Bayesian Tensor Autoregressive Models

Monica Billio

Ca' Foscari University of Venice, Italy

ECB Forecasting Conference, Frankfurt am Main

12-13 June 2023

Networks and Connectedness in Economics and Finance

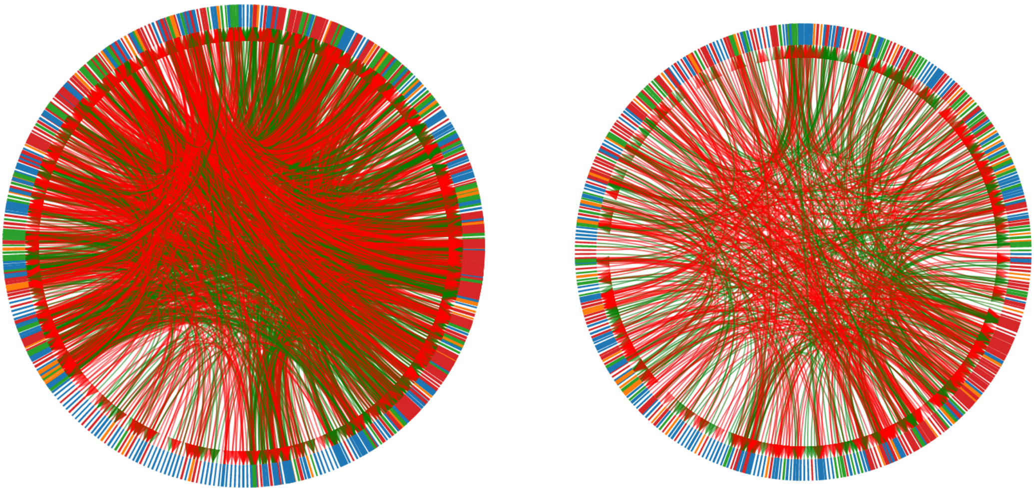


Figure: Example of a financial network in crisis and non-crisis periods.

Billio, Getmansky, Lo, Pelizzon (2012), Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors, Journal of Financial Economics, 104, 535-559

Networks @ Ca' Foscari

Better statistical tools to extract networks

Sparsity

- Ahelegbey, Billio, Casarin (2016a), "Bayesian Graphical Models for Structural Vector Autoregressive Processes" Journal of Applied Econometrics, 31(2), 357-386.
- Ahelegbey, Billio, Casarin (2016b), "Sparse Graphical Vector Autoregression: A Bayesian Approach", Annals of Economics and Statistics, 123/124, 1-30.
- Billio, Casarin, Rossini (2019), "Bayesian nonparametric sparse VAR models", Journal of Econometrics, 212(1), 97-115.

Breaks and regimes

- Bianchi, Billio, Casarin, Guidolin (2019), "Modelling Systemic Risk with Markov Switching Graphical SUR Models" Journal of Econometrics, 210(1), 58-74.
- Ahelegbey, Billio, Casarin (2021), "Modeling Turning Points in the Global Equity Market", Econometrics and Statistics, forthcoming.

Networks @ Ca' Foscari

Impact of network connectivity

Impact of connectivity

- Billio, Caporin, Panzica, Pelizzon (2022), "The impact of network connectivity on factor exposures, asset pricing, and portfolio diversification" International Review of Economics and Finance, 84, 196-223.
- Billio, Pelizzon, Frattarolo (2022), "Networks in risk spillovers: A multivariate GARCH perspective", Econometrics and Statistics, forthcoming.
- Agudze, Billio, Casarin, Ravazzolo (2022), Markov Switching Panel with Endogenous Synchronization Effects, Journal of Econometrics, 230(2), 281-298.

Networks @ Ca' Foscari

New network connectivity and complexity measures

Entropy

- Billio, Casarin, Costola, Pasqualini (2016), "An entropy-based early warning indicator for systemic risk" Journal of International Financial Markets, Institutions and Money, 45, 42-59.
- Billio, Casarin, Costola, Frattarolo (2019), "Contagion dynamics on financial networks", in J. Chevallier, S. Goutte, D. Guerreiro, S. Saglio and B. Sanhaji (Eds.) International Financial Markets (Vol 1), Routledge Advances in Applied Financial Econometrics.

Opinion Dynamics

- Billio, Casarin, Costola, Frattarolo (2018), "Disagreement in Signed Financial Networks", in M. Corazza, M. Durbán, A. Grané, C. Perna and M. Sibillo (Eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer Verlag.
- Billio, Casarin, Costola, Frattarolo (2019), Opinion Dynamics and Disagreements on Financial Networks, Advances in Decision Sciences, 23(4), 1-27.

Networks @ Ca' Foscari

From **network extraction** to **modelling** temporal sequences of networks

General research questions

Q: how to design suitable models for random networks?
Q: how to measure the impact of randomness on standard network statistics?
Q: how to model and forecast temporal networks?

Challenges

- guarantee model parsimony
- extend standard econometric models to network data (preserve interpretability)
- allow for model flexibility (exploit data structure)
- develop feasible inference methods
- deal with the computational cost

$\Rightarrow$ **New models for networks and temporal networks**

Networks @ Ca' Foscari

New models for networks and temporal networks

Matrix models

- Billio, Casarin, Costola, Iacopini (2021), "COVID-19 spreading in financial networks: A semiparametric matrix regression model", Econometrics and Statistics, forthcoming
- Billio, Casarin, Costola, Iacopini (2021) "A matrix-variate t model for networks", Frontiers in Artificial Intelligence, 4, 49.
- Billio, Casarin, Costola, Iacopini (2022), "Matrix-variate Smooth Transition Models for Temporal Networks", Innovations in Multivariate Statistical Modeling, Springer, 1, 137-167

Tensor models

In this presentation

- Billio, Casarin, Iacopini, Kaufmann (2023), "Bayesian Dynamic Tensor Regression" Journal of Business and Economic Statistics, 41(2), 429-439.
- Billio, Casarin, Iacopini (2023), "Bayesian Markov switching Tensor regression for time-varying networks" Journal of the American Statistical Association (Theory & Methods), forthcoming

Introduction

Array data

By array data we mean data occurring in the shape of matrices or multi-dimensional arrays (i.e. **tensors**).

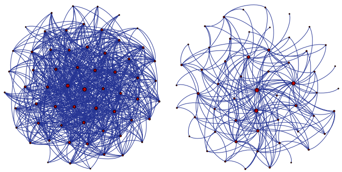
Networks

A network (or graph) $\mathcal{G} = (V, E)$ is given by a set of vertices, V, and a collection of edges, E, between them.

$\Rightarrow$ may represent the dependence between random variables (vertices).

Array and **network data:**

- high dimensionality
- meaningful and complex structure
- dynamic structure
- multiple layers
- time-varying sparse topologies



New time varying dynamic networks

Contributions

- methods & models ⇒ array data
- application ⇒ network data

Proposals for **dynamic network** modelling of **edge** information:

- [**BDTR**] Billio, Casarin, Iacopini, Kaufmann (2022), "*Bayesian Dynamic Tensor Regression*" (**this talk**)
  - ⇒ multi-layer networks with dynamic, **real-valued edges**
  - ⇒ smooth dynamics
- [**BMSTR**] Billio, Casarin, Iacopini (2022), "*Bayesian Markov Switching Tensor Regression for Time Varying Networks*"
  - ⇒ multi-layer networks with dynamic, **binary edges**
  - ⇒ discrete switching dynamics

Questions and aims

Research questions:

Q: possible to exploit information from the structure of data?

Q: how to model a time series of tensor data?

Q: more data, few relevant ⇒ how to account for sparsity?

Goals:

(i) propose dynamic models for tensors of data

(ii) account for different types of data and dynamics

(iii) explore dynamics of shock propagation (impulse-response) on real-valued networks

Our proposal:

1) use tensors ⇒ operations and representations

2) use global-local hierarchical prior distributions ⇒ sharing of information and sparsity recovery

Motivation

Q: why not vectorize?

- ✗ estimation is infeasible
- ✗ requires unclear restrictions on coefficients
- ✗ disregards topological information in the structure of data

Q: why use **tensors**?

- ✓ estimation is feasible
- ✓ preserve and exploit data structure information
- ✓ powerful decompositions and operators

General model formulation and parametrisation allows:

- generalisation of linear regression models to tensor framework
- parsimonious model specification
- learn sparsity patterns from data
- allows for flexible prior definition and efficient posterior computation

BDTR paper – Motivation

Research questions:

Q: how to exploit data structure information?

Q: how to model a time series of array data?

Q: more data, few relevant $\Rightarrow$ how to account for sparsity?

Goals:

(i) provide a model able to deal with array data

(ii) explore dynamic process of real-valued networks/graphs

(iii) analyse shock propagation through time and space

BDTR paper – Methods

- **Methods:**

$$\mathbf{y}_{\mathbf{i},t} = \beta'_{\mathbf{i}} \times_{D+1} \mathrm{vec}(\mathcal{X}_t) + \epsilon_t$$

$$\Downarrow$$

$$\mathcal{Y}_t = \mathcal{B} \times_{D+1} \mathrm{vec}(\mathcal{X}_t) + \mathcal{E}_t$$

  - linear regression model for **tensor** time series data
  - generalization of multivariate linear regression
  - tensor-valued impulse response analysis
  - PARAFAC tensor decomposition $\Rightarrow$ parsimony
  - hierarchical global-local shrinkage prior $\Rightarrow$ sparse coefficients

- **Application:**
  - 2-layer network, international trade + capital flow
  - analysis of edge-shock propagation

BMSTR paper – Motivation

Research questions:

Q: how to model a time series of binary networks?

Q: how to study structural breaks in network structure?

Q: hot to account for different sparsity patterns?

Goals:

(i) provide model for time varying binary graphs

(ii) infer regimes driving the graphical structure

(iii) uncover role of economic variables in affecting edge probability

BMSTR paper – Methods

- **Methods:**

$$x_{ijk,t}|\rho_t, \mathbf{g}_{ijk}(t) \sim \rho(t)\delta_{\{0\}}(x_{ijk,t}) + (1-\rho(t))\mathcal{B}ern(x_{ijk,t}|\psi_{ijk,t})$$
$$\psi_{ijk,t} = \frac{\exp\{\mathbf{z}_t'\mathbf{g}_{ijk}(t)\}}{1+\exp\{\mathbf{z}_t'\mathbf{g}_{ijk}(t)\}}.$$

  - zero-inflated logit for each entry
  - Markov switching dynamics for parameters
  - Bayesian inference via Pólya-Gamma data augmentation
  - PARAFAC tensor decomposition $\Rightarrow$ parsimony
  - hierarchical global-local shrinkage prior $\Rightarrow$ sparse coefficients

- **Application:**
  - financial network EU institutions
  - impact of risk factors and network topology on edge probability

Bayesian Tensor Autoregressive Models

Monica Billio

Ca' Foscari University of Venice, Italy

ECB Forecasting Conference, Frankfurt am Main

12-13 June 2023

Motivation

Availability of data:

(i) increasing size ⇒ high dimensionality
(ii) multiple data sources ⇒ multiple "layers" (e.g., cross section, time, space, ...)

⇒ gathered or meaningfully rearranged into multidimensional arrays (tensors).

Example 1.

Tensor-valued data:

- multi-country panel: m variables, n countries, t times → 3-order tensor (e.g., Hoff (2015), Canova and Ciccarelli (2004)).
- temporal networks: relations between n subjects, observed t times → 3-order tensor (e.g., financial networks Billio et al. (2012)).
- medical data: sequence of $n \times m$ brain images → 3-order tensor (e.g., Zhou et al. (2013), Li and Zhang (2017)).
- **multi-layer networks**: relations between n subjects, d attributes, observed t times → 4-order tensor (e.g., social networks Hoff et al. (2002), Hoff (2011), Hoff (2015))

Motivation: COMTRADE & BIS Multi-Layer Networks

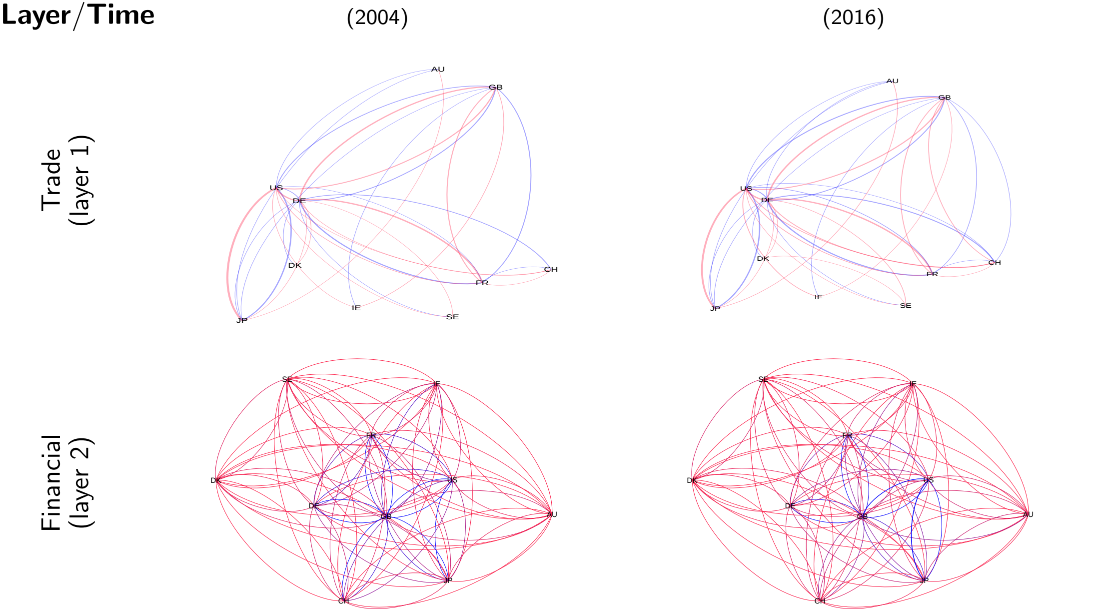


Figure: International trade and financial networks. Nodes: countries. Edges: flows.

Motivation: COMTRADE & BIS Multi-Layer Networks

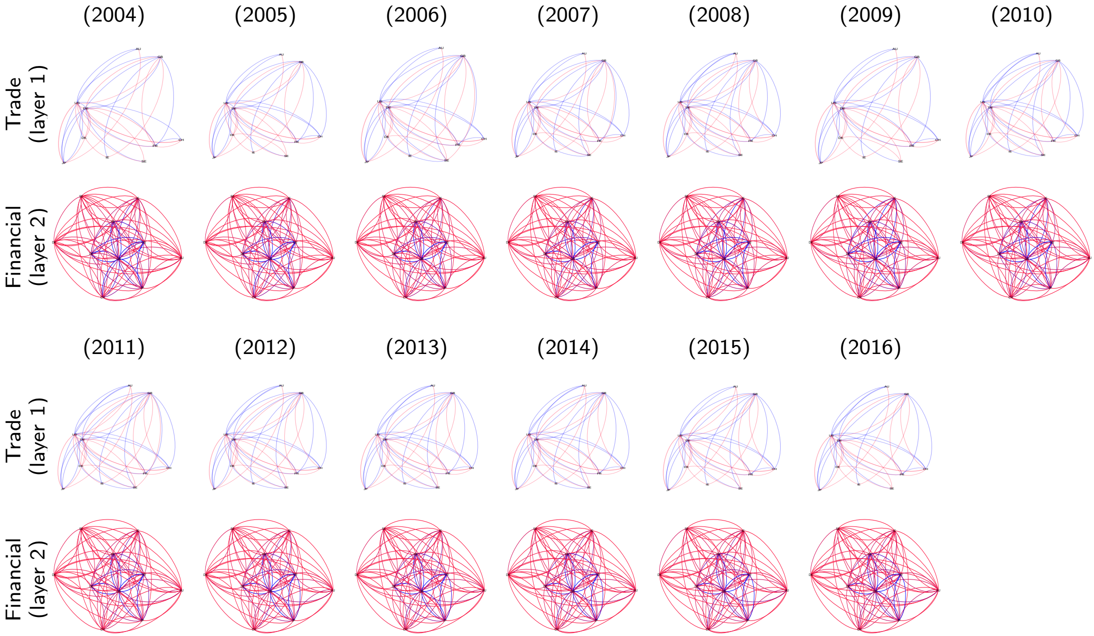


Figure: International trade and financial **temporal networks**. Nodes: countries. Edges: flows.

Questions and Aims

Research questions:

Q: how to model a time series of tensor-valued data?

Q: many variables, few relevant $\Rightarrow$ how to account for sparsity?

Q: possible to exploit information from the structure of the data?

Goals:

G: provide a **dynamic model** for tensor-valued data

G: explore dynamics of (**shock propagation**) on tensors

Our contribution:

C1: use tensors algebra (spaces, operations and representations)

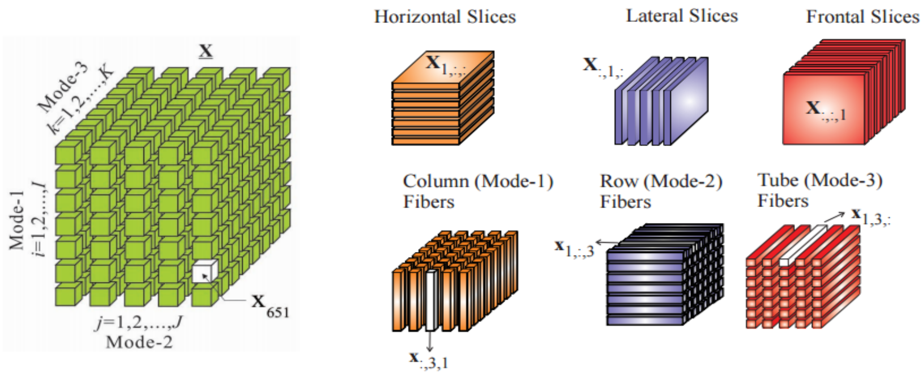
C2: use global-local hierarchical prior distributions (information sharing, sparsity)

C3: extend to tensor dynamic models the impulse response analysis

Tensors

Definition 1 (Tensor).

A real valued order-D tensor is an array $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$.



✓ Tensor algebra generalizes matrix algebra to multiple dimensions

Tensors Operations

Definition 2 (Matricisation).

Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ be a order-N tensor. The mode-k matricisation mat_k is the operator defined as:

$$\text{mat}_k : \mathbb{R}^{I_1 \times \ldots \times I_N} \to \mathbb{R}^{I_k \times I_{-k}}$$

which maps a tensor $\mathcal{X}$ of dimensions $(I_1, \ldots, I_N)$ into a matrix X of size $(I_k \times I_{-k})$, where $I_{-k} = \prod_{j \neq k} I_j$.

Remarks:

- "cut" the tensor into slices of I_k rows $\to$ stack slices horizontally
- $\text{vec}(\mathcal{X}) = \text{mat}_{I^*}(\mathcal{X})$, with $I^* = \prod_j I_j$

Tensors Operations

Definition 3 (Mode-n product).

Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ be a order-N tensor, $A \in \mathbb{R}^{J \times I_n}$ and $\mathbf{v} \in \mathbb{R}^{I_n}$.

The mode-n product $\times_n$ is defined as follows:

$$(\mathcal{X} \times_n A)_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} := \sum_{i_n=1}^{I_n} x_{i_1,\ldots,i_n,\ldots,i_N} a_{j,i_n}$$

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

Idea: compute the inner product of each mode-n fiber with the matrix/vector.

Effect: change n-th dimension of the tensor or reduces its order by one.

- Some operations performed in usual way (e.g., inner/Hadamard product, ... - see also Kolda and Bader (2009), Cichocki et al. (2016))

Tensors Operations

Definition 4 (Contracted product).

The contracted product $\mathcal{X} \bar{\times}_N \mathcal{Y}$ 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 $\mathcal{Y} \in \mathbb{R}^{I_1 \times \ldots \times I_N \times H_1 \times \ldots \times H_M}$ is a $(K+M)$-order tensor defined as

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

- It has the mode-n product as special case when $N=1$ and $M=0$ (i.e. $\mathcal{Y} = \mathbf{y}$).

Tensors Representations

Powerful tool: several **tensor representations/decompositions** available (Tucker, PARAFAC, ...)

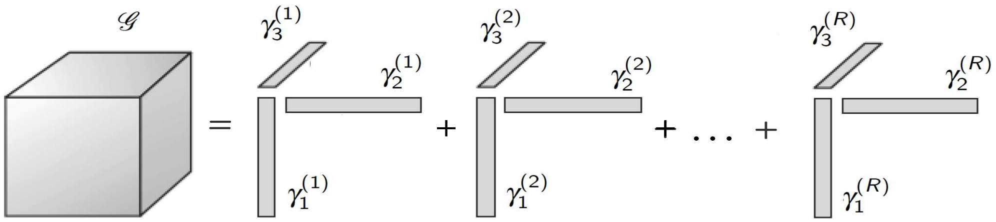
Definition 5 (PARAFAC(R) decomposition).

Let $\mathcal{G} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and let $R \in \mathbb{N}$ be the rank of $\mathcal{G}$. It holds:

$$\mathcal{G} = \sum_{r=1}^{R} \gamma_1^{(r)} \circ \ldots \circ \gamma_N^{(r)}, \qquad \gamma_j^{(r)} \in \mathbb{R}^{I_j}. \tag{1}$$

where $\circ$ is the outer product: $(\gamma_1 \circ \ldots \circ \gamma_N)_{i_1,\ldots,i_N} = \gamma_{1,i_1} \cdots \gamma_{N,i_N}$

Remark: multi-dimensional analogue of **matrix low rank** decomposition.



A Tensor Model - Idea

Tensor Regression

For each entry of the response tensor:

$$y_{\mathbf{i},t} = \beta'_{\mathbf{i}} \operatorname{vec}(\mathcal{X}_t) + \epsilon_{\mathbf{i},t}, \tag{2}$$

where $\mathbf{i} := (i_1, \ldots, i_N)$. Compactly:

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{B} \times_{N+1} \operatorname{vec}(\mathcal{X}_t) + \mathcal{E}_t \\ \mathcal{E}_t &\sim \mathcal{N}_{I_1,\ldots,I_N}(\mathbf{0}, \Sigma_1, \ldots, \Sigma_N) \end{aligned} \tag{3}$$

- $\mathcal{Y}_t, \mathcal{X}_t$: response and regressor **tensors**, with possibly <u>different order</u> and/or size
- $\mathcal{B}$: coefficient **tensor**, with $N + 1$ dimensions
- $\mathcal{E}_t$ noise, with tensor Normal distribution (see Ohlson et al. (2013))
- straightforward inclusion of other regressors: scalars, vectors, matrices, ...

A Tensor Model - Idea

Tensor regression - Vectorised form

Given the **tensor** model

$$\mathcal{Y}_t = \mathcal{B} \times_{N+1} \text{vec}(\mathcal{X}_t) + \mathcal{E}_t \,, \qquad \mathcal{E}_t \sim \mathcal{N}_{I_1,\ldots,I_N}(\mathbf{0}, \Sigma_1, \ldots, \Sigma_N) \tag{3}$$

the corresponding **vectorised** model is

$$\begin{aligned} \text{vec}(\mathcal{Y}_t) &= \text{mat}_{N+1}(\mathcal{B})\,\text{vec}(\mathcal{X}_t) + \text{vec}(\mathcal{E}_t) \\ \Leftrightarrow \mathbf{y}_t &= B'_{N+1}\mathbf{x}_t + \boldsymbol{\epsilon}_t \,, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_N \otimes \ldots \otimes \Sigma_1), \end{aligned} \tag{4}$$

where $\text{mat}_k(\cdot)$ is the mode-k *matricization* operator mapping to a matrix of size $d_k \times d_{-k}$ (where $d_{-k} = \prod_{i \neq k} d_i$).

Remarks:

- Kronecker structure of vectorised model's covariance matrix
- parametrisation for $\mathcal{B}$ mapped to parametrisation for B_{N+1}

Existing Special cases

Univariate regression

If $I_j = 1, \ \forall j \in \{1, \dots, N\}$, then model (3) reduces to:

$$y_t = \beta' \operatorname{vec}(\mathcal{X}_t) + \epsilon_t = \beta' \mathbf{x}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{5}$$

Multivariate regression

If $I_j = 1, \ \forall j \in \{2, \dots, N\}$, then model (3) reduces to:

$$\mathbf{y}_t = B \times_2 \operatorname{vec}(\mathcal{X}_t) + \boldsymbol{\epsilon}_t = B\mathbf{x}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}_{I_1}(\mathbf{0}, \Sigma) \tag{6}$$

Examples:

- **SUR**, when $\mathcal{X}_t = (I_{nm} \otimes X)$ with $X = [X_1, \dots, X_n]$, $X_i \in \mathbb{R}^{m \times k_i}$, $\mathbf{y}_t \in \mathbb{R}^{nm}$
- **VAR**, **VECM**, **MAI**, when $\mathcal{X}_t = \mathbf{y}_{t-1}$
- **Panel VAR**, when $\mathbf{y}_t = [\mathbf{y}_{1t}, \mathbf{y}_{2t}]$ and $\operatorname{vec}(\mathcal{X}_t) = \mathbf{x}_t = g(\mathbf{y}_{t-1})$

New Special cases - Tensor Autoregressive

Matrix autoregressive model

A particular case of model (3) is a MAR(1), when $\mathcal{Y}_t \in \mathbb{R}^{I \times J}$ and $\mathcal{X}_t = \mathcal{Y}_{t-1}$

$$Y_t = \mathcal{B} \times_3 \text{vec}(Y_{t-1}) + E_t, \qquad E_t \sim \mathcal{N}_{I,J}(\mathbf{0}, \Sigma_1, \Sigma_2). \tag{7}$$

More generally, a MAR(p) for $p \in \mathbb{N}$ is given by

$$Y_t = \sum_{i=1}^{p} \mathcal{B}_i \times_3 \text{vec}(Y_{t-i}) + E_t, \qquad E_t \sim \mathcal{N}_{I,J}(\mathbf{0}, \Sigma_1, \Sigma_2). \tag{8}$$

Use of matrix variate models/distributions:

- state space time series models Harrison and West (1999)
- Gaussian graphical models Carvalho et al. (2007)
- dynamic linear models Carvalho and West (2007), Wang and West (2009)
- longitudinal data classification and modelling Viroli (2011), Viroli and Anderlucci (2013)
- matrix regression Viroli (2012), Ding and Cook (2018)

New special cases - Tensor Autoregressive

Tensor autoregressive of order 1

When $\mathcal{Y}_t$ is an order-D tensor and $\mathcal{X}_t = \mathcal{Y}_{t-1}$, then we get as particular case of model (3), a **tensor autoregressive model** ART(1):

$$\mathcal{Y}_t = \mathcal{B} \times_{N+1} \text{vec}(\mathcal{Y}_{t-1}) + \mathcal{E}_t, \qquad \mathcal{E}_t \sim \mathcal{N}_{I_1,\dots,I_N}(\mathbf{0}, \Sigma_1, \dots, \Sigma_N). \tag{9}$$

Tensor autoregressive of order p

More generally, we can define a ART(p), for $p \in \mathbb{N}$, as:

$$\mathcal{Y}_t = \sum_{i=1}^{p} \mathcal{B}_i \times_{N+1} \text{vec}(\mathcal{Y}_{t-i}) + \mathcal{E}_t, \qquad \mathcal{E}_t \sim \mathcal{N}_{I_1,\dots,I_N}(\mathbf{0}, \Sigma_1, \dots, \Sigma_N). \tag{10}$$

ART(p) and its properties

Proposition 1 (Properties of ART).

The following properties of the ART(p) process in Eq. 10 can be proved (see main paper)

(1) it has an equivalent representation in terms of the contracted product

(2) it has an equivalent representation as a state-augmented ART(1) process

(3) under mild conditions on the coefficient tensor, the process is weakly stationary and has an infinite moving average representation

(4) a sufficient condition for weak stationarity can be tested on the associated VAR model

Properties of ART(1)

For studying the stability of the process, we use an equivalent compact representation of the multilinear system obtained through the contracted product that provides a natural setting for multilinear forms, decompositions and inversions:

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{A}_0 + \sum_{j=1}^{p} \widetilde{\mathcal{A}}_j \bar{\times}_N \mathcal{Y}_{t-j} + \widetilde{\mathcal{B}} \bar{\times}_M \mathcal{X}_t + \mathcal{E}_t, \\ \mathcal{E}_t &\overset{iid}{\sim} \mathcal{N}_{I_1,\ldots,I_N}(\mathcal{O}, \Sigma_1, \ldots, \Sigma_N), \end{aligned} \tag{11}$$

where $\bar{\times}_{a,b}$ is a shorthand notation for the contracted product $\times^{1\ldots a}_{a+1\ldots a+b}$ and $\bar{\times}_a$ stands for $\bar{\times}_{a,0}$.

Proposition 2 (Stationarity).

If $\rho(\widetilde{\mathcal{A}}_1) < 1$ and the process $\mathcal{X}_t$ is weakly stationary, then the ART process in eq. (11), with $p = 1$, is weakly stationary and admits the representation

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

Properties of ART(1)

By vectorising the ART(1) in (9), we get the equivalent VAR representation

$$\text{vec}(\mathcal{Y}_t) = \mathbf{B}'_{(4)} \text{vec}(\mathcal{Y}_{t-1}) + \text{vec}(\mathcal{E}_t), \quad \text{vec}(\mathcal{E}_t) \overset{iid}{\sim} \mathcal{N}_{I^*}(\mathbf{0}, \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1). \tag{12}$$

Proposition 3.

The VAR(1) in eq. (12) is weakly stationary if and only if the ART(1) in eq. (11) is weakly stationary. An equivalent result holds for any $p \geq 1$.

Proposed Parametrization

Parsimonious Parametrization of the Covariances

unrestricted VAR(1)

$$\underbrace{\prod_{j=1}^{N+1} I_j}_{\text{coeff}} + \underbrace{\frac{1}{2}\prod_{j=1}^{N} I_j \prod_{j=1}^{N} (I_j+1)}_{\text{covariance}}$$

ART(1)

$$\underbrace{\prod_{j=1}^{N+1} I_j}_{\text{coeff}} + \underbrace{\frac{1}{2}\sum_{i=1}^{N} I_j(I_j+1)}_{\text{covariance}}$$

Parsimonious Parametrization of the Coefficients

PARAFAC(R) decomposition for $\mathcal{B}$

$$\mathcal{B} = \sum_{r=1}^{R} \beta_1^{(r)} \circ \ldots \circ \beta_N^{(r)}$$

Restricted ART(1): $R \sum_{j=1}^{N+1} I_j + \sum_{j=1}^{N} I_j(I_j+1)/2 \quad \Longrightarrow \quad$ estimation **feasible**

Proposed Parametrization

Parsimonious Parametrization

unrestricted VAR(1)

$$\prod_{j=1}^{N+1} I_j + \frac{1}{2}\prod_{j=1}^{N} I_j \prod_{j=1}^{N} (I_j + 1)$$

ART(1) with PARAFAC(R)

$$R\sum_{j=1}^{N+1} I_j + \frac{1}{2}\sum_{j=1}^{N} I_j(I_j + 1)$$

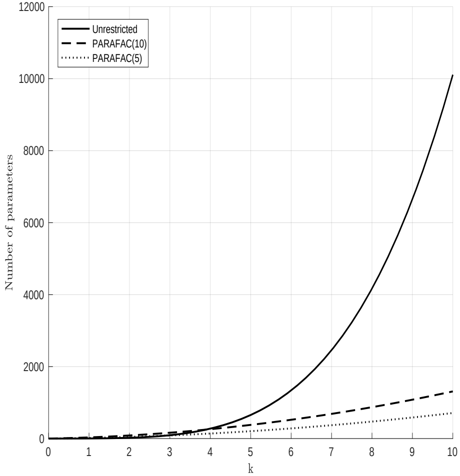


Figure: parameter reduction.

Parametrization Issues

Q: Identification of PARAFAC marginals $\beta_h^{(r)}$?

(i) scale invariance

$$\lambda_{1r}\beta_1^{(r)} \circ \ldots \circ \lambda_{Nr}\beta_N^{(r)} = \beta_j^{(r)} \circ \beta_i^{(r)}, \qquad \forall\, \lambda_{jr} : \prod_j \lambda_{jr} = 1$$

(ii) permutation invariance

$$\beta_1^{\pi(r)} \circ \ldots \circ \beta_N^{\pi(r)} = \beta_1^{(r)} \circ \ldots \circ \beta_N^{(r)}, \qquad \forall \text{ permutation } \pi(\cdot)$$

(iii) (if $N = 2$) invariance up to multiplication by orthonormal vectors

$$\left(\beta_j^{(r)}\mathbf{c}'\right) \circ \left(\beta_i^{(r)}\mathbf{c}'\right) = \beta_j^{(r)} \circ \beta_i^{(r)}, \qquad \forall\, \mathbf{c} \in \mathbb{R}^{d_j} : \mathbf{c}'\mathbf{c} = 1$$

Remark 1 (PARAFAC Parametrisation).

- reduces the size of parameter space
- coefficient tensor $\mathcal{B}$ always **identified**
- no interest in marginals $\beta_j^{(r)}$

Example - matrix autoregressive MAR(1)

Vectorised MAR(1) with PARAFAC(R) parametrisation

The **vectorised form** of the MAR(1) model (7) with a PARAFAC(R) decomposition on the tensor coefficient $\mathcal{B}$ is equivalent to a VAR(1) with restricted parameters:

$$\text{vec}(Y_t) = \text{mat}_3(\mathcal{B})' \, \text{vec}(Y_{t-1}) + \text{vec}(E_t), \qquad E_t \sim \mathcal{N}(\mathbf{0}, \Sigma_1, \Sigma_2)$$

$$\mathbf{y}_t = B_3' \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma). \tag{13}$$

The coefficient matrix B_3' and the covariance matrix Σ are given by

$$B_3' = \sum_{r=1}^{R} \boldsymbol{\beta}_3^{(r)\,\prime} \otimes \text{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)}\right), \qquad \Sigma = \Sigma_2 \otimes \Sigma_1$$

Parameters in this example:

unrestricted VAR(1)	MAR(1) with PARAFAC(R)
$\prod_{j=1}^{3} I_j + \frac{1}{2} \prod_{j=1}^{2} I_j (\prod_{j=1}^{2} I_j + 1)$	$R \sum_{j=1}^{3} I_j + \frac{1}{2} \sum_{j=1}^{2} I_j (I_j + 1)$

Prior Specification

Hierarchical **global-local** shrinkage prior for tensor marginals:

$$\pi(\beta_h^{(r)}|\tau, \phi_r, W_{h,r}) \sim \mathcal{N}_{I_h}(\mathbf{0}, \underbrace{\tau}_{\text{global}} \underbrace{\phi_r}_{\text{comp}} \underbrace{W_{h,r}}_{\text{local}}) \qquad \forall\, h, r$$

- global and component parts

$$\pi(\tau) \sim \mathcal{G}a(\overline{a}_\tau, \overline{b}_\tau), \qquad \pi(\phi) \sim \mathcal{D}ir(\overline{\boldsymbol{\alpha}})$$

- local part

$$\pi(\lambda_{h,r}) \sim \mathcal{G}a(\overline{a}_\lambda, \overline{b}_\lambda), \qquad \pi(w_{h,r,k}|\lambda_{h,r}) \sim \mathcal{E}xp(\lambda_{h,r}^2/2)$$

Noise covariances

$$\pi(\gamma) \sim \mathcal{G}a(\overline{a}_\gamma, \overline{b}_\gamma), \qquad \pi(\Sigma_h|\gamma) \sim \mathcal{IW}_{I_h}(\overline{\nu}_h, \gamma\overline{\Psi}_h)$$

Prior Specification

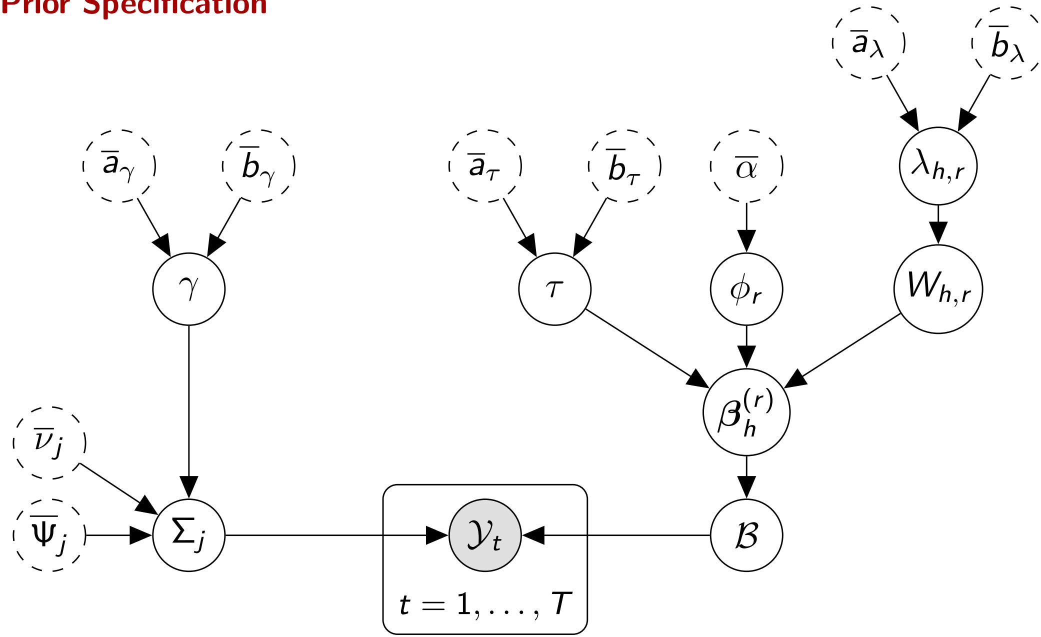


Figure: DAG of prior structure and model.

Posterior Computation - Gibbs sampler

Step 1. sample global and component variance hyper-parameters from

- collapsed Gibbs: $p(\psi_r|\mathcal{B},\mathbf{W}) \sim \text{GiG}(\alpha - d_0/2, 2b_\tau, 2C_r)$ then $\phi_r = \psi_r / \sum_l \psi_l$
- $p(\tau|\phi,\mathcal{B},\mathbf{W}) \sim \text{GiG}(a_\tau - Rd_0/2, 2b_\tau, 2\sum_r N_r)$

Step 2. sample local variance hyper-parameters and tensor marginals from

- $p(\lambda_{h,r}|\phi_r,\tau,\beta_h^{(r)}) \sim \mathcal{G}a\left(a_\lambda + I_h, b_\lambda + \left\|\beta_h^{(r)}\right\|_1 / \sqrt{\tau\phi_r}\right)$
- $p(w_{h,r,k}|\lambda_{h,r},\phi_r,\tau,\beta_h^{(r)}) \sim \text{GiG}\left(\frac{1}{2}, \lambda_{h,r}^2, \beta_{h,k}^{(r)^2}/(\tau\phi_r)\right) \ \forall\, k \in [1, I_h]$
- $p(\boldsymbol{\beta}_h^{(r)}|\boldsymbol{\beta}_{-h}^{(r)},\mathcal{B}_{-r},\boldsymbol{\phi},\tau,\mathbf{Y},\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4) \sim \mathcal{N}_{I_h}(\boldsymbol{\mu}_{\boldsymbol{\beta}_h},\Sigma_{\boldsymbol{\beta}_h})$

Step 3. sample noise covariance matrices from

- $p(\gamma|\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4) \sim \mathcal{G}a(\overline{a}_\gamma + (\sum_{h=1}^4 \overline{\nu}_h + TI_h)/2,\ \overline{b}_\gamma + \text{tr}(\sum_{h=1}^4 \overline{\Psi}_h \Sigma_h^{-1})/2)$
- $p(\Sigma_h|\gamma,\Sigma_{-h},\mathcal{B},\mathbf{Y}) \sim \mathcal{IW}_{I_h}(\overline{\nu}_h + TI_h,\ \gamma\overline{\Psi}_h + S_h)$

Application I - COMTRADE data

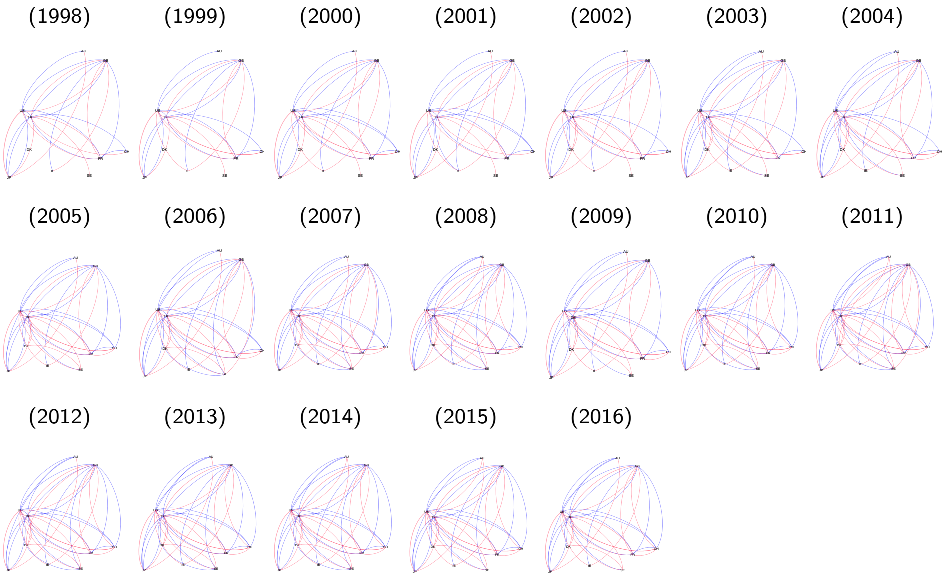


Figure: Trade network from 1998 (*top left*) to 2016 (*bottom right*). Nodes are countries, red and blue edges stand for exports and imports between two countries. Edge thickness represents flow magnitude.

Empirical Application - Single layer network

Matrix autoregressive model - MAR(1)

$$Y_t = \mathcal{B} \times_3 \text{vec}(Y_{t-1}) + E_t \,, \qquad E_t \sim \mathcal{N}_{10,10}(\mathbf{0}, \Sigma_1, \Sigma_2) \tag{14}$$

- mode-3 matricized tensor: $\text{mat}_3(\mathcal{B})' = B_3' = \left[\text{vec}(\mathcal{B}_{::1}), \text{vec}(\mathcal{B}_{::2}), \ldots, \text{vec}(\mathcal{B}_{::100})\right]$
- entry (i, j) of B_3': impact edge j $[t-1] \rightarrow i$ $[t]$

Note: vertical regularities = transaction at $t-1$ having similar impact on *all* transactions at t

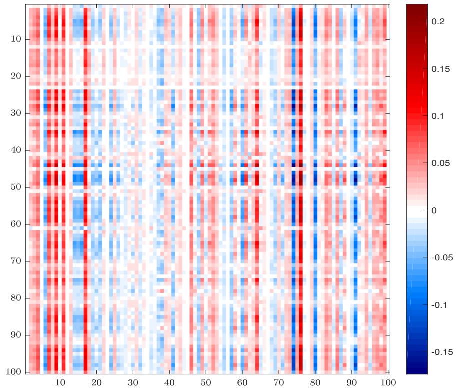


Figure: Estimated $\hat{B}_3'$.

Properties of ART(1) - Impulse Response Function

Definition 2 (Block-orthogonalized IRF for tensor models).

Denote Σ the covariance matrix of the vectorised tensor autoregressive model ART(1). We propose the **block-orthogonalised** *impulse response function* from the transformation

$$\begin{aligned} \text{vec}(\mathcal{Y}_t) &= \sum_{i=0}^{\infty} \Phi_i \boldsymbol{\epsilon}_{t-i} = \sum_{i=0}^{\infty} (\Phi_i L)(L^{-1} \boldsymbol{\epsilon}_{t-i}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma) \\ &= \sum_{i=0}^{\infty} (\Phi_h L) \boldsymbol{\eta}_{t-i} \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, D) \end{aligned} \tag{15}$$

where

$$D = L^{-1} \cdot \Sigma \cdot (L')^{-1} = \left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & S \end{array}\right], \qquad \Phi_0 = I, \quad \Phi_i = B_4' \Phi_{i-1}, \tag{16}$$

and A is a square matrix of size k equal to the number of entries to be shocked.

Single layer network - block OIRF

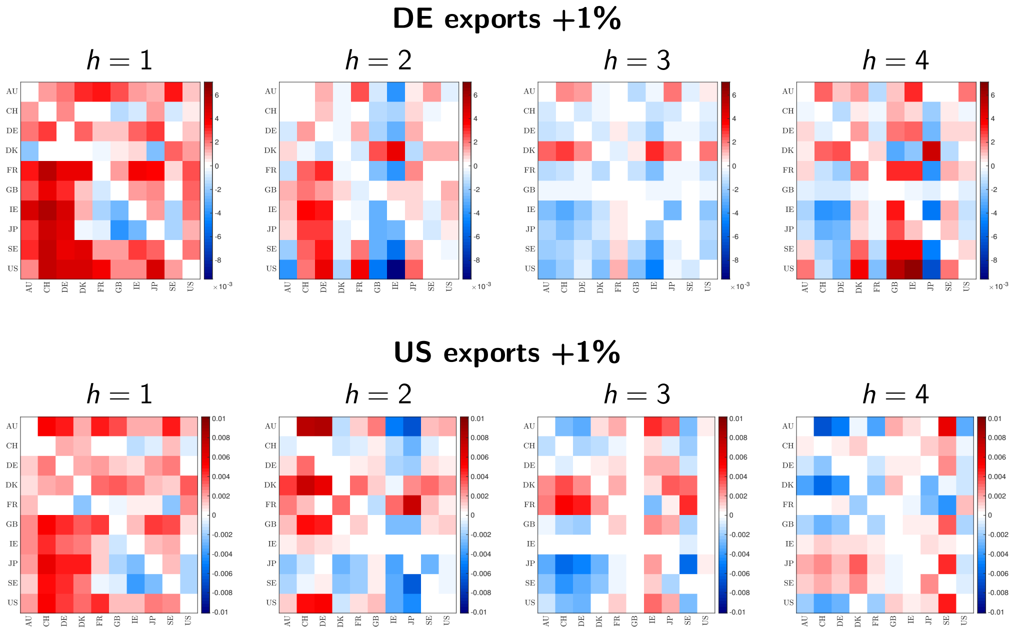
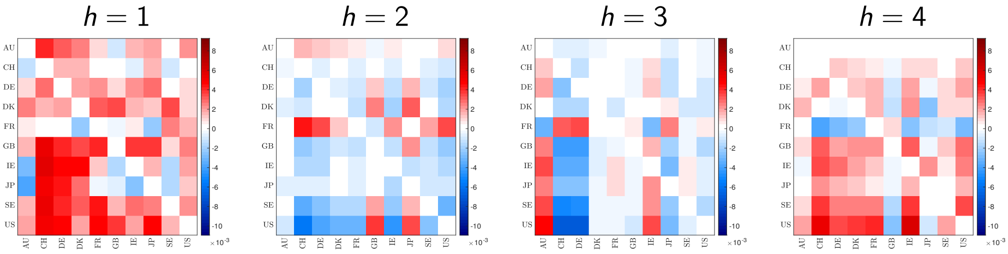


Figure: Positive effects in red, negative effects in blue.

Single layer network - block OIRF

UK exports +1%



DE imports -1%

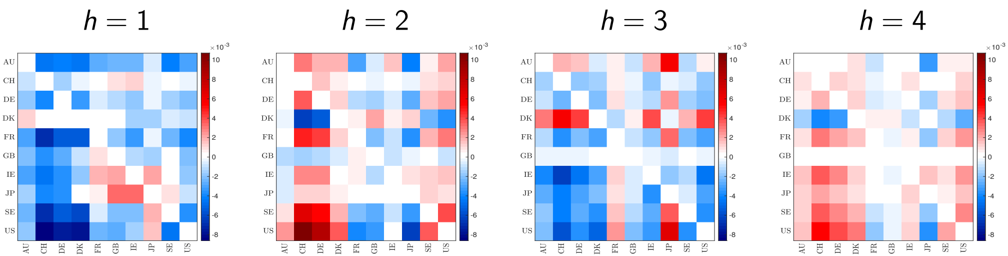


Figure: Positive effects in red, negative effects in blue.

Single layer network - IRF analysis

Comments on positive shock to US,DE,UK exports

- pos shock to US exports more effective on the network (higher average magnitude) than to DE or UK
- all cases: overall positive effect on network ⇒ stimulus to international trade
- all cases: immediate boost to imports of Switzerland, Germany and Austria

Comments on negative shock DE imports

- overall negative effect on international trade
- one lag - mostly affected: imports Austria, Switzerland, Germany and France
- more lags: alternating sign decay

▶ shock persistence ⇒ slow decay in all cases (similar decay pattern)

Application II: COMTRADE & BIS Multi-Layer Networks

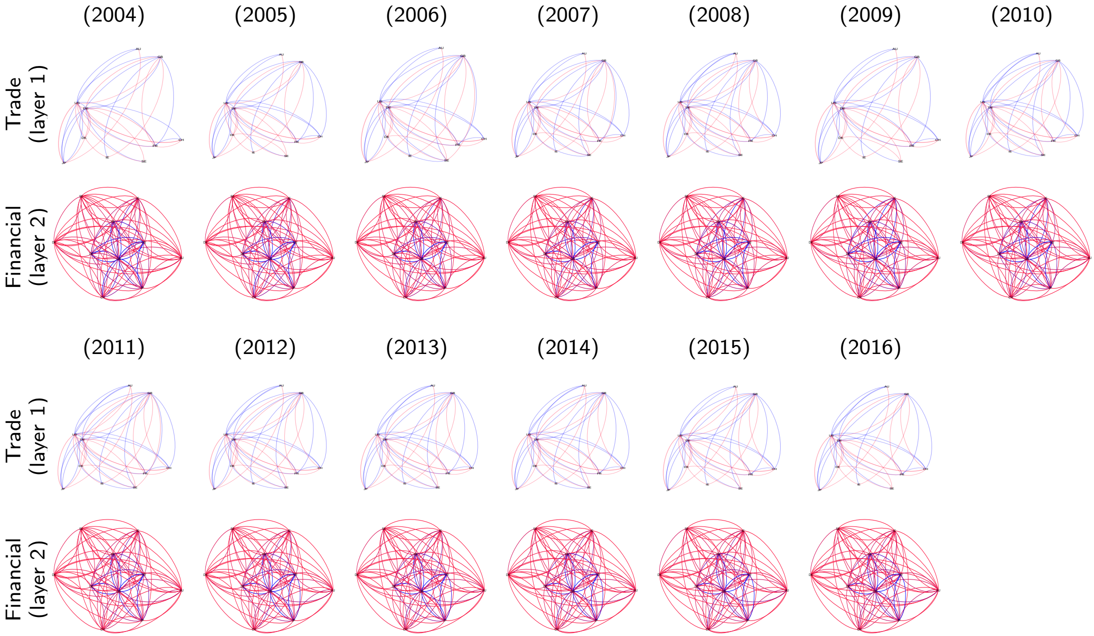


Figure: International trade and financial networks. Nodes: countries. Edges: flows.

Empirical Application - multi-layer networks

Tensor autoregressive model ART(1)

$$\mathcal{Y}_t = \mathcal{B} \times_4 \text{vec}(\mathcal{Y}_{t-1}) + \mathcal{E}_t, \qquad \mathcal{E}_t \sim \mathcal{N}_{10,10,2}(\mathbf{0}, \Sigma_1, \Sigma_2, \Sigma_3) \tag{17}$$

Parameters

unrestricted VAR(1)

$$\prod_{j=1}^{N+1} I_j + \frac{1}{2} \prod_{j=1}^{N} I_j \left(\prod_{j=1}^{N} I_j + 1 \right)$$

ART(1) with PARAFAC(R)

$$R \sum_{j=1}^{N+1} I_j + \frac{1}{2} \sum_{j=1}^{N} I_j (I_j + 1)$$

Empirical Application - multi-layer networks

Tensor autoregressive model ART(1)

$$\mathcal{Y}_t = \mathcal{B} \times_4 \text{vec}(\mathcal{Y}_{t-1}) + \mathcal{E}_t, \qquad \mathcal{E}_t \sim \mathcal{N}_{10,10,2}(\mathbf{0}, \Sigma_1, \Sigma_2, \Sigma_3) \tag{17}$$

- mode-4 matricized:
  $$B_4' = \Big[\text{vec}(\mathcal{B}_{::1,1}), \text{vec}(\mathcal{B}_{::2,1}), \ldots, \ldots, \text{vec}(\mathcal{B}_{::1,200}), \text{vec}(\mathcal{B}_{::2,200})\Big]$$
- entry (i, j) of B_4':
  impact of edge j $[t-1] \rightarrow i$ $[t]$

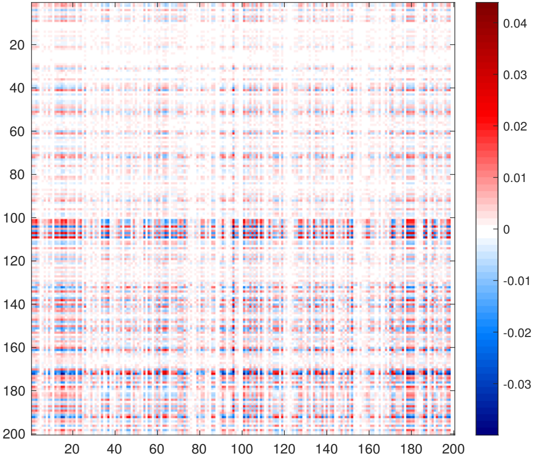


Figure: Estimated $\hat{B}_4'$.

Empirical Application - multi-layer networks

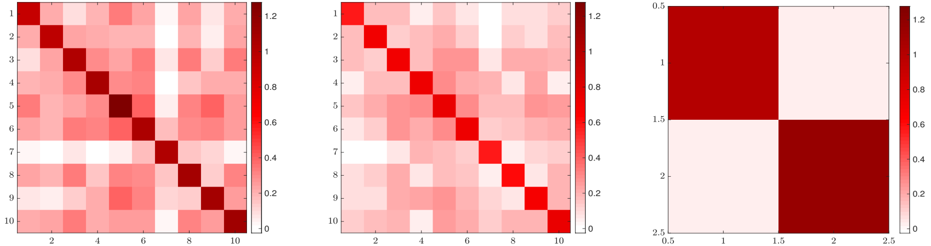


Figure: Estimated covariance matrices: $\hat{\Sigma}_1$ (*left*), $\hat{\Sigma}_2$ (*center*), $\hat{\Sigma}_3$ (*right*).

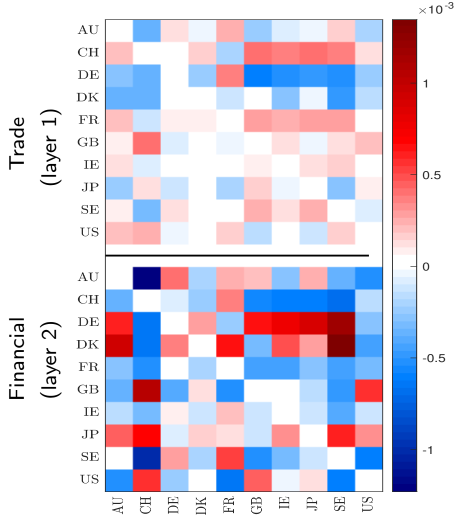
- higher values for individual variances
- mostly positive correlations

Impulse Response: US import

- overall slightly negative effect on both layers (trade and financial) of the network
- reaction of the financial layer is higher in magnitude ⇒ higher responsiveness of capital flows w.r.t. trade goods flows
- most affected real goods transactions are between Switzerland, Germany and France (the exporters) vis-á-vis UK, Ireland, Sweden and Japan (the importers)
- same relation occurs on the financial layer of the network, with opposite sign and greater magnitude
- proposed interpretation: kind of "substitution effect"
- fast decay

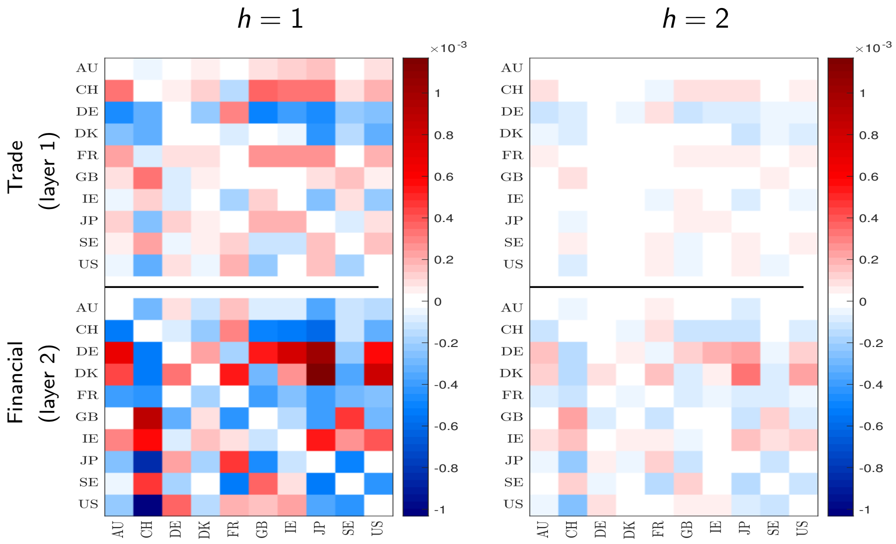
Shock to US **imports: -1%**

$h = 1$



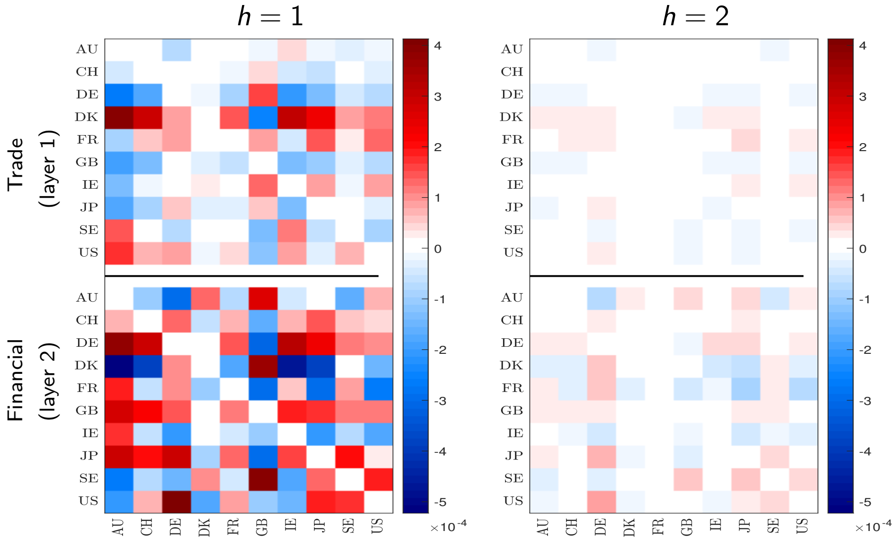
Impulse Response: UK financial flows

Shock to GB **capital inflows: -1%**



Impulse Response: UK financial flows

Shock to GB **capital inflows: -1%** and **outflows +1%**



shock capital inflows

- overall slightly negative effect on the capital (in- and out-) flows between the countries
- Austria and Japan (among the top capital exporters) ⇒ overall reduction of capital outflows
- Ireland and Germany (among the least capital exporting countries) ⇒ positive effect on outflows
- substitution effect between Switzerland and Germany
- trade layer: overall positive effect, with smaller magnitude than that on the financial layer

shock capital inflows + outflows

- one lag: positive average impact on capital flows, both in- and out- (in particular, Japan, UK, Switzerland and Denmark)
- impact on Denmark and Germany ⇒ moving in opposite directions, both on from the financial and the commercial (similar in previous case)
- overall total impact of shock is greater than in the previous two situations ⇒ due to the magnitude of the shock
- increase in UK capital outflows ⇒ overall positive cascade effect (stimulates the outflows from other countries). Impact on trade network is smaller

▶ Both cases: persistence of a financial shock greater than that of trade shock

Conclusions

Proposal: linear, dynamic tensor regression model

- generalises linear regression models to multi-dimensional regression
- PARAFAC tensor decomposition for parsimony
- hierarchical global-local shrinkage prior for sparse coefficients
- good performance against synthetic data up to 50×50

- application to COMTRADE network (**matrix AR(1)** model):
  - ✓ impact of trade links is heterogeneous and sparse
  - ✓ heterogeneous magnitude and persistence of shock propagation
  - ✓ role of network topology in shock propagation
- application to COMTRADE+BIS 2-layer networks (**tensor AR(1)** model):
  - ✓ impact of trade and financial links are heterogeneous and sparse
  - ✓ financial shock propagation has higher magnitude
  - ✓ block-orthogonal tensor IRF
  - ✓ within + between layer shock propagation
  - ✓ meaningful country-specific IRF results

References i

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

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

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

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

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

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

Cichocki, A., Lee, N., Oseledets, I., Phan, A., Zhao, Q., and Mandic, D. (2016). Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends in Machine Learning*, 9(4-5):249–429.

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

Harrison, J. and West, M. (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.

References ii

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

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

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

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

Li, L. and Zhang, X. (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.

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.

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

Springer, M. and Thompson, W. (1970). The distribution of products of beta, gamma and Gaussian random variables. *SIAM Journal on Applied Mathematics*, 18(4):721–737.

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

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

References iii

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

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

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

Proof of Proposition 2 - Part 1/3.

Denote with L the lag operator, s.t. $L\mathcal{Y}_t = \mathcal{Y}_{t-1}$, by properties of the contracted product in Lemma 3, 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

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

Proof of Proposition 2 - Part 2/3.

From the assumption $\mathcal{E}_t \overset{iid}{\sim} \mathcal{N}_{I_1,\dots,I_N}(\mathcal{O},\Sigma_1,\dots,\Sigma_N)$, 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 3, 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

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

$$= \mathbb{E}\Big(\big(\sum_{k=0}^{\infty} \mathcal{A}^{k+h} \bar{\times}_N \mathcal{E}_{t-k-h}\big) \bar{\times}_1 \big(\sum_{k=0}^{\infty} \mathcal{E}_{t-k-h}^T \bar{\times}_N (\mathcal{A}^T)^k\big)\Big),$$

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

Proof of Proposition 2 - Part 3/3.

Using $\mathbb{E}(\mathcal{E}_t) = \mathcal{O}$ and linearity of expectation and of the contracted product we get

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

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.

Lemma 3 (Properties of contracted product).

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* $\mathcal{X} \bar{\times}_N \mathcal{Y} = \langle \mathcal{X}, \mathcal{Y} \rangle = \mathrm{vec}(\mathcal{X})' \cdot \mathrm{vec}(\mathcal{Y})$.

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

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

(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*

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

(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* $\underline{\mathcal{Y}}^T_{j_1,\ldots,j_M,j_{M+1}} = \underline{\mathcal{Y}}_{j_{M+1},j_M,\ldots,j_1}$.

Lemma 4 (Relation ART(p) and ART(1)).

Every $(I_1 \times I_2 \times \ldots \times I_N)$*-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 $(pI_1 \times I_2 \times \ldots \times I_N)$*-dimensional ART(1) process*

$$\underline{\mathcal{Y}}_t = \underline{\mathcal{A}} \bar{\times}_N \underline{\mathcal{Y}}_{t-1} + \underline{\mathcal{E}}_t.$$

Proof of Lemma 4.

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 $\underline{\mathcal{E}}_t$, for $k = 0, \ldots, p$, as

$$\underline{\mathcal{Y}}_{(k-1)I_1+1:kI_1,:,\ldots,:,t} = \mathcal{Y}_{t-k}, \qquad \underline{\mathcal{E}}_{(k-1)I_1+1:kI_1,:,\ldots,:,t} = \mathcal{E}_{t-k}.$$

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

$$\begin{aligned} \underline{\mathcal{A}}_{(1:I_1,:,\ldots,:,(k-1)I_1+1:kI_1,:,\ldots,:} &= \mathcal{A}_k \qquad k = 1, \ldots, p \\ \underline{\mathcal{A}}_{(kI_1+1:(k+1)I_1,:,\ldots,:,(k-1)I_1+1:kI_1,:,\ldots,:} &= \mathcal{I} \qquad k = 1, \ldots, p-1, \end{aligned}$$

and 0 elsewhere. 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

$$\underline{\mathcal{Y}}_t = \underline{\mathcal{A}} \bar{\times}_N \underline{\mathcal{Y}}_{t-1} + \underline{\mathcal{E}}_t.$$

Proof of Proposition 3

Proof of Proposition 3.

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 $\text{mat}_{1:N,1:N}$ 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 $A\mathbf{x} = \widetilde{\lambda}\mathbf{x}$, where $A = \text{mat}_{1:N,1:N}(\mathcal{A})$. In particular, $\lambda = \widetilde{\lambda}$ and $\mathbf{x} = \text{vec}(\mathcal{X})$.

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 $p \geq 1$.

Properties of ART(1) - Impulse Response Function

Orthogonalised IRF requires **orthogonal** shocks:

$$IRF_h = \mathbb{E}[\mathcal{Y}_{t+h} | \widetilde{E}_{ij,t} = \delta_{ij}, \widetilde{E}_{-ij,t} = 0, \mathcal{F}_{t-1}] - \mathbb{E}[\mathcal{Y}_{t+h} | \widetilde{E}_t = 0, \mathcal{F}_{t-1}] \tag{18}$$

- covariance restrictions for avoiding/mitigating *compositional effect* (due to contemporaneous correlations)
- Cholesky $\Rightarrow$ not invariant to *ordering* of variables; not unique

Generalised IRF (Koop et al. (1996), Pesaran and Shin (1998)):

$$GIRF_h = \mathbb{E}[\mathcal{Y}_{t+h} | \widetilde{E}_{ij,t} = \delta_{ij}, \mathcal{F}_{t-1}] - \mathbb{E}[\mathcal{Y}_{t+h} | \mathcal{F}_{t-1}] \tag{19}$$

- unique and invariant to ordering of variables
- no covariance restrictions: when one variable is shocked, other variables also vary, then average by integrating out all other shocks
- not distinguish causes of a change in $\mathbb{E}[\mathcal{Y}_{t+h} | \mathcal{F}_{t-1}]$

Prior for entry of tensor $\mathcal{B}$ back

The joint distribution of PARAFAC marginal entries is $\prod_{h=1}^{4} \pi(\beta_{h,i}^{(r)} | \tau, \phi_r, w_{h,r})$. To obtain the conditional prior distribution for tensor entry b_{ijkp} we apply:

Theorem 5 (4 in Springer and Thompson (1970)).

The probability density function of the product $z = \prod_{j=1}^{J} x_j$ of J independent Normal random variables $x_j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, \ldots, J$, is a Meijer G-function multiplied by a normalising constant H:

$$p(z|0, \{\sigma_j^2\}_{j=1}^{J}) = H \cdot G_{J,0}^{J,0}\left(z^2 \cdot \prod_{j=1}^{J} \frac{1}{2\sigma_j} \middle| \mathbf{0}\right), \tag{20}$$

where

$$H = (2\pi)^{-J/2} \cdot \prod_{j=1}^{J} \sigma_j^{-1} \tag{21}$$

and $G_{p,q}^{m,n}(\cdot|\cdot)$ is a Meijer G-function ($c \in \mathbb{R}, s \in \mathbb{C}$, integral along vertical line in the complex plane):

$$G_{p,q}^{m,n}\left(z \middle| \begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}\right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} z^{-s} \frac{\prod_{j=1}^{m} \Gamma(s + b_j) \cdot \prod_{j=1}^{n} \Gamma(1 - a_j - s)}{\prod_{j=n+1}^{p} \Gamma(s + a_j) \cdot \prod_{j=m+1}^{q} \Gamma(1 - b_j - s)} \, \mathrm{d}s \,. \tag{22}$$

Prior for entry of tensor $\mathcal{B}$

Lemma 6.

Define $\beta_r = \beta_{1,i}^{(r)} \cdot \beta_{2,j}^{(r)} \cdot \beta_{3,k}^{(r)} \cdot \beta_{4,p}^{(r)}$. *Under the prior specification, by Theorem 5 we have the conditional prior distribution:*

$$\pi(b_{ijkp}|\tau,\phi,\mathbf{W}) = p\left(\sum_{r=1}^{R}\beta_r|-\right) = \sum_{r=1}^{R} p(\beta_r|-) = \sum_{r=1}^{R} H_r \cdot G_{4,0}^{4,0}\left(\beta_r^2 \cdot \prod_{h=1}^{4}\frac{1}{2\tau\phi_r}W_{h,r}^{-1}\middle|\mathbf{0}\right)$$
$$\propto \sum_{r=1}^{R} G_{4,0}^{4,0}\left(\beta_r^2 \cdot \prod_{h=1}^{4}\frac{1}{2\tau\phi_r}W_{h,r}^{-1}\middle|\mathbf{0}\right), \tag{23}$$

with:

$$G_{4,0}^{4,0}\left(\beta_r^2 \cdot \prod_{h=1}^{4}\frac{1}{2\tau\phi_r}W_{h,r}^{-1}\middle|\mathbf{0}\right) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left(\frac{\beta_r^2}{(2\tau\phi_r)^4}\prod_{h=1}^{4}W_{h,r}^{-1}\right)^{-s}\mathrm{d}s\,, \tag{24}$$

$$H_r = (2\pi)^{-2}\cdot(\tau\phi_r)^{-4}\prod_{j=1}^{4}W_{h,r}^{-1}\,. \tag{25}$$

Thus, the marginal prior distribution is: $\pi(b_{ijkp}) = \int \pi(b_{ijkp}|\tau,\phi,\mathbf{W})\pi(\tau)\pi(\phi)\pi(\mathbf{W})\,\mathrm{d}\tau\,\mathrm{d}\phi\,\mathrm{d}\mathbf{W}$.

Prior for entry of tensor $\mathcal{B}$

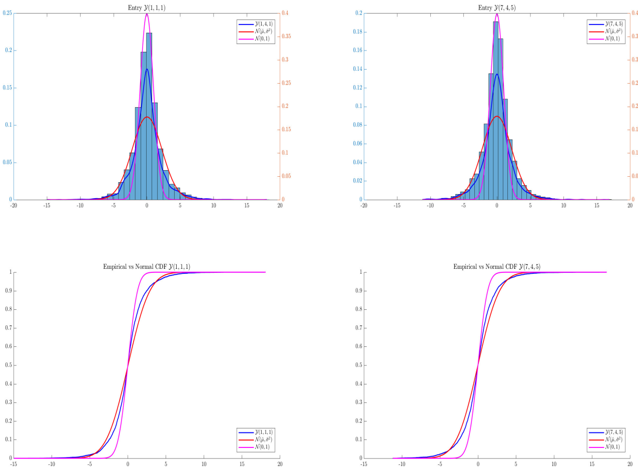


Figure: Simulated distribution of two entries of tensor (std Normal marginals) vs Normal with same mean and variance vs std Normal, for $R = 5$.

Prior for entry of tensor $\mathcal{B}$ back

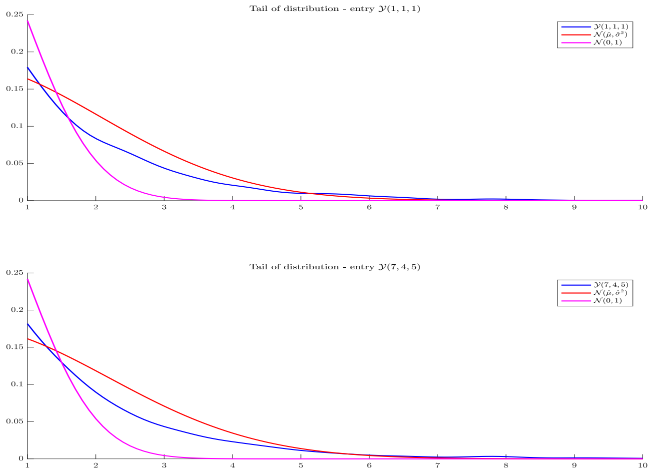


Figure: Simulated distribution (tail) of two entries of tensor (std Normal marginals) vs Normal with same mean and variance vs std Normal, for $R = 5$.

Prior for entry of tensor $\mathcal{B}$

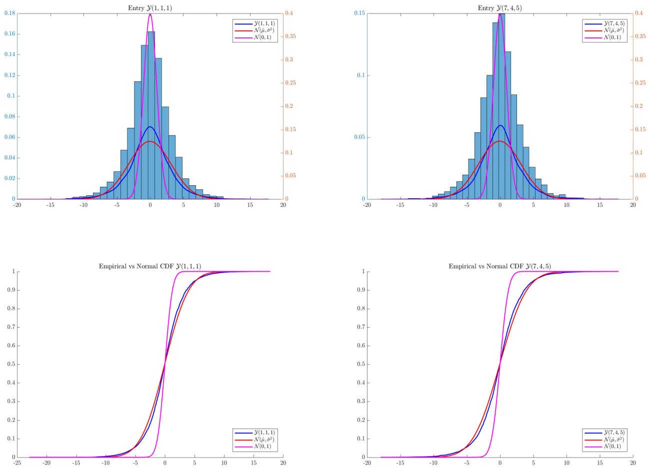


Figure: Simulated distribution of two entries of tensor (std Normal marginals) vs Normal with same mean and variance vs std Normal, for $R = 10$.

Prior for entry of tensor $\mathcal{B}$ back

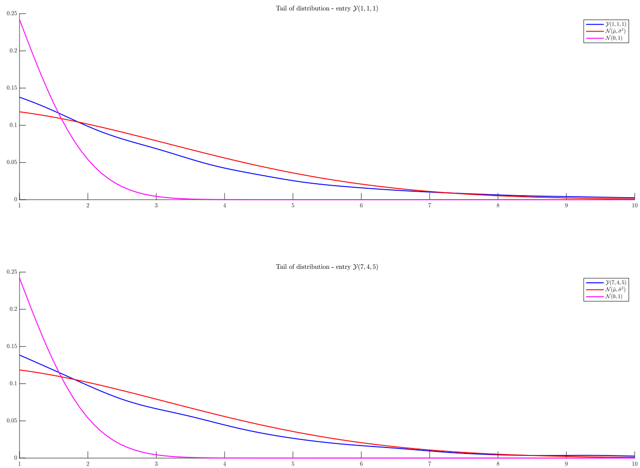


Figure: Simulated distribution (tail) of two entries of tensor (std Normal marginals) vs Normal with same mean and variance vs std Normal, for $R = 10$.

Initialisation of Gibbs sampler

Gibbs sampler sensitive to initial value of some key parameters:

- tensor PARAFAC marginals $\{\beta_j^{(r)}\}_{j,r}$ initialised via Simulated Annealing;
  Intuition: find the set of marginals generating a sufficiently sparse tensor, while allowing deviations from zero.
- other parameters initialised from prior distribution;
  Intuition: sampler not very sensitive to their starting value.

Initialisation marginals $\beta_j^{(r)}$ - back

Intuition: find the set of marginals generating a tensor with many entries close to zero, others far zero. Use Simulated Annealing for minimising the objective function:

$$f(\tilde{\mathcal{B}}^{(n)}) = \psi_0 \left\| \tilde{\mathcal{B}}^{(n)} \right\|_2 + \psi_3 \sum_{r=1}^{R} \left\| \tilde{\beta}_3^{(r),(n)} \right\|_2 . \tag{26}$$

Penalties:

- $\psi_0 > 0 \to$ tensor quadratic norm;
- $\psi_3 > 0 \to$ quadratic norm 3-order marginals.

Cooling schedule, with fixed $q > 0$:

$$C(n) = \frac{q}{1 + \log(n)} \qquad n = 1, \ldots, N_{SA} . \tag{27}$$