wise, it is defined as

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

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

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 article, 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) $\boldsymbol{\beta}_j^{(r)}\in\mathbb{R}^{I_j}$

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

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 rearranging 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 $(i_1,\dots,i_N)$ 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 Section S.1 in the supplementary materials. 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 $I_1\times I_2I_3$ 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 $\mathrm{vec}(\mathcal{X}) = \mathrm{vec}(\mathbf{X}_{(1)})$, where $\mathrm{vec}(\mathbf{X}_{(1)})$ stacks vertically into a vector the columns of the matrix $\mathbf{X}_{(1)}$. Many product operations have been defined for tensors (e.g., 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 $\mathcal{B}\in\mathbb{R}^{I_1\times\cdots\times I_N}$ with a PARAFAC(R) decomposition (with marginals $\boldsymbol{\beta}_j^{(r)}$), a $(N-1)$-order tensor $\mathcal{Y}\in\mathbb{R}^{I_1\times\cdots\times I_{N-1}}$ and a vector $\mathbf{x}\in\mathbb{R}^{I_N}$. Then

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

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

2.2. A General Dynamic Tensor Model

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

$$\mathcal{Y}_{\mathbf{i},t} = \mathcal{A}_{\mathbf{i},0} + \sum_{j=1}^{p}\sum_{\mathbf{k}\in S_y}\mathcal{A}_{\mathbf{i},\mathbf{k},j}\mathcal{Y}_{\mathbf{k},t-j} + \sum_{\mathbf{m}\in S_x}\mathcal{B}_{\mathbf{i},\mathbf{m}}\mathcal{X}_{\mathbf{m},t} + \mathcal{E}_{\mathbf{i},t},$$

$$\mathcal{E}_{\mathbf{i},t}\overset{iid}{\sim}\mathcal{N}(0,\sigma^2_{\mathbf{i}}), \tag{2}$$

$t \in \mathbb{Z}$, with given initial conditions $\mathcal{Y}_{-p+1}, \ldots, \mathcal{Y}_0 \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, 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 Equation (2) allows for measuring the effect of all the cells of the exogenous variable $\mathcal{X}_t$ and of the lagged values of $\mathcal{Y}_t$ on each endogenous variable. When exogenous variables are not included, the model is denoted by ART(p).

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

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

$t \in \mathbb{Z}$, 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 $(I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N)$ and $\mathcal{B}$ is a $(N+M)$-order tensor of size $(I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M)$. The error term $\mathcal{E}_t$ follows a N-order tensor normal distribution (Ohlson, Ahmad, and Von Rosen 2013) with probability density function

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

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

$$\mathcal{Y}_t = \mathcal{A}_0 + \textstyle\sum_{j=1}^{p} \mathcal{A}_j \times_{N+1} \text{vec}(\mathcal{Y}_{t-j}) + \mathcal{B} \times_{N+1} \text{vec}(\mathcal{X}_t) + \mathcal{E}_t,$$
$$\mathcal{E}_t \overset{\text{iid}}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{O}, \Sigma_1, \ldots, \Sigma_N), \qquad t \in \mathbb{Z}. \tag{5}$$

Let $\mathbb{T} = (\mathbb{R}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}, \bar{\times}_N)$ be the space of $(I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N)$-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 $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ are called eigenvalue and eigentensor 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 results (see Section S.4 in the supplementary materials for the proofs).

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

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

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

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

2.3. Parameterization

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

Remark 2.2 (Parameterization 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

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

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 $R(I_1 + \cdots + I_N +$

$\prod_{j=1}^{N} I_j$), while it is $2R(I_1 + \cdots + I_N)$ 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 $(I_1 \times \cdots \times I_N)$-dimensional ARTX($p$) and a $(I_1 \cdot \ldots \cdot I_N)$-dimensional VARX($p$) model. The vector form of (5) is

$$\begin{aligned}\operatorname{vec}(\mathcal{Y}_t) &= \operatorname{vec}(\mathcal{A}_0) + \sum_{j=1}^{p} \operatorname{mat}_{N+1}(\mathcal{A}_j)\operatorname{vec}(\mathcal{Y}_{t-j}) \\ &\quad + \operatorname{mat}_{N+1}(\mathcal{B})\operatorname{vec}(\mathcal{X}_t) + \operatorname{vec}(\mathcal{E}_t) \\ \mathbf{y}_t &= \boldsymbol{\alpha}_0 + \sum_{j=1}^{p} \mathbf{A}'_{(N+1),j}\mathbf{y}_{t-j} + \mathbf{B}'_{(N+1)}\mathbf{x}_t + \boldsymbol{\epsilon}_t, \\ \boldsymbol{\epsilon}_t &\sim \mathcal{N}_{I^*}(\mathbf{0}, \Sigma_N \otimes \ldots \otimes \Sigma_1), \end{aligned} \tag{6}$$

$t \in \mathbb{Z}$, 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, Ahmad, and Von Rosen 2013) given by $\mathcal{X} \sim \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_N)$ if and only if $\operatorname{vec}(\mathcal{X}) \sim \mathcal{N}_{I^*}(\operatorname{vec}(\mathcal{M}), \Sigma_N \otimes \ldots \otimes \Sigma_1)$. The restriction on the covariance structure for the vectorized tensor provides a parsimonious parameterization of the multivariate normal distribution, while allowing both within and between mode dependence. Alternative parameterizations 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 Equation (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 Equation (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 Equation (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 Figure 1.

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

1. *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 $\prod_{j=1}^{J} \lambda_{jr} = 1$;
2. *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 $\{1, \ldots, R\}$;
3. *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 parameterization.

2.4. Important Special Cases

The model in Equation (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 of the supplementary materials for further details. Dropping the covariates $\mathcal{X}_t$ from Equation (5), we obtain an autoregressive tensor model of order p (or ART(p))

$$\begin{aligned}\mathcal{Y}_t &= \mathcal{A}_0 + \sum_{j=1}^{p} \mathcal{A}_j \times_{N+1} \operatorname{vec}(\mathcal{Y}_{t-j}) + \mathcal{E}_t, \\ \mathcal{E}_t &\overset{\text{iid}}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathbf{0}, \Sigma_1, \ldots, \Sigma_N), \qquad t \in \mathbb{Z}. \end{aligned} \tag{7}$$

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 generalized IRF, exploiting the relationship between ART and VAR models. Without loss of generality, we focus on the ART(p) model in Equation (7), with $p = 1$ and $\mathcal{A}_0 = \mathbf{0}$, and introduce the following notation. Let $\mathbf{y}_t = \operatorname{vec}(\mathcal{Y}_t)$ and $\boldsymbol{\epsilon}_t = \operatorname{vec}(\mathcal{E}_t) \sim \mathcal{N}_{I^*}(\mathbf{0}, \Sigma)$ be the $(I^* \times 1)$ tensor response and noise term in vector form, respectively, where $\Sigma = \Sigma_N \otimes \cdots \otimes \Sigma_1$ 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

$$\Sigma = \begin{pmatrix} A & B \\ B' & C \end{pmatrix}, \tag{8}$$

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

$$\begin{aligned}\Sigma = &\begin{pmatrix} \mathbf{I}_n & \mathrm{O}_{n,I^*-n} \\ B'A^{-1} & \mathbf{I}_{I^*-n} \end{pmatrix}\begin{pmatrix} A & \mathrm{O}_{n,I^*-n} \\ \mathrm{O}'_{n,I^*-n} & S \end{pmatrix} \\ &\begin{pmatrix} \mathbf{I}_n & A^{-1}B \\ \mathrm{O}'_{n,I^*-n} & \mathbf{I}_{I^*-n} \end{pmatrix} = LDL', \end{aligned}$$

where $\mathbf{I}_j$ is the identity matrix of size j. Hence Σ can be block-diagonalized

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

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

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

where L_A, L_S are the Cholesky factors of A and S, respectively. Assume the vectorized ART process admits an infinite MA

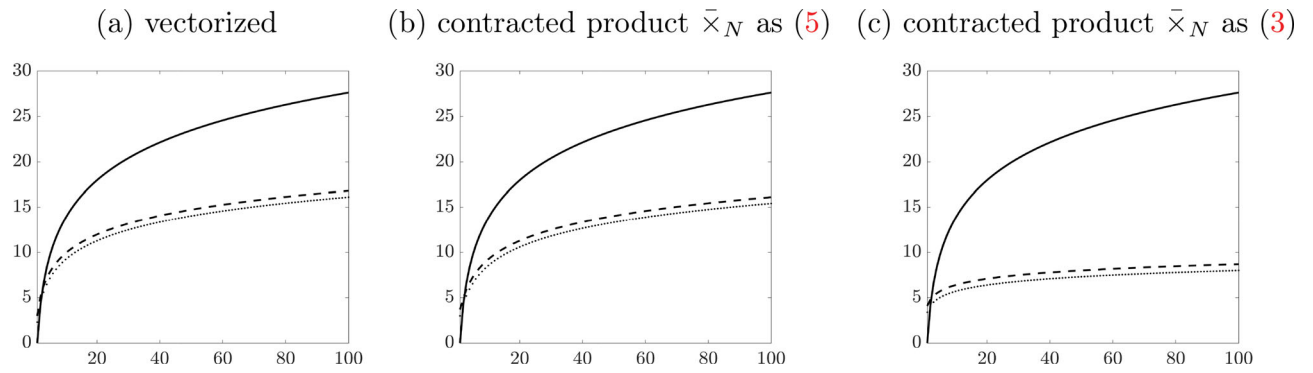

Figure 1. Number of parameters in $\mathcal{A}_0$, in log-scale (*vertical axis*) as function of the size d of the $(d \times d \times d)$-dimensional tensor $\mathcal{Y}_t$ (*horizontal axis*) in a ART(1) model. In all plots: unconstrained model (*solid* line), PARAFAC(R) parameterization with $R =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

representation, with $\Psi_0 = \mathbf{I}_{I^*}$ and $\Psi_i = \text{mat}_{(4)}(\mathcal{B})'\Psi_{i-1}$, then using the previous results we get:

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

where $\boldsymbol{\eta}_t = L^{-1}\boldsymbol{\epsilon}_t$ are the block-orthogonalized shocks and D is the block-diagonal matrix in Equation (9). Denote with E_n the $I^* \times n$ matrix that selects n columns from a premultiplied matrix, that is, 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 generalized IRF of Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998) by defining the block generalized IRF

$$\begin{aligned}\boldsymbol{\psi}^G(h;n) &= \mathbb{E}\big(\text{vec}\big(\mathcal{Y}_{t+h}\big)|\text{vec}\big(\mathcal{E}_t\big)' \\ &= (\boldsymbol{\delta}^{*\prime}, \mathbf{0}'_{I^*-n}), \mathcal{F}_{t-1}\big) - \mathbb{E}\big(\text{vec}\big(\mathcal{Y}_{t+h}\big)|\mathcal{F}_{t-1}\big) \\ &= (\Psi_h L)DE_n A^{-1}\boldsymbol{\delta}^*, \qquad h = 1, 2, \ldots \end{aligned} \tag{11}$$

where $\mathcal{F}_u$, $u \leq t$ is the natural filtration associated to the stochastic process $\mathcal{Y}_t$, $t \in \mathbb{Z}$. Starting from Equation (10) we derive the block Cholesky IRF (OIRF) as

$$\begin{aligned}\boldsymbol{\psi}^O(h;n) &= \mathbb{E}\big(\text{vec}\big(\mathcal{Y}_{t+h}\big)|\text{vec}\big(\mathcal{E}_t\big)' = (\boldsymbol{\delta}^{*\prime}, \mathbf{0}'_{I^*-n}), \mathcal{F}_{t-1}\big) \\ &\quad - \mathbb{E}\big(\text{vec}\big(\mathcal{Y}_{t+h}\big)|\text{vec}\big(\mathcal{E}_t\big)' = \mathbf{0}'_{I^*}, \mathcal{F}_{t-1}\big) \\ &= (\Psi_h L)PE_n\boldsymbol{\delta}^*, \qquad h = 1, 2, \ldots . \end{aligned} \tag{12}$$

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

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

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

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

$$\boldsymbol{\psi}_j^G(h;n) = \Psi_h LDD_{jj}^{-1}\boldsymbol{\delta}_j, \quad \boldsymbol{\psi}_j^O(h;n) = \Psi_h LP\boldsymbol{\delta}_j$$

$$\psi_{ij}^G(h;n) = \mathbf{e}_i'\Psi_h LDD_{jj}^{-1}\boldsymbol{\delta}_j, \quad \psi_{ij}^O(h;n) = \mathbf{e}_i'\Psi_h LP\boldsymbol{\delta}_j.$$

3. Bayesian Inference

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

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

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 Equation (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 vectorizing (15), we get the equivalent VAR

$$\begin{aligned}\text{vec}\big(\mathcal{Y}_t\big) &= \mathbf{B}'_{(4)}\text{vec}\big(\mathcal{Y}_{t-1}\big) + \text{vec}\big(\mathcal{E}_t\big), \\ \text{vec}\big(\mathcal{E}_t\big) &\overset{\text{iid}}{\sim} \mathcal{N}_{I^*}(\mathbf{0}, \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1), \end{aligned} \tag{16}$$

whose covariance has a Kronecker 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 heteroscedasticity.

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, Ahmad, and Von Rosen 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 toward 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 postestimation 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, Qamar, and Dunson (2017). For each $\boldsymbol{\beta}_j^{(r)}$, we define a prior distributions as a scale mixture of normals centered 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 rth component of the PARAFAC, and the local parameter $W_{j,r} = \text{diag}(\mathbf{w}_{j,r})$ drives the shrinkage of each entry of each marginal. Summarizing, for $p = 1, \ldots, I_j$, $j = 1, \ldots, J$ ($J = 4$ in Equation (15)) and $r = 1, \ldots, R$, the hierarchical prior structure (we use the shape-rate formulation for the gamma distribution) for each vector of the PARAFAC(R) decomposition in Equation (1) is

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

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 $b_{i_1,\ldots,i_J}$ of $\mathcal{B}$ is the law of a sum of product Normals (a product Normal is the distribution of the product of n independent centered 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 of the supplementary materials for further details). The peak at zero of the product Normal prior promotes shrinking effects. The following result characterizes the conditional prior distribution of an entry of the coefficient tensor $\mathcal{B}$ induced by the hierarchical prior in Equation (17). See Section S.5 for the proof, supplementary materials.

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

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

where $*$ denotes the convolution and

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

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

$$\begin{aligned}
&G_{4,0}^{4,0}\left(\beta_r^2 \prod_{h=1}^4 (2\tau\phi_r w_{h,r,m_h})^{-1} \middle| \mathbf{0}\right) \\
&\quad = \frac{1}{2\pi i}\int_{c-i^\infty}^{c+i^\infty} \left(\beta_r^2 \prod_{h=1}^4 (2\tau\phi_r w_{h,r,m_h})^{-1}\right)^{-s} \text{d}s \\
&K_r = (2\pi)^{-4/2} \prod_{h=1}^4 (2\tau\phi_r w_{h,r,m_h})^{-1}.
\end{aligned}$$

The use of Meijer G-functions and Fox H-functions is not new in econometrics. They arise as limiting distributions for the cointegrating vector in VECM models (e.g., Abadir and Paruolo 1997) and have been used for defining prior distributions in Bayesian analysis of nonconjugate Gaussian models (Andrade and Rathie 2015, 2017).

From Equation (4), we have that the covariance matrices Σ_j, $j = 1, \ldots, J$, 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 $\Sigma_j = \mathbf{I}_{I_j}$, 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 hyperparameter 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 Figure 2.

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

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

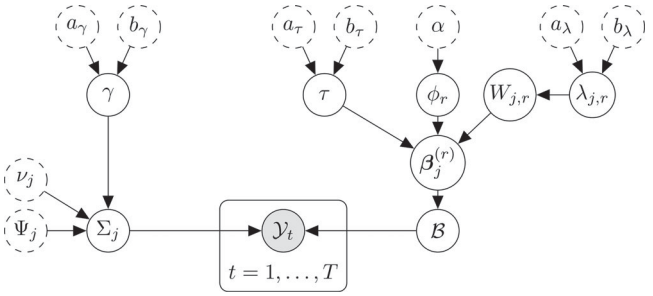

Figure 2. Directed acyclic graph of the model in Equation (15) and prior structure in Equations (17)–(18). Gray circles denote observable variables, white solid circles indicate parameters, white dashed circles indicate fixed hyperparameters. Directed edges represent the conditional independence relationships.

where $\mathbf{y}_{t-1} = \text{vec}(\mathcal{Y}_{t-1})$ and $\boldsymbol{\theta}$ denotes the collection of all parameters. Since the posterior distribution is not tractable, we adopt an MCMC procedure based on Gibbs sampling. The details of the derivation of the full conditional posterior distributions are given in Section S.6 of the supplementary materials. We articulate the sampler in three main blocks:

1. Sample the global and middle variance hyper-parameters of the marginals, from

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

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

   where $C_r = \sum_{j=1}^{J} \boldsymbol{\beta}_j^{(r)'} W_{j,r}^{-1} \boldsymbol{\beta}_j^{(r)}$, 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).
2. Sample the local variance hyper-parameters of the marginals and the marginals themselves, from

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

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

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

3. Sample the covariance matrices and the latent scale, from

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

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

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 investigated by several authors (e.g., Eaton and Kortum 2002; Fieler 2011), 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., 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 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 exposure. See Section S.9 for data description, supplementary materials.

We estimate the tensor autoregressive model in Equation (15), using the prior structure described in Section 3, and run 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 Figure 3. The (i, j)th entry of the matrix $\hat{B}_{(4)}$ reports the impact of the edge j on edge i in vectorized 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 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 Figure 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 weakly stationary. In fact, it can be shown that the stationarity of the mode-4 matricized coefficient tensor implies stationarity of the ART(1) process. Additional estimation results are provided in Section S.10 of the supplementary materials.

After estimating the ART(1) model (15), we may investigate shock propagation across the network computing generalized and orthogonalized impulse response functions presented in Equations (11) and (12), respectively. Impulse responses allow us to analyze 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 a country, by applying block Cholesky factorization to Σ, in such a way that the shocked country contemporaneously affects all others and not vice-versa (we do not report generalized IRFs, which are very similar). Thus, the matrices A and C in Equation (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 U.S. 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 Figure 4. We report the impact on the whole network (panel (a)) and, for illustrative purposes, the impact on Germany's transactions (panel (b)).

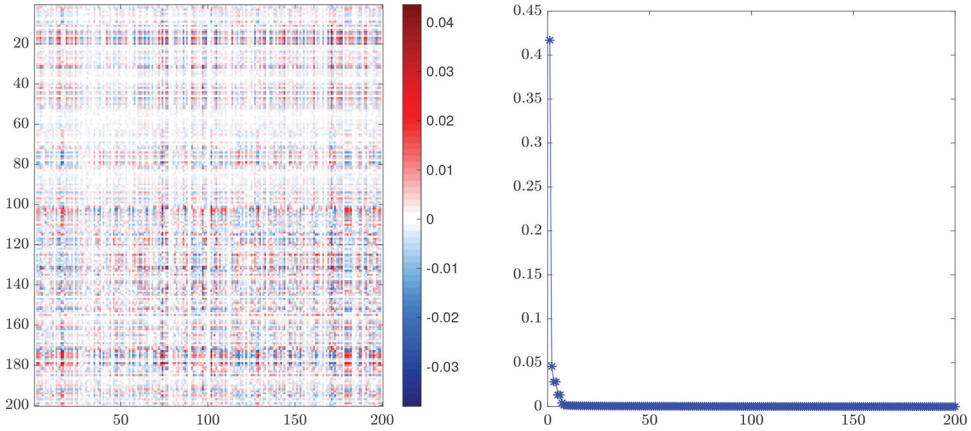

Figure 3. *Left:* mode-4 matricization of estimated coefficient tensor $\hat{B}_{(4)}$. *Right:* log-spectrum of $\hat{B}_{(4)}$, decreasing order.

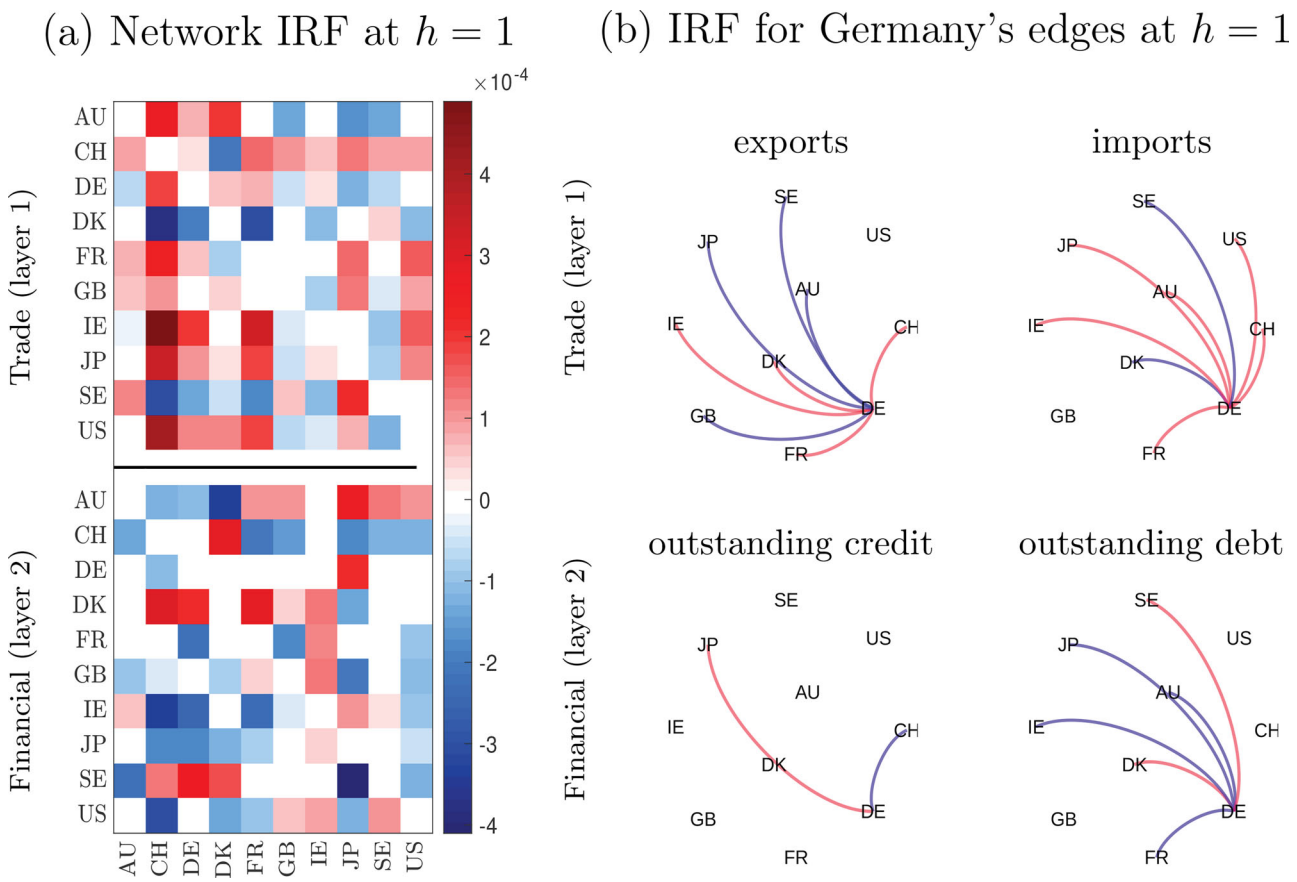

Figure 4. Shock to U.S. trade imports by −1%. IRF at horizon $h = 1$ for all (panel (a)) and Germany (panel (b)) financial and trade transactions. In each plot negative coefficients are in blue and positive in red.

Global effect on the network. The negative shock to U.S. 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 United States 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 U.S.'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 in exports to the United States, Japan and Ireland. In conclusion, model (15) permits to forecast trade creation and diversion effects (Bikker 2010).

Local effect on Germany. In panel (b) of Figure 4 we report the response of Germany's transactions to the negative shock in U.S. 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 in U.S. 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 United States (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 United States vis-à-vis almost all countries decreases after the shock. Overall, responses to a shock on U.S. 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 supplementary materials shows additional impulse responses to a (i) negative 1% shock to Great Britain's (GB) outstanding debt and (ii) 1% negative shock to GB's outstanding debt coupled with a 1% positive shock to GB's outstanding credit.

5. Conclusions

We defined a new and general statistical framework for dynamic tensor regression. It encompasses the autoregressive tensor model, called ART, and many 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 coefficient tensor to reduce the parameter space dimension and specified a global-local shrinkage prior to address 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 essential 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.

Supplementary Materials

The supplementary material to this article provides additional results about the method developed in the present article. In particular, Section S.1 contains background results on tensors, then Section S.2 provides the derivation of the tensor forecast error variance decomposition. Section S.3 gives an example of MAR. Section S.4 reports the proofs of the remarks in Section 2 of the main article. Sections S.5 and S.6 provide further details on the prior on tensor entries and on posterior computation, respectively, while Section S.7 describes the initialisation of the inferential algorithm. Section S.8 provides a summary of simulation results. Finally, Sections S.9 and S.10 contain a description of the data used in the empirical application and further plots of the estimation results.

Acknowledgments

We are grateful to Federico Bassetti, Sylvia Frühwirth-Schnatter, Christian Gouriéroux, Søren Johansen, Siem Jan Koopman, Gary Koop, André Lucas, Alain Monfort, Peter Phillips, Raquel Prado, Christian P. Robert, Mark Steel, and Mike West for their comments and suggestions. Also, we thank the seminar participants at Queen Mary University, CREST, University of Warwick, University of Southampton, Vrije University of Amsterdam, London School of Economics, Maastricht University, and Polytechnic University of Milan. We thank the participants at "ES Annual Meeting" in Milan (2020), "ICEEE" in Lecce (2019), "CFENetwork" in Pisa (2018), "EC2" in Rome (2018), "RCEA Annual meeting" in Rimini (2018), "CFENetwork" in London (2017), "ICEEE" in Messina (2017), "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), for their constructive comments.

Funding

This research used the SCSCF and HPC multiprocessor cluster systems and is part of the project Venice Center for Risk Analytics (VERA) at Ca' Foscari University of Venice. Monica Billio and Roberto Casarin acknowledge financial support from the Italian Ministry MIUR under the PRIN project Hi-Di NET – Econometric Analysis of High Dimensional Models with Network Structures in Macroeconomics and Finance (grant agreement no. 2017TA7TYC). Matteo Iacopini acknowledges financial support from the EU Horizon 2020 programme under the Marie Skłodowska-Curie scheme (grant agreement no. 887220).

References

Abadir, K. M., and Paruolo, P. (1997), "Two Mixed Normal Densities from Cointegration Analysis," *Econometrica*, 65, 671–680. [435]

Abraham, R., Marsden, J. E., and Ratiu, T. (2012), *Manifolds, Tensor Analysis, and Applications* (Vol. 75), New York: Springer. [431]

Aldasoro, I., and Alves, I. (2018), "Multiplex Interbank Networks and Systemic Importance: An Application to European Data," *Journal of Financial Stability*, 35, 17–37. [429]

Anacleto, O., and Queen, C. (2017), "Dynamic Chain Graph Models for Time Series Network Data," *Bayesian Analysis*, 12, 491–509. [430]

Andrade, J., and Rathie, P. (2017), "Exact Posterior Computation in Non-conjugate Gaussian Location-Scale Parameters Models," *Communications in Nonlinear Science and Numerical Simulation*, 53, 111–129. [435]

Andrade, J., and Rathie, P. N. (2015), "On Exact Posterior Distributions Using H-functions," *Journal of Computational and Applied Mathematics*, 290, 459–475. [435]

Aris, R. (2012), *Vectors, Tensors and the Basic Equations of Fluid Mechanics*, Courier Corporation, New York: Dover Publications Inc. [431]

Balazsi, L., L. Matyas, and T. Wansbeek (2015), "The Estimation of Multidimensional Fixed Effects Panel Data Models," *Econometric Reviews*, 37, 212–227. [429]

Behera, R., Nandi, A. K., and Sahoo,J. K. (2020), "Further Results on the Drazin Inverse of Even Order Tensors," *Numerical Linear Algebra with Applications*, 27, e2317. [430]

Bikker, J. A. (2010), "An Extended Gravity Model with Substitution Applied to International Trade," in *The Gravity Model in International Trade: Advances and Applications*, eds. P. A. G. van Bergeijk and S. Brakmanpp, pp. 135–164. Cambridge: Cambridge University Press. [436,437]

Canova, F., and Ciccarelli, M. (2004), "Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model," *Journal of Econometrics*, 120, 327–359. [435]

Cichocki, A. (2014), "Era of Big Data Processing: A New Approach via Tensor Networks and Tensor Decompositions," in *Proceedings of the International Workshop on Smart Info-Media Systems in Asia (SISA2013)*. [429]

Dawid, A. P., and Lauritzen, S. L. (1993), "Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models," *The Annals of Statistics*, 21, 1272–1317. [435]

De Paula, A. (2017), "Econometrics of Network Models," in *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, eds. B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, pp. 268–323, Cambridge: Cambridge University Press. [430]

Dobra, A. (2015), "Graphical Modeling of Spatial Health Data," in *Handbook of Spatial Epidemiology*, eds. A. B. Lawson, S. Banerjee, R. P. Haining, M. D. Ugarte, pp. 575–594, Boca Raton, FL: Chapman and Hall /CRC. [435]

Eaton, J., and Kortum, S. (2002), "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779. [436]

Fieler, A. C. (2011), "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, 79, 1069–1101. [436]

George, E. I., and McCulloch, R. E. (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373. [435]

Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014), "Smooth Scalar-on-Image Regression via Spatial Bayesian Variable Selection," *Journal of Computational and Graphical Statistics*, 23, 46–64. [430]

Guha, S., and Guhaniyogi, R. (2021), "Bayesian Generalized Sparse Symmetric Tensor-on-Vector Regression," *Technometrics*, 63, 160–170. [430]

Guha, S., and Rodriguez, A. (2020), "Bayesian Regression with Undirected Network Predictors with an Application to Brain Connectome Data," *Journal of the American Statistical Association*, 116, 581–593. [430]

Guhaniyogi, R. (2020), "Bayesian Methods for Tensor Regression," *Wiley StatsRef: Statistics Reference Online*, 1–18. [430]

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017), "Bayesian Tensor Regression," *Journal of Machine Learning Research*, 18, 1–31. [429,430,435]

Guhaniyogi, R., and Spencer, D. (2021), "Bayesian Tensor Response Regression with an Application to Brain Activation Studies," *Bayesian Analysis*, 16, 1221–1249. [430]

Hackbusch, W. (2012), *Tensor Spaces and Numerical Tensor Calculus*, Berlin: Springer. [429,430,431]

Hoff, P. D. (2011), "Separable Covariance Arrays via the Tucker Product, with Applications to Multivariate Relational Data," *Bayesian Analysis*, 6, 179–196. [435]

Hoff, P. D. (2015), "Multilinear Tensor Regression for Longitudinal Relational Data," *The Annals of Applied Statistics*, 9, 1169–1193. [430]

Holme, P., and Saramäki, J. (2012), "Temporal Networks," *Physics Reports*, 519, 97–125. [430]

Ishwaran, H., and Rao, J. S. (2005), "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *The Annals of Statistics*, 33, 730–773. [435]

Ji, J., and Wei, Y. (2018), "The Drazin Inverse of an Even-Order Tensor and its Application to Singular Tensor Equations," *Computers & Mathematics with Applications*, 75, 3402–3413. [430]

Kapetanios, G., Serlenga, L., and Shin, Y. (2021), "Estimation and Inference for Multi-Dimensional Heterogeneous Panel Datasets with Hierarchical Multi-Factor Error Structure," *Journal of Econometrics*, 220, 504–531. [429]

Kolda, T. G., and Bader, B. W. (2009), "Tensor Decompositions and Applications," *SIAM Review*, 51, 455–500. [431]

Koop, G., Pesaran, M. H., and Potter, S. M. (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147. [434]

Kostakos, V. (2009), "Temporal Graphs," *Physica A: Statistical Mechanics and its Applications*, 388, 1007–1023. [430]

Lee, N., and Cichocki, A. (2018), "Fundamental Tensor Operations for Large-Scale Data Analysis in Tensor Train Formats," *Multidimensional Systems and Signal Processing*, 29, 921–960. [431]

Li, L., and Zhang, X. (2017), "Parsimonious Tensor Response Regression," *Journal of the American Statistical Association*, 112, 1131–1146. [429,430]

Li, X., Xu, D., Zhou, H., and Li, L. (2018), "Tucker Tensor Regression and Neuroimaging Analysis," *Statistics in Biosciences*, 10, 520–545. [429]

Lock, E. F. (2018), "Tensor-on-Tensor Regression," *Journal of Computational and Graphical Statistics*, 27, 638–647. [429]

Mitchell, T. J., and Beauchamp, J. J. (1988), "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83, 1023–1032. [435]

Neal, R. M. (2011), "MCMC Using Hamiltonian Dynamics," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, pp. 113–162, Boca Raton, FL: Chapman & Hall /CRC. [436]

Ohlson, M., Ahmad, M. R., and Von Rosen, D. (2013), "The Multilinear Normal Distribution: Introduction and Some Basic Properties," *Journal of Multivariate Analysis*, 113, 37–47. [432,433,434]

Park, T., and Casella, G. (2008), "The Bayesian Lasso," *Journal of the American Statistical Association*, 103, 681–686. [435]

Pesaran, H. H., and Shin, Y. (1998), "Generalized Impulse Response Analysis in Linear Multivariate Models," *Economics Letters*, 58, 17–29. [434]

Rabusseau, G., and Kadri, H. (2016), "Low-Rank Regression with Tensor Responses," *Advances in Neural Information Processing Systems*, 29, 1867–1875. [429]

Raskutti, G., Yuan, M., and Chen, H. (2019), "Convex Regularization for High-Dimensional Multiresponse Tensor Regression," *The Annals of Statistics*, 47, 1554–1584. [430]

Spencer, D., Guhaniyogi, R., and Prado, R. (2020), "Joint Bayesian Estimation of Voxel Activation and Inter-Regional Connectivity in fMRI Experiments," *Psychometrika*, 85, 845–869. [430]

Sun, W. W., and Li, L. (2017), "Store: Sparse Tensor Response Regression and Neuroimaging Analysis," *The Journal of Machine Learning Research*, 18, 4908–4944. [430]

Wang, B., Du, H., and Ma, H. (2020), "Perturbation Bounds for DMP and CMP Inverses of Tensors via Einstein Product," *Computational and Applied Mathematics*, 39, 1–17. [430]

Wang, H., and West, M. (2009), "Bayesian Analysis of Matrix Normal Graphical Models," *Biometrika*, 96, 821–834. [435]

Xu, T., Yin, Z., Siliang, T., Jian, S., Fei, W., and Yueting, Z. (2013), "Logistic Tensor Regression for Classification," in *Intelligent Science and Intelligent Data Engineering*, eds. J. Yang, F. Fang, C. Sun, pp. 573–581, Berlin: Springer. [429]

Yang, Y., and Dunson, D. B. (2016), "Bayesian Conditional Tensor Factorizations for High-Dimensional Classification," *Journal of the American Statistical Association*, 111, 656–669. [430]

Yuan, M., and Zhang, C.-H. (2016), "On Tensor Completion via Nuclear Norm Minimization," *Foundations of Computational Mathematics*, 16, 1031–1068. [429]

Zhang, Y. D., Naughton, B. P., Bondell, H. D., and Reich, B. J. (2020), "Bayesian Regression Using a Prior on the Model Fit: The r2-d2 Shrinkage Prior," *Journal of the American Statistical Association*, 1–13. [430]

Zhou, H., Li, L., and Zhu, H. (2013), "Tensor Regression with Applications in Neuroimaging Data Analysis," *Journal of the American Statistical Association*, 108, 540–552. [429,430]

Zhou, J., Bhattacharya, A., Herring, A. H., and Dunson, D. B. (2015), "Bayesian Factorizations of Big Sparse Tensors," *Journal of the American Statistical Association*, 110, 1562–1576. [430]