

1 **Web-based supporting materials for “Bayesian**
2 **Dynamic Tensor Regression” by M. Billio, R.**
3 **Casarini, M. Iacobini and S. Kaufmann**

4 Monica Billio

5 *Ca’ Foscari University of Venice, Italy*

6 Roberto Casarin[†]

7 *Ca’ Foscari University of Venice, Italy*

8 E-mail: r.casarin@unive.it

9 Matteo Iacobini

10 *Scuola Normale Superiore of Pisa, Italy*

11 Sylvia Kaufmann

12 *Study Center Gerzensee, Foundation of the Swiss National Bank, Switzerland*

13 **Supplementary material.** This appendix contains background results on
14 tensors in Section S.1, and the derivation of the tensor forecast error variance
15 decomposition in Section S.2. An example of MAR in Section S.3 and proofs of
16 the remarks in Section 2 of the main paper in Section S.4. Details on the prior on
17 tensor entries are given in Section S.5. Also, Section S.6 reports the details on
18 posterior computation and Section S.7 describes the initialisation of the inferential
19 algorithm. A summary of simulation results is provided in Section S.8. The data
20 used in the empirical application is described in Section S.9, and further plots of
21 the estimation results are given in Section S.10.

22 **S.1. Background Material on Tensor Calculus**

23 This section provides additional results on tensor operators.

Let \mathcal{X}, \mathcal{Y} be two $(I_1 \times \dots \times I_N)$ -dimensional tensors. The *Hadamard product* between them, $\mathcal{Z} = \mathcal{X} \odot \mathcal{Y}$, is the $(I_1 \times \dots \times I_N)$ -dimensional tensor \mathcal{Z} defined by the element-wise multiplication

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

We introduce two multilinear operators acting on tensors (see [Kolda \(2006\)](#) for more details).

[†]*Address for correspondence:* Roberto Casarin, Department of Economics, Ca’ Foscari University of Venice, Venice, 30123, Italy

DEFINITION S.1.1 (TUCKER OPERATOR). Let $\mathcal{Y} \in \mathbb{R}^{J_1 \times \dots \times J_N}$ and $\mathbf{N} = \{1, \dots, N\}$. Let $(A_n)_n$ be a collection of N matrices such that $A_n \in \mathbb{R}^{I_n \times J_n}$. The Tucker operator is defined as

$$\llbracket \mathcal{Y}; A_1, \dots, A_N \rrbracket = \mathcal{Y} \bar{\times}_1 A_1 \bar{\times}_1 A_2 \dots \bar{\times}_1 A_N,$$

²⁶ and the resulting tensor has size $I_1 \times \dots \times I_N$.

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

$$\mathcal{B} = \mathcal{G} \bar{\times}_1 A^{(1)} \bar{\times}_1 \dots \bar{\times}_1 A^{(N)} = \sum_{i_1=1}^{g_1} \dots \sum_{i_N=1}^{g_N} \mathcal{G}_{i_1, \dots, i_N} \mathbf{a}_{i_1}^{(1)} \circ \dots \circ \mathbf{a}_{i_N}^{(N)} \quad (\text{S1})$$

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

$$\mathcal{B}_{j_1, \dots, j_N} = \sum_{i_1=1}^{g_1} \dots \sum_{i_N=1}^{g_N} \mathcal{G}_{i_1, \dots, i_N} \cdot A_{i_1, j_1}^{(1)} \dots A_{i_N, j_N}^{(N)} \quad (\text{S2})$$

The PARAFAC(R) decomposition[†], is rank- R decomposition which represents a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ as a finite sum of R rank-1 tensors obtained as the outer products of N vectors (called marginals) $\boldsymbol{\beta}_j^{(r)} \in \mathbb{R}^{I_j}$, $j = 1, \dots, J$

$$\mathcal{B} = \sum_{r=1}^R \mathcal{B}_r = \sum_{r=1}^R \boldsymbol{\beta}_1^{(r)} \circ \dots \circ \boldsymbol{\beta}_J^{(r)}. \quad (\text{S3})$$

²⁷ Fig. 1 provides a graphical representation of this decomposition for a 3-order
²⁸ tensor.

DEFINITION S.1.2 (KRUSKAL OPERATOR). Let $\mathbf{N} = \{1, \dots, N\}$ and $(A_n)_n$ be a collection of N matrices such that $A_n \in \mathbb{R}^{I_n \times R}$ for $n \in \mathbf{N}$. Let \mathcal{I} be the identity tensor of size $R \times \dots \times R$, i.e. a tensor having ones along the superdiagonal and zeros elsewhere. The Kruskal operator is defined as

$$\mathcal{X} = \llbracket A_1, \dots, A_N \rrbracket = \llbracket \mathcal{I}; A_1, \dots, A_N \rrbracket,$$

with \mathcal{X} a tensor of size $I_1 \times \dots \times I_N$. An alternative representation is obtained

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

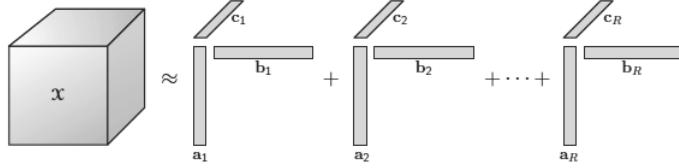


Figure 1: PARAFAC decomposition of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with $\mathbf{a}_r \in \mathbb{R}^{I_1}$, $\mathbf{b}_r \in \mathbb{R}^{I_2}$ and $\mathbf{c}_r \in \mathbb{R}^{I_3}$, $r = 1, \dots, R$. Figure from Kolda and Bader (2009).

by defining $\mathbf{a}_n^{(r)}$ the r -th column of the matrix A_n and using the outer product

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

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

$$\begin{aligned} \mathbf{X}_{(n)} &= A_n (A_N \odot^K \dots \odot^K A_{n+1} \odot^K A_{n-1} \odot^K \dots \odot^K A_1)' \\ \text{vec}(\mathcal{X}) &= (A_N \odot^K \dots \odot^K A_1) \mathbf{1}_R \end{aligned}$$

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

REMARK S.1.1. Let \mathcal{X} be a N -order tensor of dimensions $I_1 \times \dots \times I_N$ and let $I^* = \prod_{i=1}^N I_i$. Then there exists a $I^* \times I^*$ vec-permutation (or commutation) matrix $K_{1 \rightarrow n}$ such that

$$K_{1 \rightarrow n} \text{vec}(\mathcal{X}) = K_{1 \rightarrow n} \text{vec}(\mathbf{X}_{(1)}) = \text{vec}(\mathbf{X}_{(n)}) = \text{vec}(\mathbf{X}_{(1)}^{T_\sigma}) = \text{vec}(\mathcal{X}^{T_\sigma}),$$

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

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

LEMMA S.1.1 (TENSOR – MATRIX NORMAL). Let \mathcal{X} be a N -order random tensor with dimensions I_1, \dots, I_N and let $\mathbf{N} = \{1, \dots, N\}$ be partitioned by the index sets $\mathcal{R} = \{r_1, \dots, r_m\} \subset \mathbf{D}$ and $\mathcal{C} = \{c_1, \dots, c_p\} \subset \mathbf{N}$, i.e. $\mathbf{N} = \mathcal{R} \cup \mathcal{C}$,

$\mathcal{R} \cap \mathcal{C} = \emptyset$ and $N = m + p$. Then

$$\mathcal{X} \sim \mathcal{N}_{I_1, \dots, I_N}(\mathcal{M}, \Sigma_1, \dots, \Sigma_N) \iff \mathbf{X}_{(\mathcal{R} \times \mathcal{C})} \sim \mathcal{N}_{m,p}(\mathbf{M}_{(\mathcal{R} \times \mathcal{C})}, \Sigma_1, \Sigma_2),$$

³⁰ with $\Sigma_1 = \Sigma_{r_m} \otimes \dots \otimes \Sigma_{r_1}$ and $\Sigma_2 = \Sigma_{c_p} \otimes \dots \otimes \Sigma_{c_1}$.

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

$$(2\pi)^{-\frac{\prod_i I_i}{2}} |\Sigma_1|^{-\frac{I_{-1}}{2}} \cdots |\Sigma_n|^{-\frac{I_{-n}}{2}} \cdots |\Sigma_N|^{-\frac{I_{-N}}{2}} = \quad (\text{S4})$$

$$\begin{aligned} &= (2\pi)^{-\frac{\prod_i I_i}{2}} |\Sigma_1|^{-\frac{I_{-1}}{2}} \cdots |\Sigma_{n-1}|^{-\frac{I_{-(n-1)}}{2}} |\Sigma_{n+1}|^{-\frac{I_{-(n+1)}}{2}} \cdots |\Sigma_N|^{-\frac{I_{-N}}{2}} |\Sigma_n|^{-\frac{I_{-n}}{2}} \\ &= (2\pi)^{-\frac{\prod_i I_i}{2}} |\Sigma_N \otimes \dots \otimes \Sigma_{n-1} \otimes \Sigma_{n+1} \otimes \dots \otimes \Sigma_N|^{-\frac{n}{2}} |\Sigma_n|^{-\frac{I_{-n}}{2}}. \end{aligned} \quad (\text{S5})$$

Concerning the exponent, let $\mathbf{i} = (i_1, \dots, i_N)$ and, for ease of notation, define $\mathcal{Y} = \mathcal{X} - \mathcal{M}$ and $\mathcal{U} = (\Sigma_N^{-1} \circ \dots \circ \Sigma_1^{-1})$. By the definition of contracted and outer products, it holds

$$\mathcal{Y} \bar{\times}_N \mathcal{U} \bar{\times}_N \mathcal{Y} = \sum_{i_1, \dots, i_n, \dots, i_N} \sum_{i'_1, \dots, i'_n, \dots, i'_N} y_{i_1, \dots, i_N} (u_{i_1, i'_1}^{-1} \cdots u_{i_n, i'_n}^{-1} \cdots u_{i_N, i'_N}^{-1}) y_{i'_1, \dots, i'_n, \dots, i'_N}. \quad (\text{S6})$$

Define $\mathbf{j} = \sigma(\mathbf{i})$, where σ is the permutation defined in Remark S.1.1 exchanging i_1 with i_n , $n \in \{2, \dots, N\}$. Then the previous equation can be rewritten as

$$\begin{aligned} \mathcal{Y} \bar{\times}_N \mathcal{U} \bar{\times}_N \mathcal{Y} &= \sum_{j_1, \dots, j_N} \sum_{j'_1, \dots, j'_N} y_{j_n, \dots, j_1, \dots, i_N} (u_{j_n, j'_n}^{-1} \cdots u_{j_1, j'_1}^{-1} \cdots u_{i_N, i'_N}^{-1}) y_{j'_n, \dots, j'_1, \dots, i'_N} \\ &= \mathcal{Y}^\sigma \bar{\times}_N (\Sigma_1^{-1} \circ \dots \circ \Sigma_N^{-1})^\sigma \bar{\times}_N \mathcal{Y}^\sigma \end{aligned}$$

where \mathcal{Y}^σ is the transpose tensor of \mathcal{Y} (see Pan (2014)) obtained by permuting the first and the n -th modes and similarly for the N -order tensor $(\Sigma_1^{-1} \circ \dots \circ \Sigma_N^{-1})^\sigma$. Let $(\mathcal{S}_1, \mathcal{S}_2)$, with $\mathcal{S}_1 = \{1, \dots, N\}$ and $\mathcal{S}_2 = \{N+1, \dots, 2N\}$, be a partition of $\{1, \dots, 2N\}$. By vectorizing eq. (S6) and exploiting the results in Lemma A.1 and Lemma A.2, we have

$$\begin{aligned} \mathcal{Y} \bar{\times}_N \mathcal{U} \bar{\times}_N \mathcal{Y} &= \text{vec}(\mathcal{Y})' \cdot \mathcal{U}_{(\mathcal{S}_1, \mathcal{S}_2)} \cdot \text{vec}(\mathcal{Y}) \\ &= \text{vec}(\mathcal{Y})' \cdot (\Sigma_N^{-1} \otimes \dots \otimes \Sigma_n^{-1} \otimes \dots \otimes \Sigma_1^{-1}) \cdot \text{vec}(\mathcal{Y}) \\ &= \text{vec}(\mathcal{Y}^\sigma)' \cdot (\Sigma_N^{-1} \otimes \dots \otimes \Sigma_1^{-1} \otimes \Sigma_n^{-1}) \cdot \text{vec}(\mathcal{Y}^\sigma) \\ &= \text{vec}(\mathbf{Y}_{(n)})' \cdot (\Sigma_N^{-1} \otimes \dots \otimes \Sigma_1^{-1} \otimes \Sigma_n^{-1}) \cdot \text{vec}(\mathbf{Y}_{(n)}) \end{aligned} \quad (\text{S7})$$

$$\begin{aligned}
&= \text{vec}(\mathbf{Y}_{(n)})' \cdot \text{vec}(\Sigma_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot (\Sigma_N^{-1} \otimes \dots \otimes \Sigma_1^{-1})) \\
&= \text{tr}(\mathbf{Y}'_{(n)} \cdot \Sigma_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot (\Sigma_N^{-1} \otimes \dots \otimes \Sigma_1^{-1})) \\
&= \text{tr}((\Sigma_N^{-1} \otimes \dots \otimes \Sigma_1^{-1})(\mathbf{X}_{(n)} - \mathbf{M}_{(n)})' \Sigma_n^{-1} (\mathbf{X}_{(n)} - \mathbf{M}_{(n)})). \quad (\text{S8})
\end{aligned}$$

Since the term in (S4) and (S7) are the normalizing constant and the exponent of the tensor normal distribution, whereas (S5) and (S8) are the corresponding expressions for the desired matrix normal distribution, the result is proved for the case $\#\mathcal{R} = 1$. In the general case $\#\mathcal{R} = r > 1$ the proof follows from the same reasoning, by substituting the permutation σ with another permutation σ' which exchanges the modes of the tensor such that the first r modes of the transpose tensor $\mathcal{Y}^{\sigma'}$ correspond to the elements of \mathcal{R} .

38 S.2. Forecast error variance decomposition

From the results in eqs. (11)-(12), we obtain the forecast error variance decomposition (tFEVD) for the tensor autoregressive model in each of the two cases. The tFEVD $\theta_{i,j}(h)$ measures the proportion of the h -step ahead forecast error variance of variable i that is accounted for by the innovations in variable j , in the VAR formulation of the model. Recently, Lanne and Nyberg (2016) have introduced a modification to the FEVD obtained from the GIRF of Koop et al. (1996), $\theta_{i,j}^*(h)$, which has unit sum. Denoting by $IRF(h)$ an impulse response function at horizon h , the corresponding tFEVD and its modification are, respectively,

$$\theta_{ij}(h) = \frac{\sum_{k=0}^h IRF_{ij}^2(k)}{\sum_{k=0}^h \sum_{j=0}^{I^*} IRF_{ij}^2(k)}, \quad \theta_{i,j}^*(h) = \frac{\sum_{k=0}^h (\psi_{ij}^G(k; n))^2}{\sum_{k=0}^h \sum_{j=0}^{I^*} (\psi_{ij}^G(k; n))^2}.$$

The orthogonalised tensor forecast error variance decomposition (OtFEVD) by construction sums (over j) to 1. In this case $\delta_j^* = 1$, and all the other $I^* - 1$ entries are zero (equivalent to $\boldsymbol{\delta}^* = \mathbf{e}_j$). The OtFEVD is given by

$$\theta_{i,j}^O(h) = \frac{\sum_{k=0}^h (\psi_{ij}^O(k; n))^2}{\sum_{k=0}^h \sum_{j=0}^{I^*} (\psi_{ij}^O(k; n))^2} = \frac{\sum_{k=0}^h (\mathbf{e}'_i \Psi_k L P \mathbf{e}_j)^2}{\sum_{k=0}^h \mathbf{e}'_i (\Psi_k L) D(\Psi_k L)' \mathbf{e}_i}.$$

Consider the case $\delta_j^* = \sqrt{D_{jj}}$, with all the other $I^* - 1$ entries being zero (equivalent to $\boldsymbol{\delta}^* = \sqrt{D_{jj}} \mathbf{e}_j$). The generalised tensor forecast error variance decomposition (GtFEVD) does not sum to 1, and is

$$\theta_{i,j}^G(h) = \frac{\sum_{k=0}^h (\psi_{ij}^G(k; n))^2}{\sum_{k=0}^h \sum_{j=0}^{I^*} (\psi_{ij}^G(k; n))^2} = \frac{\sum_{k=0}^h (\mathbf{e}'_i \Psi_k L D D_{jj}^{-1/2} \mathbf{e}_j)^2}{\sum_{k=0}^h \mathbf{e}'_i (\Psi_k L) D(\Psi_k L)' \mathbf{e}_i} \quad (\text{S9})$$

Finally, the modified tFEVD applied to the tensor GIRF (S9) yields

$$\theta_{i,j}^{G*}(h) = \frac{\sum_{k=0}^h (\mathbf{e}'_i \Psi_k L D D_{jj}^{-1/2} \mathbf{e}_j)^2}{\sum_{k=0}^h \sum_{j=1}^{I^*} \mathbf{e}'_i \Psi_k L D D_{jj}^{-1/2} \mathbf{e}_j} = \frac{\sum_{k=0}^h (\mathbf{e}'_i \Psi_k L D D_{jj}^{-1/2} \mathbf{e}_j)^2}{\sum_{k=0}^h \mathbf{e}'_i (\Psi_k L) D \Lambda D' (\Psi_k L)' \mathbf{e}_i},$$

³⁹ where $\Lambda = \text{diag}(D_{11}^{-1}, \dots, D_{I^* I^*}^{-1})$.

⁴⁰ S.3. Example: MAR(1)

To facilitate the understanding of the model in eq. (5), this section shows a special case of the general model in eq. (5), that we call the matrix autoregressive model, or MAR(p). We illustrate in a toy example the case with only the lagged dependent variable (i.e., \mathcal{Y}_{t-1}) as regressor. Assuming $N = 2$, $p = 1$ and $I_1 = I_2 = 2$, we obtain a matrix autoregressive model (i.e. with $\mathcal{Y}_t = Y_t$, $\mathcal{E}_t = E_t$) with one lag. Denoting $\text{vec}(\mathcal{Y}_t) = \mathbf{y}_t$ and $\text{vec}(\mathcal{E}_t) = \boldsymbol{\epsilon}_t$, as follows

$$\begin{aligned} \mathcal{Y}_t &= \begin{pmatrix} y_{11,t} & y_{12,t} \\ y_{21,t} & y_{22,t} \end{pmatrix} \implies \text{vec}(\mathcal{Y}_t) = (y_{11,t}, y_{12,t}, y_{21,t}, y_{22,t})' = (y_{1,t}, y_{2,t}, y_{3,t}, y_{4,t})' \\ \mathcal{B} &= (\mathcal{B}_{::1}, \mathcal{B}_{::2}, \mathcal{B}_{::3}, \mathcal{B}_{::4}), \quad \text{with } \mathcal{B}_{::k} = B_k = \begin{pmatrix} b_{11k} & b_{12k} \\ b_{21k} & b_{22k} \end{pmatrix} \\ \mathcal{E}_t &= \begin{pmatrix} \epsilon_{11,t} & \epsilon_{12,t} \\ \epsilon_{21,t} & \epsilon_{22,t} \end{pmatrix} \implies \text{vec}(\mathcal{E}_t) = (\epsilon_{11,t}, \epsilon_{12,t}, \epsilon_{21,t}, \epsilon_{22,t})' = (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{4,t})'. \end{aligned}$$

Therefore, model (5) becomes

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{B} \bar{x}_1 \mathcal{Y}_{t-1} + \mathcal{E}_t \implies Y_t = \mathcal{B} \bar{x}_1 Y_{t-1} + E_t \\ \begin{pmatrix} y_{11,t} & y_{12,t} \\ y_{21,t} & y_{22,t} \end{pmatrix} &= \mathcal{B}_{::1} \mathbf{y}_{1,t-1} + \dots + \mathcal{B}_{::4} \mathbf{y}_{4,t-1} + \begin{pmatrix} \epsilon_{11,t} & \epsilon_{12,t} \\ \epsilon_{21,t} & \epsilon_{22,t} \end{pmatrix} \\ &= \begin{pmatrix} b_{11,1} & b_{12,1} \\ b_{21,1} & b_{22,1} \end{pmatrix} \mathbf{y}_{1,t-1} + \dots + \begin{pmatrix} b_{11,4} & b_{12,4} \\ b_{21,4} & b_{22,4} \end{pmatrix} \mathbf{y}_{4,t-1} + \begin{pmatrix} \epsilon_{11,t} & \epsilon_{12,t} \\ \epsilon_{21,t} & \epsilon_{22,t} \end{pmatrix}. \end{aligned}$$

Assuming a PARAFAC(R) decomposition on the tensor coefficient \mathcal{B} yields

$$\mathcal{B} = \sum_{r=1}^R \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)} = \sum_{r=1}^R \begin{pmatrix} \beta_{1,1}^{(r)} \\ \beta_{1,2}^{(r)} \end{pmatrix} \circ \begin{pmatrix} \beta_{2,1}^{(r)} \\ \beta_{2,2}^{(r)} \end{pmatrix} \circ \begin{pmatrix} \beta_{3,1}^{(r)} \\ \beta_{3,2}^{(r)} \\ \beta_{3,3}^{(r)} \\ \beta_{3,4}^{(r)} \end{pmatrix}$$

$$\begin{aligned}
&= \sum_{r=1}^R \begin{pmatrix} \beta_{1,1}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,1}^{(r)} \beta_{2,2}^{(r)} \\ \beta_{1,2}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,2}^{(r)} \beta_{2,2}^{(r)} \end{pmatrix} \circ \begin{pmatrix} \beta_{3,1}^{(r)} \\ \beta_{3,2}^{(r)} \\ \beta_{3,3}^{(r)} \\ \beta_{3,4}^{(r)} \end{pmatrix} \\
&= \left(\sum_{r=1}^R \beta_{3,1}^{(r)} \begin{pmatrix} \beta_{1,1}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,1}^{(r)} \beta_{2,2}^{(r)} \\ \beta_{1,2}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,2}^{(r)} \beta_{2,2}^{(r)} \end{pmatrix}, \dots, \sum_{r=1}^R \beta_{3,4}^{(r)} \begin{pmatrix} \beta_{1,1}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,1}^{(r)} \beta_{2,2}^{(r)} \\ \beta_{1,2}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,2}^{(r)} \beta_{2,2}^{(r)} \end{pmatrix} \right) \\
&= (\mathcal{B}_{::1}, \mathcal{B}_{::2}, \mathcal{B}_{::3}, \mathcal{B}_{::4}),
\end{aligned}$$

where, for each $i = 1, \dots, 4$, we have

$$\mathcal{B}_{::k} = \sum_{r=1}^R \beta_{3,k}^{(r)} \begin{pmatrix} \beta_{1,1}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,1}^{(r)} \beta_{2,2}^{(r)} \\ \beta_{1,2}^{(r)} \beta_{2,1}^{(r)} & \beta_{1,2}^{(r)} \beta_{2,2}^{(r)} \end{pmatrix} = \begin{pmatrix} b_{11k} & b_{12k} \\ b_{21k} & b_{22k} \end{pmatrix},$$

hence, by choosing a PARAFAC(R) decomposition, we are assuming

$$b_{ijk} = \sum_{r=1}^R \beta_{1,i}^{(r)} \beta_{2,j}^{(r)} \beta_{3,k}^{(r)}, \quad i = 1, 2, \quad j = 1, 2, \quad k = 1, \dots, 4.$$

41 S.4. Special cases and corresponding proofs

42 In this section we provide the derivation of the results in the main paper.
43 We start by recalling a relationship between the outer product, the
44 Kronecker product and the ordinary matrix product. For two vectors $\mathbf{u} \in \mathbb{R}^n$
45 and $\mathbf{v} \in \mathbb{R}^m$ it holds $\mathbf{u} \otimes \mathbf{v}' = \mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}'$.

46 The model in eq. (5) is a generalization of several well-known econometric
47 models, as shown in the following remarks. See Section S.4 for the proofs of
48 these results.

REMARK S.4.1 (UNIVARIATE). If $I_i = 1$ for $i = 1, \dots, N$, then model (5)
reduces to a univariate regression

$$y_t = \alpha_0 + \sum_{j=1}^p \alpha_j y_{t-j} + \boldsymbol{\beta}' \text{vec}(\mathcal{X}_t) + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \quad (\text{S10})$$

49 where the coefficients of (5) become $\mathcal{A}_j = \alpha_j \in \mathbb{R}$, $j = 0, \dots, p$ and $\mathcal{B} = \boldsymbol{\beta} \in \mathbb{R}^{J^*}$.

50 PROOF. Consider model (5) when $I_j = 1$, for $j = 1, \dots, N$. Note that a N -
51 order tensor whose modes have all unit length is equivalent to a 1-order tensor,
52 i.e. a scalar. As a consequence, the dependent variable becomes $y_t \in \mathbb{R}$ and the
53 autoregressive coefficient tensors reduce to $\alpha_j \in \mathbb{R}$, $j = 0, \dots, p$. The coefficient
54 tensor related to the covariates \mathcal{X}_t becomes a vector $\boldsymbol{\beta} \in \mathbb{R}^{J^*}$. Finally, the error

term distribution reduces to a univariate normal with 0 mean and variance σ^2 . In this framework, the mode- $N + 1$ product reduces to the standard inner product between vectors.

The PARAFAC(R) decomposition can still be applied in this case. We get

$$\alpha_j = \sum_{r=1}^R \alpha_{j,1}^{(r)} \circ \dots \circ \alpha_{j,N}^{(r)} = \sum_{r=1}^R \alpha_{j,1}^{(r)} \cdots \alpha_{j,N}^{(r)},$$

for each $j = 0, \dots, p$, where the outer product reduces to the ordinary scalar multiplication and all $\alpha_{j,k}^{(r)}$, $k = 1, \dots, N$, $r = 1, \dots, R$ are scalars. Similarly, we have

$$\beta = \sum_{r=1}^R \beta_1^{(r)} \circ \dots \circ \beta_N^{(r)} \circ \beta_{N+1}^{(r)} = \sum_{r=1}^R \beta_1^{(r)} \cdot \dots \cdot \beta_N^{(r)} \cdot \beta_{N+1}^{(r)}$$

since again the outer product reduces to the ordinary scalar multiplication and all $\beta_{j,k}^{(r)}$, $k = 1, \dots, N$, $r = 1, \dots, R$ are scalars, while the marginal corresponding to the last mode $N + 1$ is a vector of length J^* .

REMARK S.4.2 (SUR). If $I_i = 1$ for $i = 2, \dots, N$ and define by $\mathbf{1}_n$ the unit vector of length n , then model (5) reduces to a *Seemingly Unrelated Regression (SUR)* model (Zellner (1962))

$$\mathbf{y}_t = \boldsymbol{\alpha}_0 + B \times_2 \text{vec}(\mathcal{X}_t) + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \Sigma), \quad (\text{S11})$$

where $I_1 = m$ and the coefficients of (5) become $\mathcal{A}_j = 0$, $j = 1, \dots, p$, $\mathcal{A}_0 = \boldsymbol{\alpha}_0 \in \mathbb{R}^m$ and $\mathcal{B} = B \in \mathbb{R}^{m \times J^*}$. Note that, by definition, $B \times_2 \text{vec}(\mathcal{X}_t) = B \text{vec}(\mathcal{X}_t)$.

REMARK S.4.3 (VARX AND PANEL VAR). Consider the setup of Remark S.4.2. If $\mathbf{z}_t = \mathbf{y}_{t-1}$, then we obtain a VARX(1) model, with restricted covariance matrix. Another vector of regressors $\mathbf{w}_t = \text{vec}(W_t) \in \mathbb{R}^q$ may enter the regression (S11) pre-multiplied (along mode-3) by a tensor $\mathcal{D} \in \mathbb{R}^{m \times n \times q}$. Therefore, model (5) encompasses as a particular case also the panel VAR models of Canova and Ciccarelli (2004), Canova et al. (2007), Canova and Ciccarelli (2009), provided that we make the same restriction on Σ .

PROOF. Consider model (5) with $I_1 = m$ and $I_j = 1$, for $j = 2, \dots, N$. Denote by $\mathbf{x}_t = \text{vec}(\mathcal{X}_t)$ the external covariates. Note that the mode- $N + 1$ product become mode-2 product and the distribution of the error term reduces to the multivariate (m -dimensional) normal. The dependent variable reduces to the vector $\mathbf{y}_t \in \mathbb{R}^m$ while the coefficient tensors become $\boldsymbol{\alpha}_0 \in \mathbb{R}^m$, $A_j \in \mathbb{R}^{m \times m}$, for $j = 1, \dots, p$ and $B \in \mathbb{R}^{m \times J^*}$.

Assuming a PARAFAC(R) decomposition, we get the same result for $\boldsymbol{\alpha}_0$ as in the previous proof, having in this case $N - 1$ scalar marginals and one vector

marginal. For the remaining tensors, it holds

$$A_j = \sum_{r=1}^R \boldsymbol{\alpha}_{j,1}^{(r)} \circ (\alpha_{j,2}^{(r)} \cdot \dots \cdot \alpha_{j,N-1}^{(r)}) \circ \boldsymbol{\alpha}_N^{(r)} = \sum_{r=1}^R A_j^{(r)} \cdot (\alpha_{j,2}^{(r)} \cdot \dots \cdot \alpha_{j,N-1}^{(r)}).$$

Similarly, for the matrix B one gets

$$B = \sum_{r=1}^R \boldsymbol{\beta}_1^{(r)} \circ (\beta_2^{(r)} \cdot \dots \cdot \beta_N^{(r)}) \circ \boldsymbol{\beta}_{N+1}^{(r)} = \sum_{r=1}^R B^{(r)} \cdot (\beta_2^{(r)} \cdot \dots \cdot \beta_N^{(r)}).$$

It remains to prove that the structure imposed by standard VARX and Panel VAR models holds also in the model of eq. (5). Notice that the latter does not impose any restriction on the coefficients, other than the PARAFAC(R) decomposition. It must be stressed that it is not possible to achieve the desired structure of the coefficients, in terms of the location of the zeros, by means of an accurate choice of the marginals. In fact, the decomposition we are assuming does not allow to create a particular structure on the resulting tensor.

Nonetheless, it is still possible to achieve the desired result by a slight modification of the model in eq. (5). For example, consider the coefficient tensor \mathcal{B} , then to create a tensor whose entries are non-zero only in some pre-specified (hence *a-priori* known) cells, it suffices to multiply \mathcal{B} by a binary tensor (i.e. one where all entries are either 0 or 1) via the Hadamard product. In formulas, let $\mathcal{H} \in \{0, 1\}^{I_1 \times \dots \times I_N \times J}$, such that it has 0 only in those cells which are known to be null. Then $\bar{\mathcal{B}} = \mathcal{H} \odot \mathcal{B}$ has the desired structure. The same way of reasoning holds for any coefficient tensor as well as for the covariance matrices.

To conclude, in Panel VAR models one generally has as regressors in each equation a function of the endogenous variables (for example their average). Since this does not affect the coefficients of the model, it is possible to re-create it in our framework by simply rearranging the regressors in eq. (5) accordingly. In terms of the model, none of the issues described invalidates the formulation of eq. (5), which is able to encompass all of them by suitable rearrangements of the covariates and/or the coefficients, which are consistent with the general model.

REMARK S.4.4 (VECM). *The model in eq. (5) generalises the Vector Error Correction Model (VECM) widely used in multivariate time series analysis (see [Engle and Granger \(1987\)](#), [Schotman and Van Dijk \(1991\)](#)). Consider a K -dimensional VAR(1) model*

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

Defining $\Delta\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ and $\Pi = (B - I) = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $K \times R$

matrices of rank $R < K$, we obtain the associated VECM

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t. \quad (\text{S12})$$

This is used for studying the cointegration relations among the components of \mathbf{y}_t . Since $\Pi = \boldsymbol{\alpha} \boldsymbol{\beta}' = \sum_{r=1}^R \boldsymbol{\alpha}_{:,r} \boldsymbol{\beta}_{:,r}' = \sum_{r=1}^R \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)}$, we can interpret the VECM model in eq. (S12) as a particular case of the model in eq. (5) where the coefficient \mathcal{B} is the matrix $\Pi = \boldsymbol{\alpha} \boldsymbol{\beta}'$. Furthermore by writing $\Pi = \sum_{r=1}^R \tilde{\boldsymbol{\beta}}_1^{(r)} \circ \tilde{\boldsymbol{\beta}}_2^{(r)}$ we can interpret this relation as a rank- R PARAFAC decomposition of \mathcal{B} . Following this analogy, the PARAFAC rank corresponds to the cointegration rank, $\tilde{\boldsymbol{\beta}}_1^{(r)}$ are the mean-reverting coefficients and $\tilde{\boldsymbol{\beta}}_2^{(r)} = (\tilde{\beta}_{2,1}^{(r)}, \dots, \tilde{\beta}_{2,K}^{(r)})'$ are the cointegrating vectors. See Section S.4 for details. This interpretation opens the way to reparametrization of \mathcal{B} based on tensor SVD representations, and to the application of regularization methods in the spirit of [Baştürk et al. \(2017\)](#). This is beyond the scope of the paper, thus we leave it for further research.

REMARK S.4.5 (FOLLOWS FROM [REMARK S.4.4](#)). From the VECM in eq. (S12) and denoting $\mathbf{y}_{t-1} = \text{vec}(Y_{t-1})$ we can obtain an explicit form for the long run equilibrium (or cointegrating) relations, as follows

$$\boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} = \left(\sum_{r=1}^R \boldsymbol{\gamma}_1^{(r)} \circ \boldsymbol{\gamma}_2^{(r)} \right) \bar{x}_1 \mathbf{y}_{t-1} = \left(\sum_{r=1}^R \boldsymbol{\gamma}_1^{(r)} \boldsymbol{\gamma}_2^{(r)'} \right) \mathbf{y}_{t-1} = \sum_{r=1}^R \boldsymbol{\gamma}_1^{(r)} (\boldsymbol{\gamma}_2^{(r)'} \mathbf{y}_{t-1}),$$

where $\boldsymbol{\gamma}_1^{(r)}$ and $\boldsymbol{\gamma}_2^{(r)}$ are vectors of length K . The marginals $(\boldsymbol{\gamma}_2^{(r)})_r$ can thus be interpreted as the long run cointegrating relationships, and the marginals $(\boldsymbol{\gamma}_2^{(r)})_r$ are the corresponding loadings.

REMARK S.4.6 (MAI OF [CARRIERO ET AL. \(2016\)](#)). The multivariate autoregressive index model (MAI) of [Carriero et al. \(2016\)](#) is another special case of model (5). A MAI is a VAR model with a low rank decomposition imposed on the coefficient matrix, as follows

$$\mathbf{y}_t = \mathbf{A} \mathbf{B}_0 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

where \mathbf{y}_t is a $(n \times 1)$ vector, whereas \mathbf{A}, \mathbf{B}_0 are $(n \times R)$ and $(R \times n)$ matrices, respectively. In [Carriero et al. \(2016\)](#), the authors assumed $R = 1$. This corresponds to our parametrization using $R = 1$ and defining $\mathbf{A} \boldsymbol{\beta}_1^{(1)}$ and $\mathbf{B}_0' = \boldsymbol{\beta}_2^{(1)}$, which leads us to $\mathbf{A} \mathbf{B}_0 = \boldsymbol{\beta}_1^{(1)} \circ \boldsymbol{\beta}_2^{(1)}$.

S.5. Prior distribution on tensor entries

PROOF (PROOF OF [LEMMA 3.1](#)). The distribution of each of these products has been characterised by [Springer and Thompson \(1970\)](#), who proved the

¹²⁰ following theorem.

THEOREM S.5.1 (4 IN SPRINGER AND THOMPSON (1970)).

The probability density function of the product $z = \prod_{h=1}^H x_h$ of H independent Normal random variables $x_h \sim \mathcal{N}(0, \sigma_h^2)$, $h = 1, \dots, H$, is proportional to a Meijer G-function

$$p(z | (\sigma_h^2)_{h=1}^H) = K \cdot G_{H,0}^{H,0} \left(z^2 \prod_{h=1}^H \frac{1}{2\sigma_h} \middle| 0 \right),$$

where the normalising constant is

$$K = \left((2\pi)^{H/2} \prod_{h=1}^H \sigma_h \right)^{-1}$$

and $G_{p,q}^{m,n}(\cdot | \cdot)$ is a Meijer G-function (with $c \in \mathbb{R}$ and $s \in \mathbb{C}$)

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

¹²¹ The integral is taken over a vertical line in the complex plane. Note that in
¹²² the special case $H = 2$ we have $z \sim c_1 P_1 - c_2 P_2$, with $P_1, P_2 \sim \chi_1^2$ and
¹²³ $c_1 = \text{Var}(x_1 + x_2)/4$, $c_2 = \text{Var}(x_1 - x_2)/4$. In this case, the resulting distribution
¹²⁴ is called product Normal distribution.

¹²⁵ Therefore, the result follows from **Theorem S.5.1**, with $z = \beta_r$, $H = 4$,
¹²⁶ $\sigma_h = \tau \phi_r w_{h,r,m_h}$ and where the parameters of the G-function are $m = p = 4 =$,
¹²⁷ $n = q = 0$, $(a_1, \dots, a_p) = (0, \dots, 0)$ and $(b_1, \dots, b_q) = (0, \dots, 0)$.

We assessed the shape of this marginal distribution in a simulated setting, and found that it has fatter tails than the Gaussian distribution. In particular, Fig. 4 show the empirical distribution of two randomly chosen entries of a 3-order tensor \mathcal{B} whose PARAFAC decomposition is assumed with $R = 5$. The probability density function of a Laplace (or double exponential) distribution with mean $\mu \in \mathbb{R}$ and variance $2b^2$, with $b > 0$, is

$$f(x | \mu, b) = \frac{1}{2b} \exp \left(-\frac{|x - \mu|}{2b} \right) \quad x \in \mathbb{R}.$$

¹²⁸ Compared to the standard normal and standard Laplace distribution, the prior
¹²⁹ distribution induced on the single entries of the tensor has fatter tails.

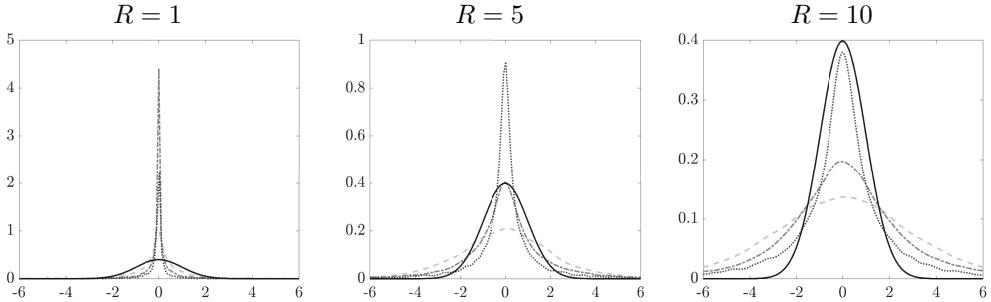


Figure 2: Monte Carlo simulation from the prior distribution of entry b_{i_1, \dots, i_N} of a generic N -order tensor, for varying rank R . In column: simulation with $R = 1$ (left), $R = 5$ (middle) and $R = 10$ (right). In all plots: standard Normal (continuous line) and prior for b_{i_1, \dots, i_N} , for $N = 2$ (dashed line), $N = 4$ (dash-dotted line) and $N = 6$ (dotted line).

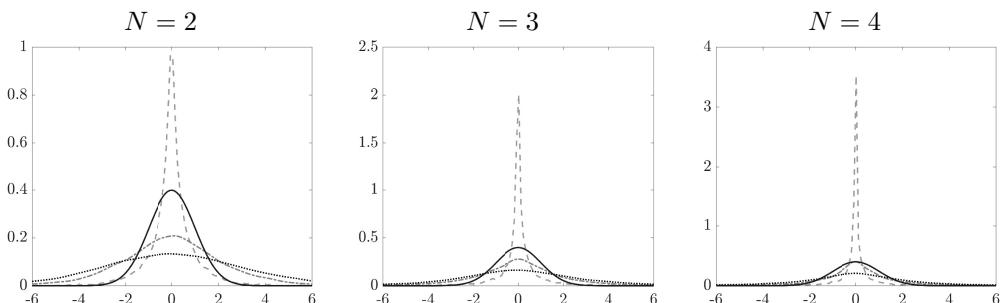


Figure 3: Monte Carlo simulation from the prior distribution of entry b_{i_1, \dots, i_N} of a N -order tensor, with rank R , for varying N . In column: simulation with $N = 2$ (left), $N = 3$ (middle) and $N = 4$ (right). In all plots: standard Normal (continuous line) and prior for b_{i_1, \dots, i_N} , for $R = 1$ (dashed line), $R = 5$ (dash-dotted line) and $R = 10$ (dotted line).

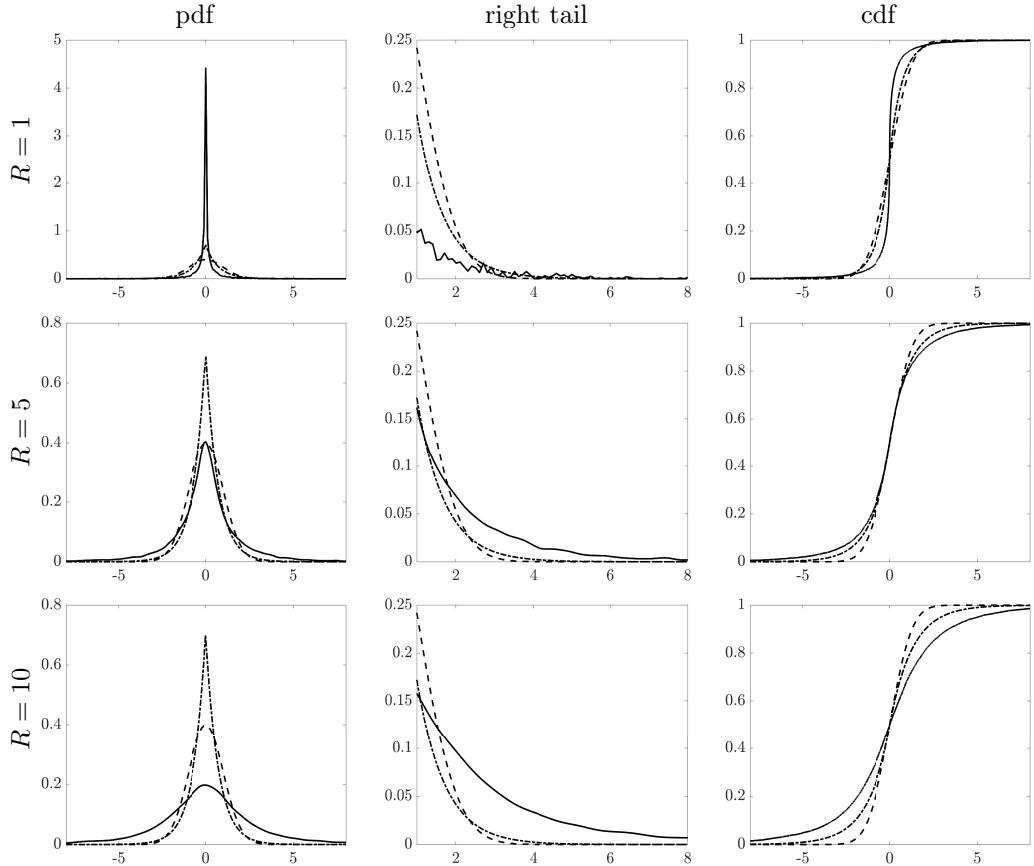


Figure 4: Monte Carlo simulation from the prior distribution of a generic 4-order tensor entry b_{ijkp} (continuous line), standard Normal distribution (dashed line) and standard Laplace distribution (dash-dotted line). In column: probability density function (*left*), right tail the probability density function (*middle*), cumulative distribution function (*right*). In row: simulations with $R = 1$, $R = 5$ and $R = 10$ (*first, second and third*, respectively).

130 S.6. Computational details - ART(1) model

131 In this section we will follow the convention of denoting the prior distributions
132 with $\pi(\cdot)$. In addition, let $\mathbf{W} = (W_{j,r})_{j,r}$ be the collection of all (local variance)
133 matrices $W_{j,r}$, for $j = 1, \dots, J$ and $r = 1, \dots, R$, let $I_0 = \sum_{j=1}^J I_j$ be the sum
134 of the length of each mode of the tensor \mathcal{B} and let $\mathbf{Y} = (\mathcal{Y}_t)_t$ the collection of
135 observed variables. Recall that in the TAR(1) model in eq. (15), the variable \mathcal{Y}_t
136 is a 3-order tensor, thus we have $J = 4$.

¹³⁷ **S.6.1. Full conditional distribution of ϕ_r**

In order to derive this posterior distribution, we use Guhaniyogi et al. (2017, Lemma 7.9). Recall that: $a_\tau = \alpha R$ and $b_\tau = \alpha(R)^{1/J}$. The posterior full conditional distribution of ϕ is

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

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

$$\begin{aligned} p(\phi|\mathcal{B}, \mathbf{W}) &\propto \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau-1-\frac{RI_0}{2}} e^{-b_\tau\tau} \left(\prod_{r=1}^R \exp \left(-\frac{1}{2\tau\phi_r} C_r \right) \right) d\tau \\ &= \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau-1-\frac{Rd_0}{2}} \exp \left(-b_\tau\tau - \sum_{r=1}^R \frac{C_r}{2\tau\phi_r} \right) d\tau. \quad (\text{S13}) \end{aligned}$$

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

$$x \sim \text{GiG}(a, b, c) \iff p(x|a, b, c) = \frac{(a/b)^{\frac{c}{2}}}{2K_c(\sqrt{ab})} x^{c-1} \exp \left(-\frac{1}{2}(ax + b/x) \right),$$

with $K_c(\cdot)$ a modified Bessel function of the second type. Our goal is to reconcile eq. (S13) to the kernel of this distribution. Since by definition $\sum_{r=1}^R \phi_r = 1$, it holds that $\sum_{r=1}^R (b_\tau \tau \phi_r) = (b_\tau \tau) \sum_{r=1}^R \phi_r = b_\tau \tau$. This allows to rewrite the exponential as

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

$$= \int_0^{+\infty} \left(\prod_{r=1}^R \phi_r^{\alpha - \frac{I_0}{2} - 1} \right) \tau^{\left(\alpha R - \frac{RI_0}{2} \right) - 1} \exp \left(- \sum_{r=1}^R \left(\frac{C_r}{2\tau\phi_r} + b_\tau \tau \phi_r \right) \right) d\tau$$

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

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

then, we obtain ϕ_r by renormalizing (see Kruijer et al. (2010)): $\phi_r = \psi_r / \sum_{l=1}^R \psi_l$.

140 S.6.2. Full conditional distribution of τ

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

$$\begin{aligned} p(\tau | \mathcal{B}, \mathbf{W}, \phi) &\propto \pi(\tau) p(\mathcal{B} | \mathbf{W}, \phi, \tau) \\ &\propto \tau^{a_\tau - 1} e^{-b_\tau \tau} \left(\prod_{r=1}^R (\tau\phi_r)^{-\frac{I_0}{2}} \exp \left(- \frac{1}{2\tau\phi_r} \sum_{j=1}^4 \beta_j^{(r)'} (W_{j,r})^{-1} \beta_j^{(r)} \right) \right) \\ &\propto \tau^{a_\tau - \frac{RI_0}{2} - 1} \exp \left(- b_\tau \tau - \left(\sum_{r=1}^R \frac{C_r}{\phi_r} \frac{1}{\tau} \right) \right). \end{aligned}$$

This is the kernel of a generalized inverse Gaussian

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

141 S.6.3. Full conditional distribution of $\lambda_{j,r}$

Start by observing that, for $j = 1, \dots, 4$ and $r = 1, \dots, R$, the prior distribution on the vector $\beta_j^{(r)}$ defined in eq. (17) implies that each component follows a double exponential distribution

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

with probability density function given by

$$\pi(\beta_{j,p}^{(r)} | \lambda_{j,r}, \phi_r, \tau) = \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\left(-\frac{|\beta_{j,p}^{(r)}|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}}\right). \quad (\text{S14})$$

Then, exploiting the Gamma prior and eq. (S14)

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

Thus, the full conditional distribution of $\lambda_{j,r}$ is given by

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

¹⁴² S.6.4. Full conditional distribution of $w_{j,r,p}$

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

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

where the second row comes from the fact that $w_{j,r,p}$ influences only the p -th component of the vector $\boldsymbol{\beta}_j^{(r)}$. Hence, we get

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

¹⁴³ S.6.5. Full conditional distributions of PARAFAC marginals

Define $\boldsymbol{\alpha}_1 \in \mathbb{R}^I$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^J$ and $\boldsymbol{\alpha}_3 \in \mathbb{R}^K$ and let $\mathcal{A} = \text{vec}(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3)$. Then, from Lemma A.3 it holds

$$\text{vec}(\mathcal{A}) = \text{vec}(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3) = \boldsymbol{\alpha}_3 \otimes \text{vec}(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2')$$

$$= \boldsymbol{\alpha}_3 \otimes (\boldsymbol{\alpha}_2 \otimes \mathbf{I}_I) \text{vec}(\boldsymbol{\alpha}_1) = (\boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \mathbf{I}_I) \boldsymbol{\alpha}_1 \quad (\text{S15})$$

$$= \boldsymbol{\alpha}_3 \otimes ((\mathbf{I}_J \otimes \boldsymbol{\alpha}_1) \text{vec}(\boldsymbol{\alpha}'_2)) = (\boldsymbol{\alpha}_3 \otimes \mathbf{I}_J \otimes \boldsymbol{\alpha}_1) \boldsymbol{\alpha}_2 \quad (\text{S16})$$

$$\begin{aligned} &= \text{vec}(\text{vec}(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}'_2) \boldsymbol{\alpha}'_3) = (\mathbf{I}_K \otimes \text{vec}(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}'_2)) \text{vec}(\boldsymbol{\alpha}'_3) \\ &= (\mathbf{I}_K \otimes \text{vec}(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}'_2)) \boldsymbol{\alpha}_3 = (\mathbf{I}_K \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1) \boldsymbol{\alpha}_3. \end{aligned} \quad (\text{S17})$$

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

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{B} \bar{x}_1 \mathbf{x}_t + \mathcal{E}_t \\ \text{vec}(\mathcal{Y}_t) &= \text{vec}(\mathcal{B} \bar{x}_1 \mathbf{x}_t + \mathcal{E}_t) \\ &= \text{vec}(\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t) + \text{vec}(\mathcal{B}_r \bar{x}_1 \mathbf{x}_t) + \text{vec}(\mathcal{E}_t), \end{aligned}$$

where the term in the middle can be re-written as

$$\text{vec}(\mathcal{B}_r \bar{x}_1 \mathbf{x}_t) = \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)}.$$

It is then possible to make explicit the dependence on each PARAFAC marginal by exploiting the results in eq. (S15)-(S17), as follows

$$\text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \mathbf{b}_4 \boldsymbol{\beta}_4^{(r)} \quad (\text{S18})$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_I) \boldsymbol{\beta}_1^{(r)} = \mathbf{b}_1 \boldsymbol{\beta}_1^{(r)} \quad (\text{S19})$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_2^{(r)} = \mathbf{b}_2 \boldsymbol{\beta}_2^{(r)} \quad (\text{S20})$$

$$= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\mathbf{I}_K \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)} = \mathbf{b}_3 \boldsymbol{\beta}_3^{(r)}. \quad (\text{S21})$$

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

$$\begin{aligned} L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &= \prod_{t=1}^T (2\pi)^{-\frac{I_1 I_2 I_3}{2}} |\Sigma_3|^{-\frac{I_1 I_2}{2}} |\Sigma_2|^{-\frac{I_1 I_3}{2}} |\Sigma_1|^{-\frac{I_2 I_3}{2}} \\ &\quad \cdot \exp\left(-\frac{1}{2} (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t) \bar{x}_3 (\circ_{j=1}^3 \Sigma_j^{-1}) \bar{x}_3 (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t)\right) \\ &\propto \exp\left(-\frac{1}{2} \sum_{t=1}^T \tilde{\mathcal{E}}_t \bar{x}_3 (\Sigma_1^{-1} \circ \Sigma_2^{-1} \circ \Sigma_3^{-1}) \bar{x}_3 \tilde{\mathcal{E}}_t\right), \end{aligned}$$

with

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

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

$$\begin{aligned} L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &= \prod_{t=1}^T (2\pi)^{-\frac{I_1 I_2 I_3}{2}} |\Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1|^{-\frac{1}{2}} \\ &\cdot \exp \left(-\frac{1}{2} \text{vec} (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t)' (\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \text{vec} (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t) \right) \\ &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T \text{vec} (\tilde{\mathcal{E}}_t)' (\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \text{vec} (\tilde{\mathcal{E}}_t) \right). \end{aligned}$$

Thus, defining with $\mathbf{y}_t = \text{vec} (\mathcal{Y}_t)$ and $\boldsymbol{\Sigma}^{-1} = \Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}$, we obtain

$$\begin{aligned} L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &\propto \\ &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T \text{vec} (\tilde{\mathcal{E}}_t)' (\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \text{vec} (\tilde{\mathcal{E}}_t) \right) \\ &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T \left(\text{vec} (\mathcal{Y}_t) - \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t) - \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \right. \right. \\ &\quad \cdot \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \left. \right)' \boldsymbol{\Sigma}^{-1} \left(\text{vec} (\mathcal{Y}_t) - \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t) - \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \right) \Bigg) \\ &= \exp \left(-\frac{1}{2} \sum_{t=1}^T \mathbf{y}'_t \boldsymbol{\Sigma}^{-1} \mathbf{y}_t - 2 \mathbf{y}'_t \boldsymbol{\Sigma}^{-1} \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t) \right. \\ &\quad + \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t)' \boldsymbol{\Sigma}^{-1} \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t) \\ &\quad - 2 \mathbf{y}'_t \boldsymbol{\Sigma}^{-1} \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \\ &\quad + 2 \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t)' \boldsymbol{\Sigma}^{-1} \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)})' \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \\ &\quad + \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)})' \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \Bigg) \\ &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec} (\mathcal{B}_{-r} \bar{x}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \right. \\ &\quad + \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)})' \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \text{vec} (\beta_1^{(r)} \circ \beta_2^{(r)} \circ \beta_3^{(r)}) \langle \beta_4^{(r)}, \mathbf{x}_t \rangle \Bigg). \end{aligned} \tag{S22}$$

Now, we focus on a specific $j = 1, 2, 3, 4$ and derive proportionality results that will be necessary to obtain the posterior full conditional distributions of the

PARAFAC marginals of the tensor \mathcal{B} . Consider the case $j = 1$. By exploiting eq. (S19) we get

$$\begin{aligned}
 L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &\propto \\
 &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \Sigma^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t \boldsymbol{\beta}_4^{(r)} \right. \\
 &\quad \left. + \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)})' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \Sigma^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \right) \\
 &= \exp \left(-\frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \Sigma^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \boldsymbol{\beta}_1^{(r)} \right. \\
 &\quad \left. + (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \boldsymbol{\beta}_1^{(r)})' \Sigma^{-1} (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \boldsymbol{\beta}_1^{(r)}) \right) \\
 &= \exp \left(-\frac{1}{2} \sum_{t=1}^T \boldsymbol{\beta}_1^{(r)'} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1})' \Sigma^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \boldsymbol{\beta}_1^{(r)} \right. \\
 &\quad \left. - 2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \Sigma^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \boldsymbol{\beta}_1^{(r)} \right) \\
 &= \exp \left(-\frac{1}{2} \boldsymbol{\beta}_1^{(r)'} \mathbf{S}_1^L \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L \boldsymbol{\beta}_1^{(r)} \right), \tag{S23}
 \end{aligned}$$

with

$$\begin{aligned}
 \mathbf{S}_1^L &= \sum_{t=1}^T (\boldsymbol{\beta}_3^{(r)'} \otimes \boldsymbol{\beta}_2^{(r)'} \otimes \mathbf{I}_{I_1}) \Sigma^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \\
 \mathbf{m}_1^L &= \sum_{t=1}^T (\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \Sigma^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle.
 \end{aligned}$$

Consider the case $j = 2$. From eq. (S20) we get

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

$$\begin{aligned}
& + (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_2^{(r)})' \boldsymbol{\Sigma}^{-1} (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_2^{(r)}) \\
& = \exp \left(- \frac{1}{2} \sum_{t=1}^T \boldsymbol{\beta}_2^{(r)'} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_2^{(r)} \right. \\
& \quad \left. - 2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_2^{(r)} \right) \\
& = \exp \left(- \frac{1}{2} \boldsymbol{\beta}_2^{(r)'} \mathbf{S}_2^L \boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L \boldsymbol{\beta}_2^{(r)} \right), \tag{S24}
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{S}_2^L &= \sum_{t=1}^T (\boldsymbol{\beta}_3^{(r)'} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)'}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \\
\mathbf{m}_2^L &= \sum_{t=1}^T (\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle.
\end{aligned}$$

Consider the case $j = 3$, by exploiting eq. (S21) we get

$$\begin{aligned}
L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &\propto \\
&\propto \exp \left(- \frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t \boldsymbol{\beta}_4^{(r)} \right. \\
&\quad \left. + (\text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle)' \boldsymbol{\Sigma}^{-1} (\text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle) \right) \\
&= \exp \left(- \frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)} \right. \\
&\quad \left. + (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)})' \boldsymbol{\Sigma}^{-1} (\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)}) \right) \\
&= \exp \left(- \frac{1}{2} \sum_{t=1}^T \boldsymbol{\beta}_3^{(r)'} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\Sigma}^{-1} (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)} \right. \\
&\quad \left. - 2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \boldsymbol{\beta}_3^{(r)} \right) \\
&= \exp \left(- \frac{1}{2} \boldsymbol{\beta}_3^{(r)'} \mathbf{S}_3^L \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L \boldsymbol{\beta}_3^{(r)} \right), \tag{S25}
\end{aligned}$$

with

$$\begin{aligned}\mathbf{S}_3^L &= \sum_{t=1}^T (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)'} \otimes \boldsymbol{\beta}_1^{(r)'}) \boldsymbol{\Sigma}^{-1} (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \\ \mathbf{m}_3^L &= \sum_{t=1}^T (\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} (\mathbf{I}_{I_3} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle.\end{aligned}$$

Finally, in the case $j = 4$. From eq. (S22) we get

$$\begin{aligned}L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) &\propto \\ &\propto \exp \left(-\frac{1}{2} \sum_{t=1}^T -2(\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t \boldsymbol{\beta}_4^{(r)} \right. \\ &\quad \left. + \boldsymbol{\beta}_4^{(r)'} \mathbf{x}_t \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)})' \boldsymbol{\Sigma}^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t \boldsymbol{\beta}_4^{(r)} \right) \\ &= \exp \left(-\frac{1}{2} \boldsymbol{\beta}_4^{(r)'} \mathbf{S}_4^L \boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L \boldsymbol{\beta}_4^{(r)} \right),\end{aligned}\tag{S26}$$

with

$$\begin{aligned}\mathbf{S}_4^L &= \sum_{t=1}^T \mathbf{x}_t \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)})' \boldsymbol{\Sigma}^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t \\ \mathbf{m}_4^L &= \sum_{t=1}^T (\mathbf{y}'_t - \text{vec}(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t)') \boldsymbol{\Sigma}^{-1} \text{vec}(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}) \mathbf{x}'_t.\end{aligned}$$

¹⁴⁴ It is now possible to derive the full conditional distributions for the PARAFAC
¹⁴⁵ marginals $\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}, \boldsymbol{\beta}_4^{(r)}$, as shown in the following.

¹⁴⁶ S.6.5.1. Full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$ is obtained by combining the prior distribution in eq. (17) and the likelihood in eq. (S23) as follows

$$\begin{aligned}p(\boldsymbol{\beta}_1^{(r)} | \boldsymbol{\beta}_{-1}^{(r)}, \mathcal{B}_{-r}, W_{1,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) &\propto L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \pi(\boldsymbol{\beta}_1^{(r)} | W_{1,r}, \phi_r, \tau) \\ &\propto \exp \left(-\frac{1}{2} \boldsymbol{\beta}_1^{(r)'} \mathbf{S}_1^L \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L \boldsymbol{\beta}_1^{(r)} \right) \cdot \exp \left(-\frac{1}{2} \boldsymbol{\beta}_1^{(r)'} (W_{1,r} \phi_r \tau)^{-1} \boldsymbol{\beta}_1^{(r)} \right) \\ &= \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_1^{(r)'} \mathbf{S}_1^L \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L \boldsymbol{\beta}_1^{(r)} + \boldsymbol{\beta}_1^{(r)'} (W_{1,r} \phi_r \tau)^{-1} \boldsymbol{\beta}_1^{(r)}) \right) \\ &= \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_1^{(r)'} (\mathbf{S}_1^L + (W_{1,r} \phi_r \tau)^{-1}) \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L \boldsymbol{\beta}_1^{(r)}) \right)\end{aligned}$$

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

where

$$\bar{\Sigma}_{\boldsymbol{\beta}_1^r} = ((W_{1,r}\phi_r\tau)^{-1} + \mathbf{S}_1^L)^{-1}, \quad \bar{\mu}_{\boldsymbol{\beta}_1^r} = \bar{\Sigma}_{\boldsymbol{\beta}_1^r}(\mathbf{m}_1^L)'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$ is given by

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

¹⁴⁷ *S.6.5.2. Full conditional distribution of $\beta_2^{(r)}$*

The posterior full conditional distribution of $\beta_2^{(r)}$ is obtained by combining the prior distribution in eq. (17) and the likelihood in eq. (S24) as follows

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

where

$$\bar{\Sigma}_{\boldsymbol{\beta}_2^r} = ((W_{2,r}\phi_r\tau)^{-1} + \mathbf{S}_2^L)^{-1}, \quad \bar{\mu}_{\boldsymbol{\beta}_2^r} = \bar{\Sigma}_{\boldsymbol{\beta}_2^r}(\mathbf{m}_2^L)'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$ is given by

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

¹⁴⁸ *S.6.5.3. Full conditional distribution of $\beta_3^{(r)}$*

The posterior full conditional distribution of $\beta_3^{(r)}$ is obtained by combining the prior distribution in eq. (17) and the likelihood in eq. (S25) as follows

$$\begin{aligned} p(\boldsymbol{\beta}_3^{(r)} | \boldsymbol{\beta}_{-3}^{(r)}, \mathcal{B}_{-r}, W_{3,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) &\propto L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) \pi(\boldsymbol{\beta}_3^{(r)} | W_{3,r}, \phi_r, \tau) \\ &\propto \exp \left(-\frac{1}{2} \boldsymbol{\beta}_3^{(r)'} \mathbf{S}_3^L \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L \boldsymbol{\beta}_3^{(r)} \right) \cdot \exp \left(-\frac{1}{2} \boldsymbol{\beta}_3^{(r)'} (W_{3,r}\phi_r\tau)^{-1} \boldsymbol{\beta}_3^{(r)} \right) \end{aligned}$$

$$\begin{aligned}
&= \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_3^{(r)'} \mathbf{S}_3^L \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L \boldsymbol{\beta}_3^{(r)} + \boldsymbol{\beta}_3^{(r)'} (W_{3,r} \phi_r \tau)^{-1} \boldsymbol{\beta}_3^{(r)}) \right) \\
&= \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_3^{(r)'} (\mathbf{S}_3^L + (W_{3,r} \phi_r \tau)^{-1}) \boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L \boldsymbol{\beta}_3^{(r)}) \right) \\
&= \exp \left(-\frac{1}{2} (\boldsymbol{\beta}_3^{(r)'} \bar{\Sigma}_{\boldsymbol{\beta}_3^r}^{-1} \boldsymbol{\beta}_3^{(r)} - 2\bar{\mu}_{\boldsymbol{\beta}_3^r} \boldsymbol{\beta}_3^{(r)}) \right),
\end{aligned}$$

where

$$\bar{\Sigma}_{\boldsymbol{\beta}_3^r} = ((W_{3,r} \phi_r \tau)^{-1} + \mathbf{S}_3^L)^{-1}, \quad \bar{\mu}_{\boldsymbol{\beta}_3^r} = \bar{\Sigma}_{\boldsymbol{\beta}_3^r} (\mathbf{m}_3^L)'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$ is given by

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

¹⁴⁹ **S.6.5.4. Full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$**

The posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$ is obtained by combining the prior distribution in eq. (17) and the likelihood in eq. (S26) as follows

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

where

$$\bar{\Sigma}_{\boldsymbol{\beta}_4^r} = ((W_{4,r} \phi_r \tau)^{-1} + \mathbf{S}_4^L)^{-1}, \quad \bar{\mu}_{\boldsymbol{\beta}_4^r} = \bar{\Sigma}_{\boldsymbol{\beta}_4^r} (\mathbf{m}_4^L)'.$$

Thus the posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$ is given by

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

¹⁵⁰ **S.6.6. Full conditional distribution of Σ_1**

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

properties of the vectorization and trace operators, we obtain

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

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

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

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

$$p(\Sigma_1 | \mathcal{B}, \mathbf{Y}, \Sigma_2, \Sigma_3, \gamma) \sim \mathcal{IW}_{I_1}(\nu_1 + T I_2 I_3, \gamma \Psi_1 + S_1).$$

¹⁵¹ S.6.7. Full conditional distribution of Σ_2

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_2 is conjugate. For ease of notation, define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \bar{\mathbf{x}}_1 \mathbf{x}_t$ and $\tilde{\mathbf{E}}_{(2),t}$ the mode-2 matricization of $\tilde{\mathcal{E}}_t$. By exploiting the relation between the tensor

normal distribution and the matrix normal distribution and the properties of the Kronecker product and of the vectorization and trace operators we obtain

$$\begin{aligned}
 p(\Sigma_2 | \mathcal{B}, \mathbf{Y}, \Sigma_1, \Sigma_3, \gamma) &\propto L(\mathbf{Y} | \mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) \pi(\Sigma_2 | \gamma) \\
 &\propto |\Sigma_2|^{-\frac{T I_1 I_3}{2}} \exp \left(-\frac{1}{2} \sum_{t=1}^T (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t) \bar{x}_3 (\Sigma_1^{-1} \circ \Sigma_2^{-1} \circ \Sigma_3^{-1}) \right. \\
 &\quad \left. \bar{x}_3 (\mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t) \right) \cdot |\Sigma_2|^{-\frac{\nu_2 + I_2 + 1}{2}} \exp \left(-\frac{1}{2} \text{tr} (\Psi_2 \Sigma_2^{-1}) \right) \\
 &\propto |\Sigma_2|^{-\frac{\nu_2 + I_2 + T I_1 I_3 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_2 \Sigma_2^{-1}) \right. \\
 &\quad \left. + \sum_{t=1}^T \tilde{\mathcal{E}}_t \bar{x}_3 (\Sigma_1^{-1} \circ \Sigma_2^{-1} \circ \Sigma_3^{-1}) \bar{x}_3 \tilde{\mathcal{E}}_t) \right) \\
 &\propto |\Sigma_2|^{-\frac{\nu_2 + I_2 + T I_1 I_3 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_2 \Sigma_2^{-1}) \right. \\
 &\quad \left. + \sum_{t=1}^T \text{tr} (\tilde{\mathbf{E}}'_{(2),t} (\Sigma_3^{-1} \otimes \Sigma_1^{-1} \otimes \Sigma_2^{-1}) \tilde{\mathbf{E}}_{(2),t})) \right) \\
 &\propto |\Sigma_2|^{-\frac{\nu_2 + I_2 + T I_1 I_3 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_2 \Sigma_2^{-1}) \right. \\
 &\quad \left. + \sum_{t=1}^T \text{tr} ((\Sigma_3^{-1} \otimes \Sigma_1^{-1}) \tilde{\mathbf{E}}'_{(2),t} \Sigma_2^{-1} \tilde{\mathbf{E}}_{(2),t})) \right) \\
 &\propto |\Sigma_2|^{-\frac{\nu_2 + I_2 + T I_1 I_3 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_2 \Sigma_2^{-1}) + \text{tr} (\sum_{t=1}^T \tilde{\mathbf{E}}_{(2),t} (\Sigma_3^{-1} \otimes \Sigma_1^{-1}) \tilde{\mathbf{E}}'_{(2),t} \Sigma_2^{-1})) \right) \\
 &\propto |\Sigma_2|^{-\frac{\nu_2 + I_2 + T I_1 I_3 + 1}{2}} \exp \left(-\frac{1}{2} \text{tr} (\gamma \Psi_2 \Sigma_2^{-1} + S_2 \Sigma_2^{-1}) \right),
 \end{aligned}$$

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

$$p(\Sigma_2 | \mathcal{B}, \mathbf{Y}, \Sigma_1, \Sigma_3) \sim \mathcal{IW}_{I_2}(\nu_2 + T I_1 I_3, \gamma \Psi_2 + S_2).$$

152 S.6.8. Full conditional distribution of Σ_3

Given a inverse Wishart prior, the posterior full conditional distribution for Σ_3 is conjugate. For ease of notation, define $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B} \bar{x}_1 \mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t}$ the mode-1 matricization of $\tilde{\mathcal{E}}_t$ and $\mathbf{Z}_3 = \Sigma_2^{-1} \otimes \Sigma_1^{-1}$. By exploiting the relation between the tensor normal distribution and the multivariate normal distribution and the properties of the vectorization and trace operators, we obtain

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

$$\begin{aligned}
&\propto |\Sigma_3|^{-\frac{T I_1 I_2}{2}} \exp \left(-\frac{1}{2} \sum_{t=1}^T \text{vec} (\mathcal{Y}_t - \mathcal{B} \bar{\times}_1 \mathbf{x}_t)' (\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \right. \\
&\quad \cdot \text{vec} (\mathcal{Y}_t - \mathcal{B} \bar{\times}_1 \mathbf{x}_t) \Big) \cdot |\Sigma_3|^{-\frac{\nu_3 + I_3 + 1}{2}} \exp \left(-\frac{1}{2} \text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) \right) \\
&\propto |\Sigma_3|^{-\frac{\nu_3 + I_3 + T I_1 I_2 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) + \sum_{t=1}^T \text{vec} (\tilde{\mathcal{E}}_t)' (\Sigma_3^{-1} \otimes \mathbf{Z}_3) \text{vec} (\tilde{\mathcal{E}}_t)) \right) \\
&\propto |\Sigma_3|^{-\frac{\nu_3 + I_3 + T I_1 I_2 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) \right. \\
&\quad \left. + \sum_{t=1}^T \text{vec} (\tilde{\mathbf{E}}_{(1),t})' (\Sigma_3^{-1} \otimes \mathbf{Z}_3) \text{vec} (\tilde{\mathbf{E}}_{(1),t})) \right) \\
&\propto |\Sigma_3|^{-\frac{\nu_3 + I_3 + T I_1 I_2 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) \right. \\
&\quad \left. + \sum_{t=1}^T \text{tr} (\text{vec} (\tilde{\mathbf{E}}_{(1),t})' \text{vec} (\mathbf{Z}_3 \tilde{\mathbf{E}}_{(1),t} \Sigma_3^{-1}))) \right) \\
&\propto |\Sigma_3|^{-\frac{\nu_3 + I_3 + T I_1 I_2 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) + \sum_{t=1}^T \text{tr} (\tilde{\mathbf{E}}_{(1),t}' \mathbf{Z}_3 \tilde{\mathbf{E}}_{(1),t} \Sigma_3^{-1})) \right).
\end{aligned}$$

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

$$\begin{aligned}
p(\Sigma_3 | \mathcal{B}, \mathbf{Y}, \Sigma_1, \Sigma_2) &\propto |\Sigma_3|^{-\frac{\nu_3 + I_3 + T I_1 I_2 + 1}{2}} \exp \left(-\frac{1}{2} (\text{tr} (\gamma \Psi_3 \Sigma_3^{-1}) + \text{tr} (S_3 \Sigma_3^{-1})) \right) \\
&\propto |\Sigma_3|^{-\frac{(\nu_3 + T I_1 I_2) + I_3 + 1}{2}} \exp \left(-\frac{1}{2} \text{tr} ((\gamma \Psi_3 + S_3) \Sigma_3^{-1}) \right),
\end{aligned}$$

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

$$p(\Sigma_3 | \mathcal{B}, \mathbf{Y}, \Sigma_1, \Sigma_2) \sim \mathcal{IW}_{I_3}(\nu_3 + T I_1 I_2, \gamma \Psi_3 + S_3).$$

153 S.6.9. Full conditional distribution of γ

Using a gamma prior distribution we have

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

thus

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

¹⁵⁴ **S.7. Initialisation details**

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

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

¹⁵⁵ Here $\Delta f = f(\boldsymbol{\theta}^*) - f(\boldsymbol{\theta}^{(i)})$ and T is a parameter called temperature. The key
¹⁵⁶ of the SA method is in the cooling scheme, which describes the deterministic,
¹⁵⁷ decreasing evolution of the temperature over the iterations of the algorithm: it
¹⁵⁸ has been proved that under sufficiently slow decreasing schemes, the SA yields
¹⁵⁹ a global optimum.

We use the SA algorithm for minimising the objective function

$$f((\boldsymbol{\beta}_j^{(r)})_{j,r}) = \kappa_N \psi_N + \kappa_J \psi_J,$$

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

$$\psi_N = \|\mathcal{B}^{SA}\|_2 \quad \psi_J = \sum_{r=1}^R \|\boldsymbol{\beta}_J^{(r)}\|_2.$$

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

cooling scheme which updates the temperature at each iteration of the SA

$$T_i = \frac{k}{1 + \log(i)} \quad i = 1, \dots, I^{SA},$$

where $k > 0$ is a tuning parameter, which can be interpreted as the initial value of the temperature. In order to perform the initialisation of the margins, we run the SA algorithm for $I^{SA} = 1,200$ iterations, then we took the vectors which gave the best fit in terms of minimum value of the objective function.

S.8. Simulation Results

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

We have fixed $I_1 = I_2 = I$ and performed experiments for different sizes I of the response and covariate matrices. We have generated a matrix-valued time series $(Y_t, X_t)_{t=1}^T$ by simulating each entry of X_t from

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

for $i = 1, \dots, I_1$, $j = 1, \dots, I_2$ and $t = 1, \dots, T$. Then, we have generated the matrix-valued response Y_t according to

$$Y_t = \mathcal{B} \bar{X}_1 \text{vec}(X_t) + E_t, \quad E_t \sim \mathcal{N}_{I_1, I_2}(\mathbf{0}, \Sigma_1, \mathbf{I}_{I_2}).$$

where $\mathbb{E}(\eta_{ij,t}\eta_{kl,v}) = 0$, $\mathbb{E}(\eta_{ij,t}E_v) = 0$, $\forall (i, j) \neq (k, l)$, $\forall t \neq v$, and $\alpha_{ij} \sim \mathcal{U}(-1, 1)$. We randomly draw \mathcal{B} using the PARAFAC decomposition in eq. (1), with rank $R = 5$ and marginals sampled from the prior distribution in eq. (17). The matrices X_t, Y_t in each simulated dataset have size $I \in \{10, 20, 30, 40\}$, and $T = 60$ in each simulation. We initialized the Gibbs sampler by setting the PARAFAC marginals $\beta_1^{(r)}, \beta_2^{(r)}, \beta_3^{(r)}$, $r = 1, \dots, R$ (with $R = 5$), via simulated annealing (see section S.7). We chose a burn-in period of 10,000 iterations and, due to autocorrelation in the sample, we applied thinning and selected every 2nd iteration, thus obtaining 5,000 draws from the posterior distribution after convergence.

Fig. 5 shows the accuracy of the sampler in estimating the coefficient tensor, in the four experiments corresponding to $I \in \{10, 20, 30, 40\}$. The efficiency

[§]<http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html>

decreases with I (recall that the number of cells of the coefficient tensor is I^4). The estimation error is mainly due to the over-shrinking to zero, which is a known drawback of global-local hierarchical prior distributions (e.g., see Carvalho et al. (2010)). Note that we expected a decrease of efficiency with I , since the sample size was held fixed ($T = 60$) across all simulation experiments, while increasing the size of the parameter space. In Figg. 6, 8, 10, 12 we report the estimation results for some randomly chosen cells of the coefficient tensor. We find that, after removing burn-in iterations and performing thinnig, the autocorrelation wipes out.

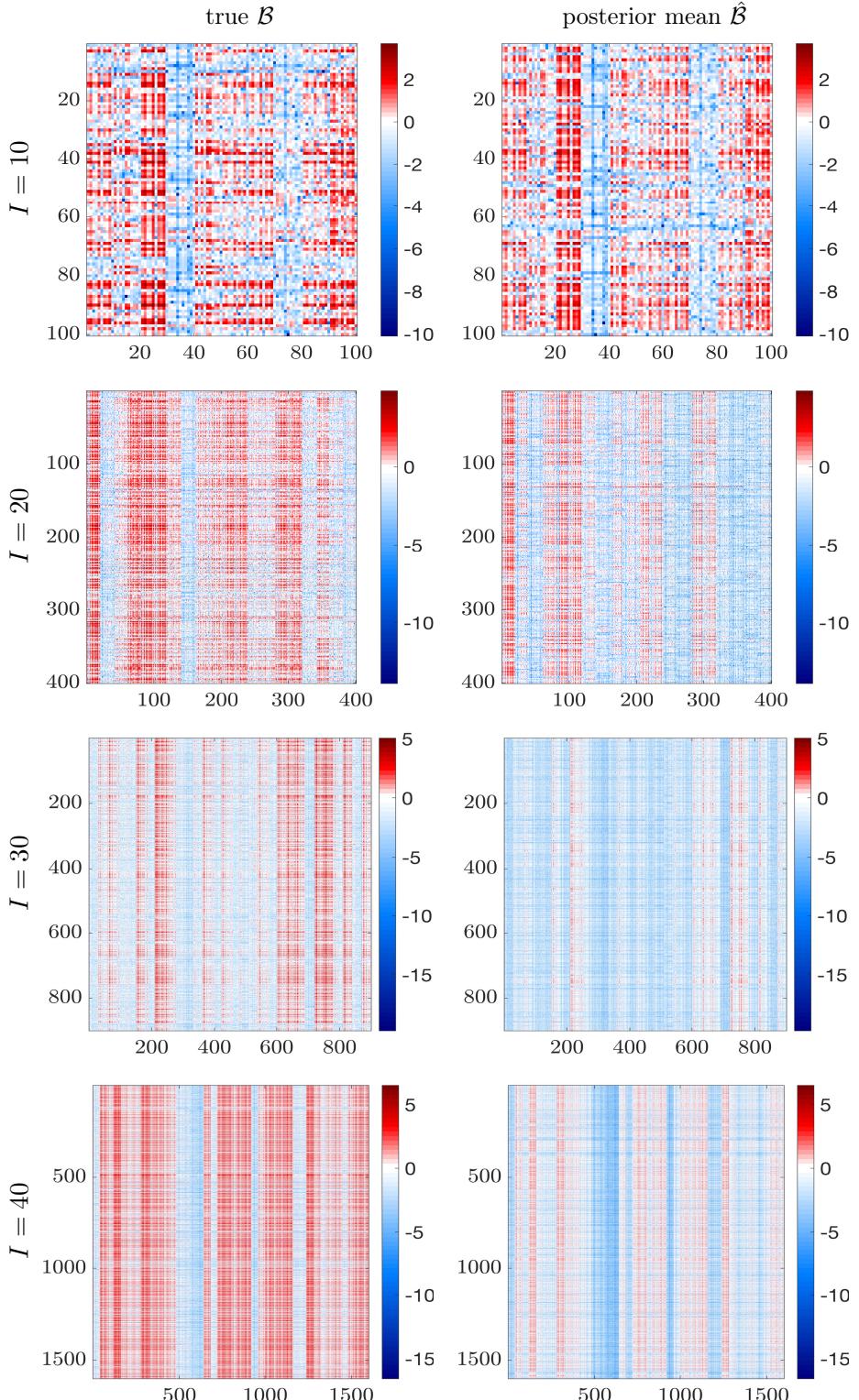


Figure 5: Logarithm of the absolute value of the coefficient tensors (in matricized form): true \mathcal{B} (left) and posterior mean estimate $\hat{\mathcal{B}}$ (right), for four experiments with different size I (in row).

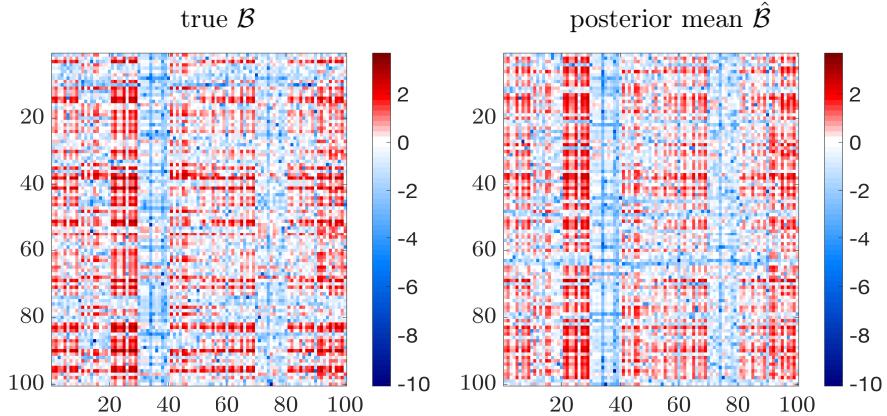
194 S.8.1. Experiment: $I=10$ 

Figure 6: Experiment $I = 10$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true \mathcal{B} (left) and posterior mean estimate $\hat{\mathcal{B}}$ (right).

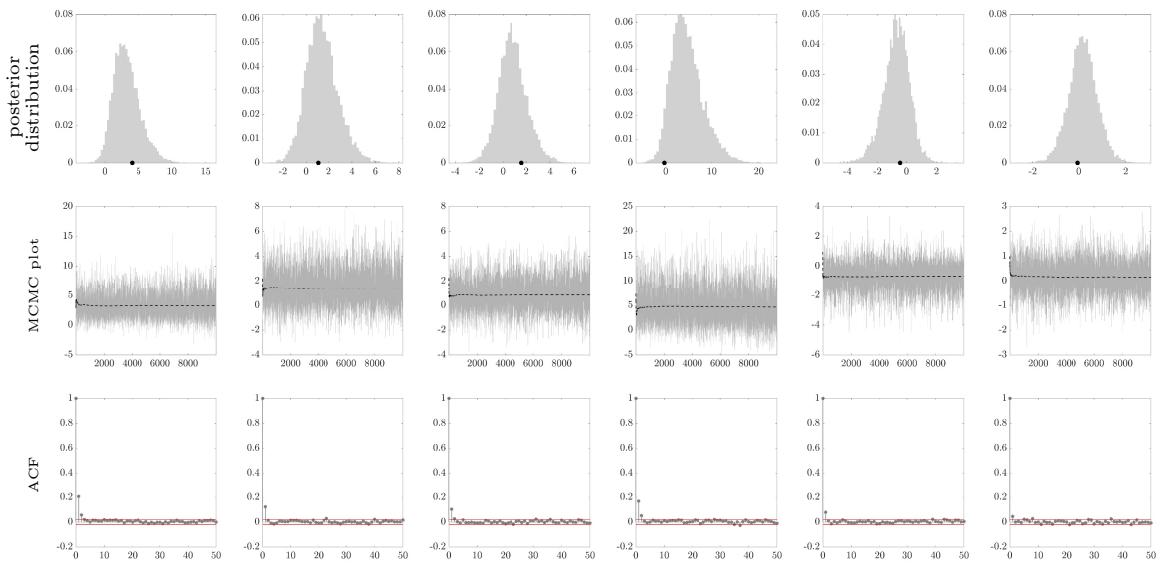


Figure 7: Experiment $I = 10$. Posterior distribution (first row, the black dot is the true value), MCMC plot (second row, dashed line represents the progressive mean) and autocorrelation function (third row) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each column).

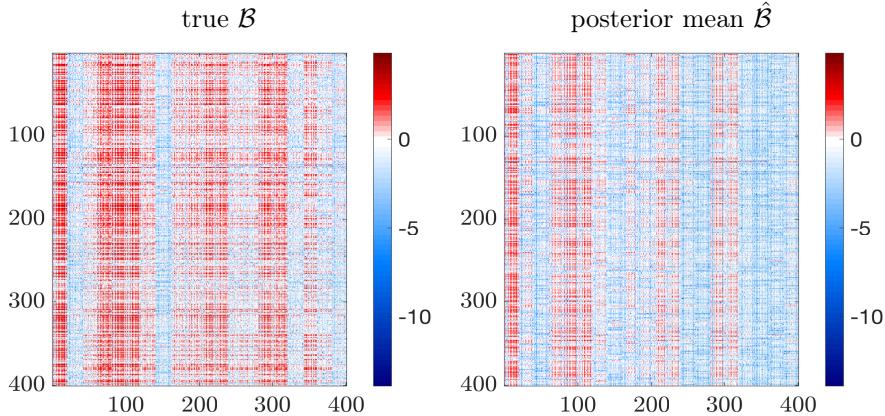
195 S.8.2. Experiment: $I=20$ 

Figure 8: Experiment $I = 20$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true \mathcal{B} (left) and posterior mean estimate $\hat{\mathcal{B}}$ (right).

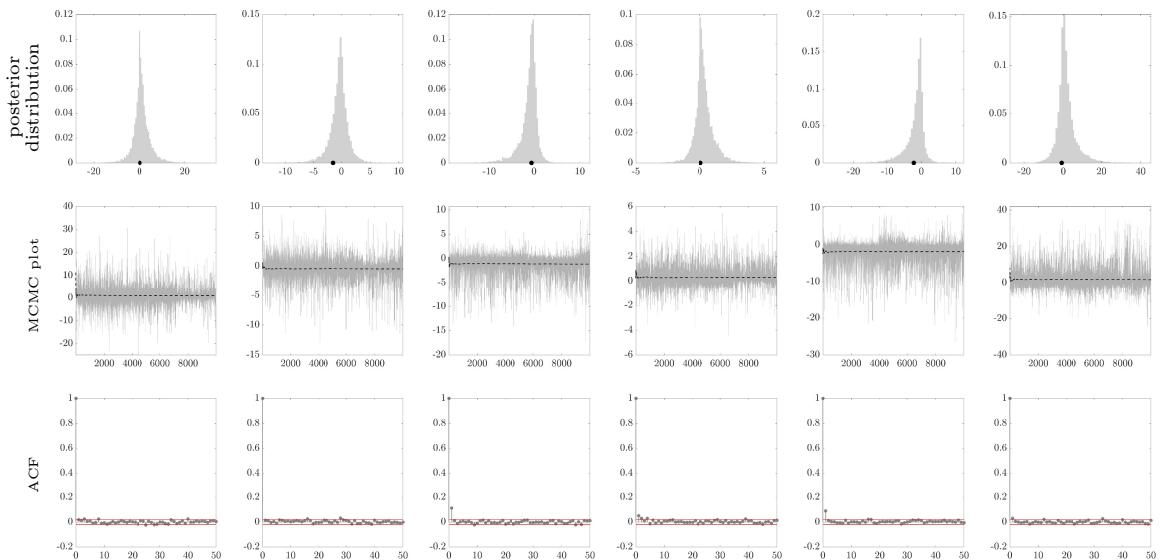


Figure 9: Experiment $I = 20$. Posterior distribution (first row, the black dot is the true value), MCMC plot (second row, dashed line represents the progressive mean) and autocorrelation function (third row) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each column).

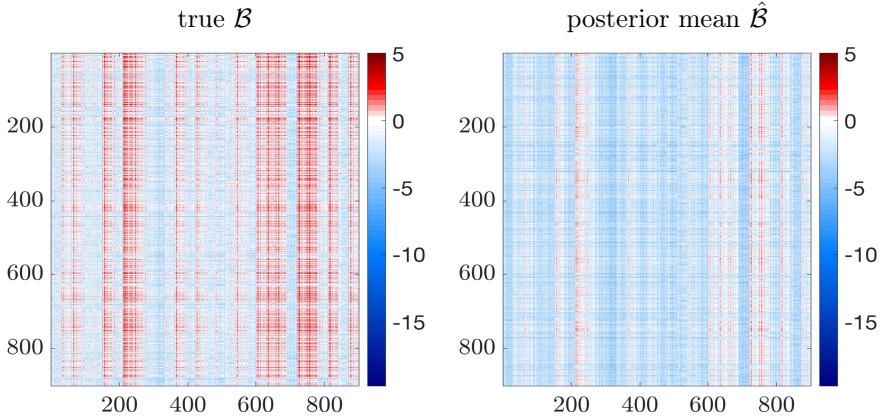
196 S.8.3. Experiment: $I=30$ 

Figure 10: Experiment $I = 30$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true \mathcal{B} (left) and posterior mean estimate $\hat{\mathcal{B}}$ (right).

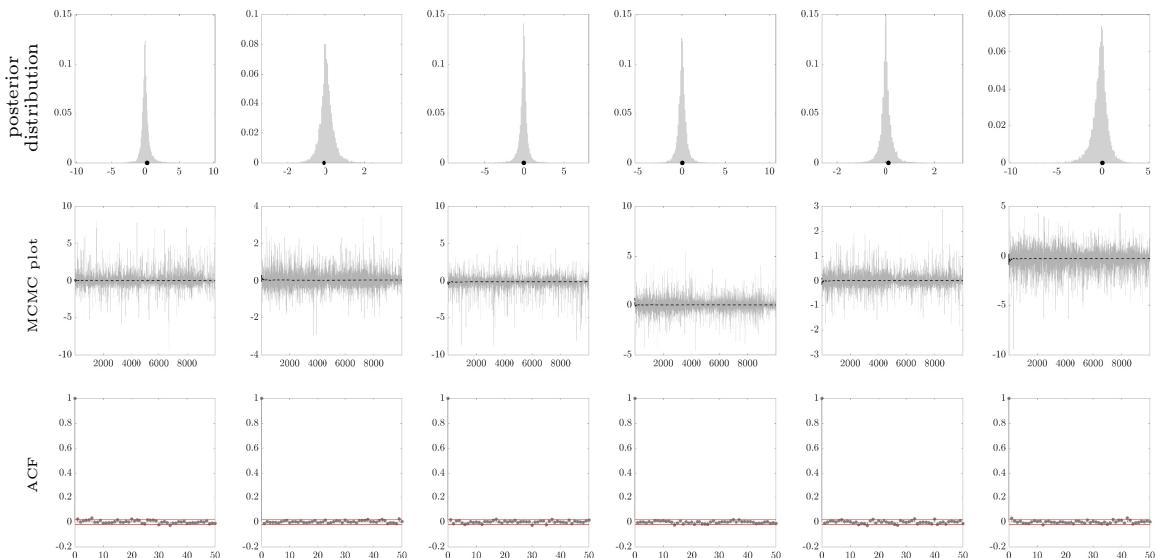


Figure 11: Experiment $I = 30$. Posterior distribution (first row, the black dot is the true value), MCMC plot (second row, dashed line represents the progressive mean) and autocorrelation function (third row) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each column).

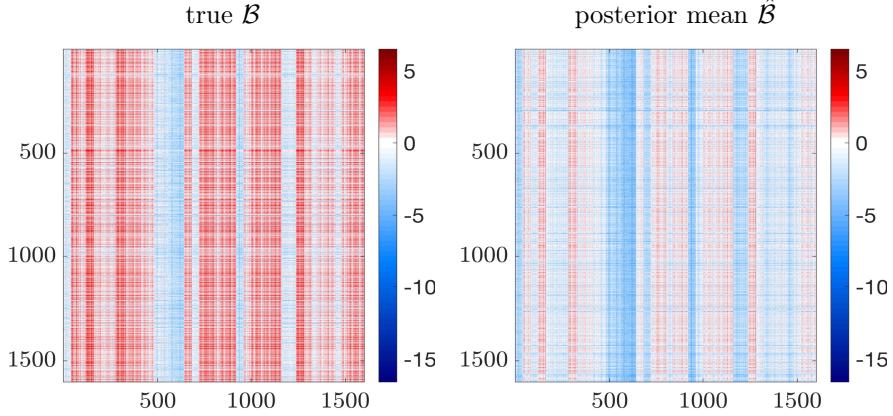
197 S.8.4. Experiment: $I=40$ 

Figure 12: Experiment $I = 40$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true \mathcal{B} (left) and posterior mean estimate $\hat{\mathcal{B}}$ (right).

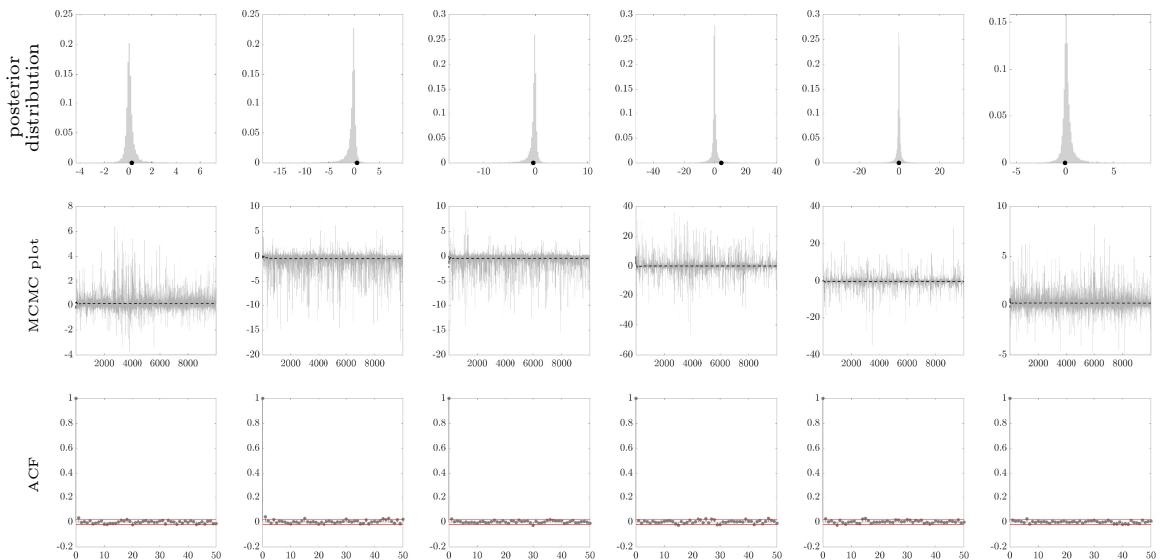


Figure 13: Experiment $I = 40$. Posterior distribution (first row, the black dot is the true value), MCMC plot (second row, dashed line represents the progressive mean) and autocorrelation function (third row) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each column).

198 **S.9. Data Description**

199 As put forward by Schweitzer et al. (2009), the analysis of economic networks is
 200 one of the most recent and complex challenges that the econometric community
 201 is facing nowadays. We contribute to the econometric literature about complex
 202 networks by applying the proposed methodology to the study jointly the
 203 dynamics of international trade and credit networks. The international trade and
 204 financial networks have been previously studied by several authors (e.g., Fiebler
 205 (2011), Eaton and Kortum (2002), Anundsen et al. (2016), Hidalgo and
 206 Hausmann (2009), Fagiolo et al. (2009), Kharrazi et al. (2017), Meyfroidt et al.
 207 (2010), Zhu et al. (2014), Squartini et al. (2011)), who investigated its topological
 208 properties and identified its main communities. To the best of our knowledge,
 209 this is the first attempt to model the dynamics of two networks jointly.

210 The bilateral trade data come from the United Nations COMTRADE
 211 database¶, whereas the data on bilateral outstanding capital come from the Bank
 212 of International Settlements database||, both are publicly available resources. For
 213 each couple (i, j) of countries, the international trade data from COMTRADE
 214 report total exports from country i to country j occurred during year t , while
 215 the BIS dataset gives the total amount of claims (i.e., credit) of country i vis-
 216 à-vis country j in year t . We use a subset of the COMTRADE database. Our
 217 sample of yearly observations for 10 countries ($I_1 = I_2 = I = 10$) runs from 2003
 218 to 2016. In order to remove potential non-linearities in the data, we take the
 219 logarithm all variables of interest. We thus consider the international trade and
 220 financial network in each period as one observation from a real-valued tensor-
 221 valued stochastic process. To sum up, our dataset consists in a 3-order tensor-
 222 valued time series of length $T = 13$. At each time t , the 3-order tensor \mathcal{Y}_t has
 223 dimension (I_1, I_2, I_3) , with $I_1 = I_2 = I = 10$ and $I_3 = 2$, and it represents a
 224 2-layer node-aligned network (or multiplex) with 10 vertices (countries), where
 225 each edge is given by a bilateral trade flow or financial stock. The entry $(i, j, 1, t)$
 226 of \mathcal{Y}_t reports the total exports of country i vis-à-vis country j , in year t , whereas
 227 entry $(i, j, 2, t)$ contains the total outstanding credit from country i towards
 228 country j , in year t . The series $(\mathcal{Y}_t)_t$, $t = 1, \dots, T$, has been standardized (over
 229 the temporal dimension).

230 **S.10. Additional results for the empirical application**

231 **S.10.1. Estimation results**

232 Fig. 16 shows the estimated covariance matrices. In all cases, the highest values
 233 correspond to individual variances, while the estimated covariances are lower
 234 in magnitude and heterogeneous. We also find evidence of heterogeneity in the
 235 dependence structure, since Σ_1 , which captures the covariance between rows (i.e.,

¶<https://comtrade.un.org>

||<http://stats.bis.org/statx/toc/LBS.html>

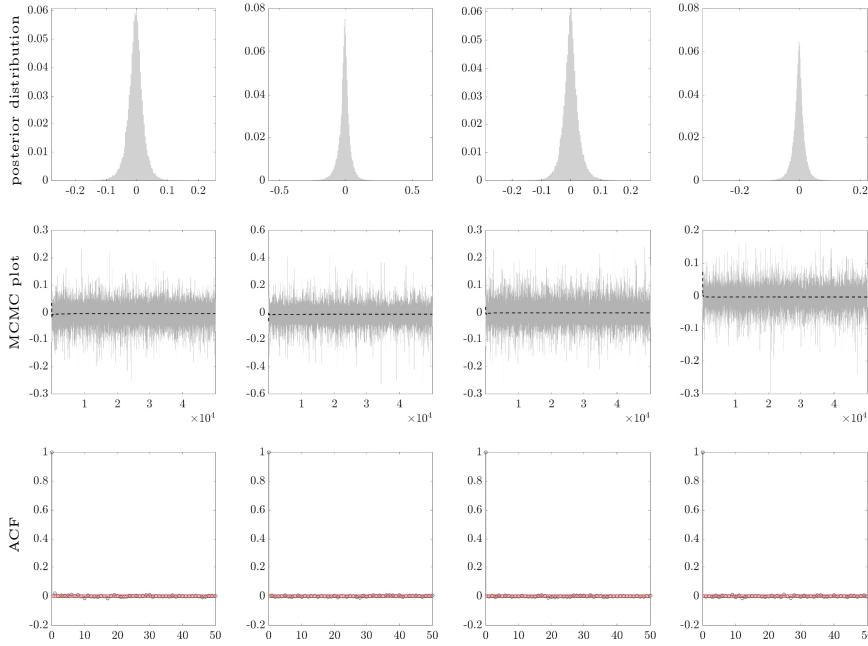


Figure 14: Posterior distribution (*first row*), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for four randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each *column*).

236 exporting and creditor countries), differs from Σ_2 , which describes the covariance
 237 between columns (i.e., importing and debtor countries). With few exceptions,
 238 estimated covariances are positive.

239 *S.10.2. Impulse response analysis*

240 Fig. 17 shows the block Cholesky IRF at horizon $h = 1, 2$, resulting from a
 241 negative 1% shock to GB's outstanding debt**. The main findings follow.

242 *Global effect* on the network. We observe heterogeneous effects across
 243 countries. Effects on the trade layer at horizon 1 are equally heterogeneous,
 244 but smaller in magnitude compared with the financial layer.

245 *Local effect on Germany.* Compared with other countries, the shock has
 246 smaller effects on Germany's trade. The negative shock to GB's outstanding
 247 debt has a negative impact on Germany's exports and imports to all countries
 248 but Ireland and Sweden for exports and Denmark for imports. Germany's
 249 outstanding credit increases vis-à-vis Denmark, GB, Japan and US. Germany's

**Again, the shock is allocated across countries to reflect country-specific shares of the last period in the sample.

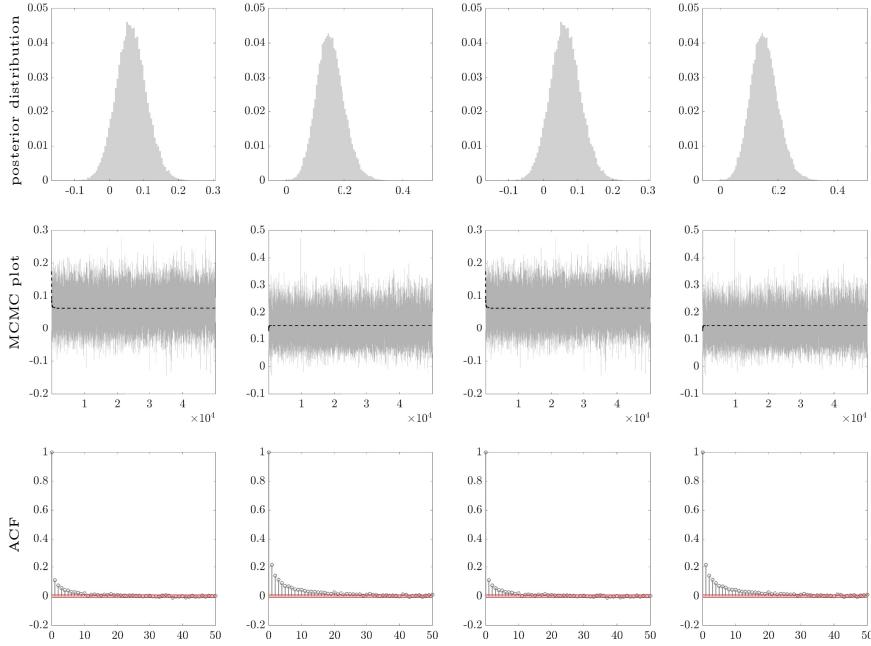


Figure 15: Posterior distribution (*first row*), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for two randomly chosen cells of the estimated covariance matrix $\hat{\Sigma}_1$ (*first and second column*) and $\hat{\Sigma}_2$ (*third and fourth column*).

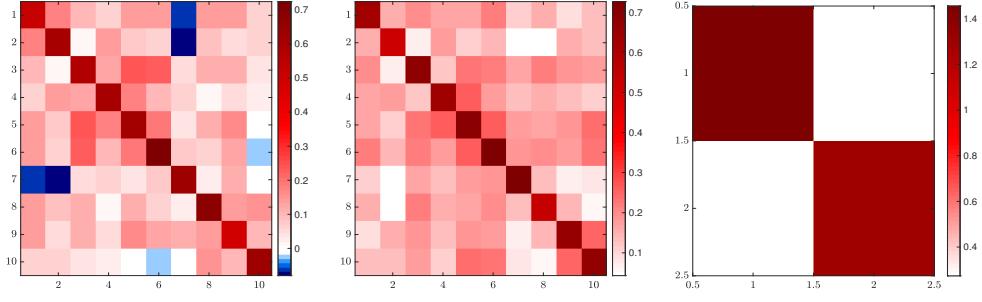


Figure 16: Estimated covariance matrices: $\hat{\Sigma}_1$ (*left*), $\hat{\Sigma}_2$ (*center*), $\hat{\Sigma}_3$ (*right*).

outstanding debt increases against all countries but Denmark and Sweden, in particular against France, Japan and Ireland. At horizon 2 responses are not reverted, but nearly all effects turn insignificant, providing evidence of monotone and fast decay of the IRFs.

Local effect on other countries. On the trade layer at horizon 1, we observe a positive response in Denmark's exports and on average a negative response of Switzerland's, Ireland's and Japan's exports. France and Sweden are the most

257 affected countries on the financial layer: The increase in outstanding credit of
 258 France towards Germany, Denmark and GB is counterbalanced by a reduction
 259 in Sweden's outstanding credit towards the same countries. We observe reverse
 260 effects concerning France's and Sweden's outstanding credit towards Switzerland
 261 and Ireland. Finally, Ireland's outstanding credit reacts positively towards most
 262 other countries.

263 Compared with responses to the shock to US imports, the persistence of
 264 a negative shock to GB's outstanding debt is slightly stronger, see impulse
 265 responses at horizon 2 in Fig. 17. The decay is monotonic. However, the speed
 266 of decay is heterogeneous across countries. For some countries, there are small
 267 effects at horizon 2, while for others the effects are completely wiped already.
 268 Overall, we do not find evidence of a relation between the size of a country
 269 in terms of exports or outstanding credit and the persistence in the impulse
 270 response. At the most, persistence seems determined by the origin of the shock,
 271 the effects of a financial shock being more persistent than those of a trade shock.

272 Finally, in Fig. 18 we plot the block Cholesky IRF, respectively, at horizon
 273 $h = 1, 2$, resulting from a 1% negative shock to GB's outstanding debt coupled
 274 with a 1% positive shock to GB's outstanding credit. The main findings follow.

275 *Global effect* on the network. The results remarkably differ from the previous
 276 ones (see Fig. 17). The responses to this simultaneous shock in GB's outstanding
 277 debt and credit are larger, in particular in the trade layer. However, already at
 278 horizon 2 responses are nearly fully decayed. The results in Fig. 17 and Fig. 18
 279 suggest that an increase in GB's outstanding credit has an overall positive effect
 280 on trade, stimulating export/import activities of most other countries.

281 *Local effect on Germany.* One period after the shock, we observe an overall
 282 positive effect on German exports, the exception being towards GB, Ireland and
 283 Sweden. Imports react mostly positively. Imports from US and Ireland react
 284 most, while those from Denmark react negatively. The responses of Germany's
 285 outstanding debt vis-à-vis most countries but Denmark and Sweden are negative,
 286 especially against France. At horizon 2 Germany's responses have nearly faded
 287 away, suggesting a rapid monotone decay of the shock's effect.

288 *Local effect* on other countries. In particular, the reactions of Switzerland's
 289 imports and outstanding debt are strikingly different from the previous case,
 290 compare with Fig. 17. Imports from US and Ireland, and to a lesser extent
 291 from France and Austria, are strongly boosted, while those from Denmark and
 292 Sweden decrease strongly. Moreover, we note that Japan's outstanding debt
 293 increases significantly against most countries. We interpret this as a signal for
 294 Japan's attractiveness for foreign capital. Compared with the previous exercise,
 295 France's financial responses are now mostly insignificant, or of opposite sign.
 296 Finally, the reactions of GB's exports and outstanding credit are heterogeneous,
 297 the latter ones being larger in absolute magnitude.

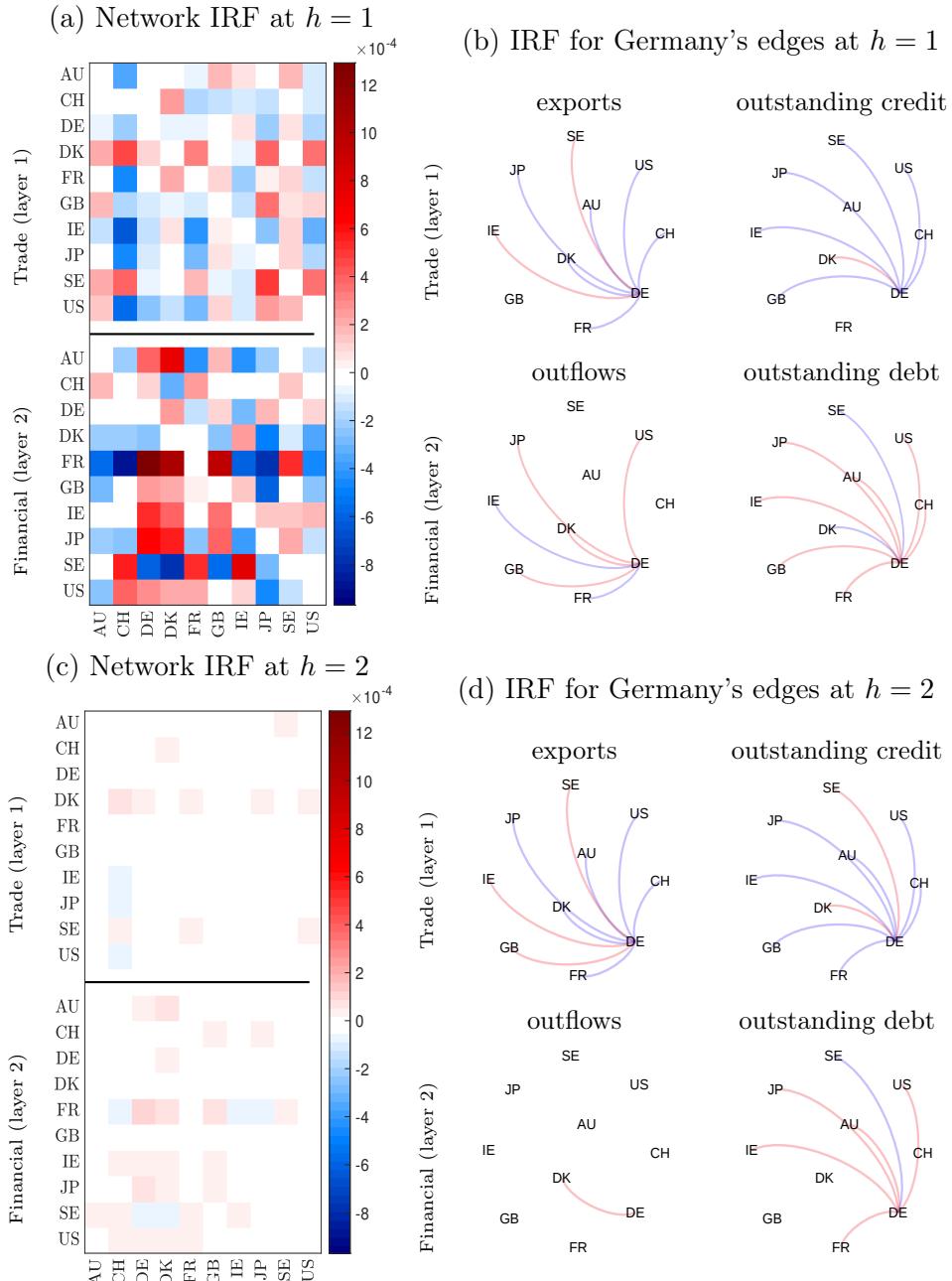


Figure 17: Shock to GB capital inflows by -1%. IRF at horizon $h = 1$ for all (panel a) and Germany (panel b) financial and trade transactions. IRF at horizon $h = 2$ for all (panel c) and Germany (panel d) financial and trade transactions. In each plot negative coefficients are in blue and positive in red.

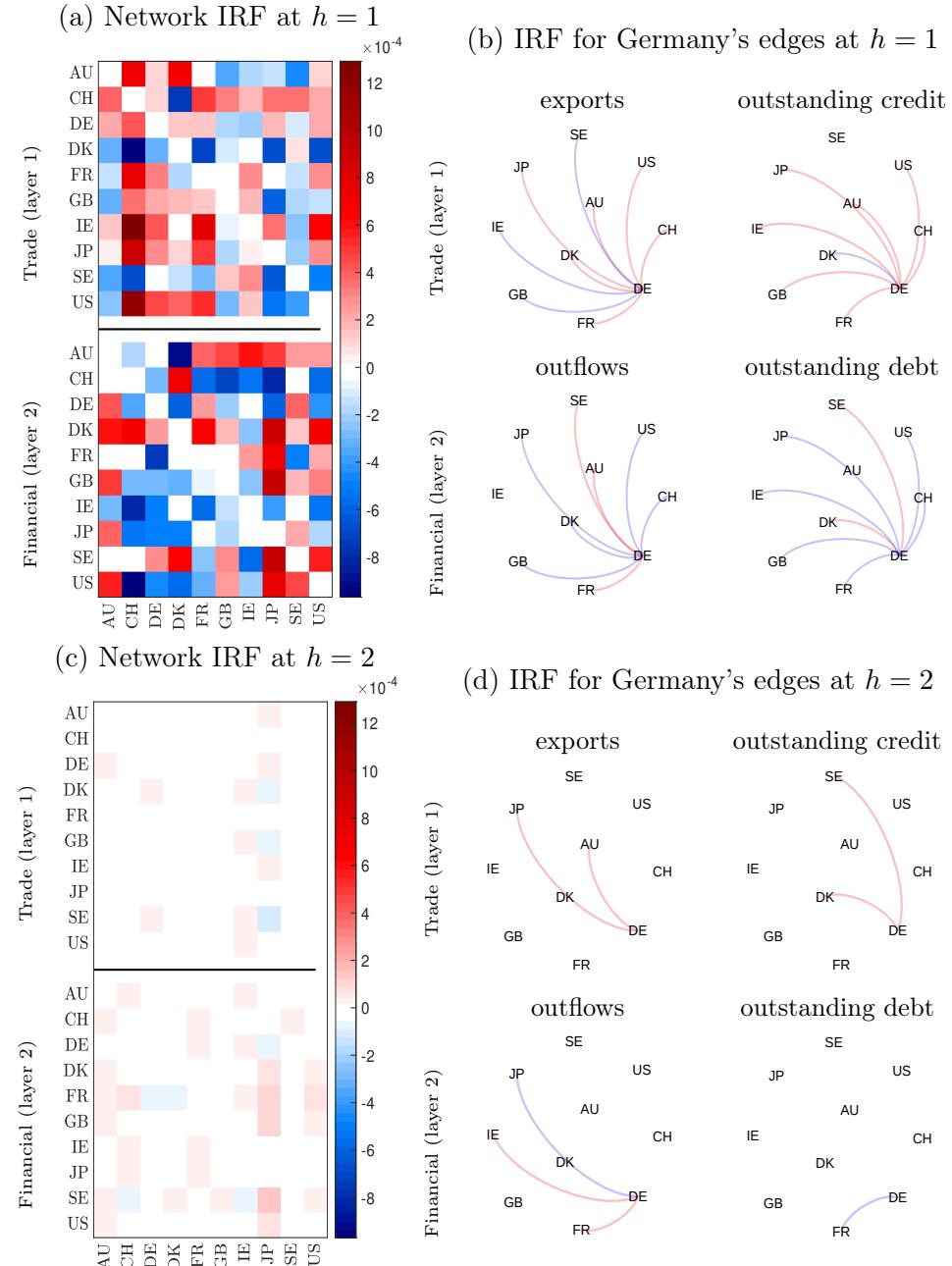


Figure 18: Shock to GB capital inflows by -1% and outflows by $+1\%$. IRF at horizon $h = 1$ for all (panel a) and Germany (panel b) financial and trade transactions. IRF at horizon $h = 2$ for all (panel c) and Germany (panel d) financial and trade transactions. In each plot negative coefficients are in blue and positive in red.

298 **References**

- 299 Anundsen, A. K., K. Gerdrup, F. Hansen, and K. Kragh-Sørensen (2016).
300 Bubbles and crises: The role of house prices and credit. *Journal of Applied
301 Econometrics* 31(7), 1291–1311.
- 302 Baştürk, N., L. Hoogerheide, and H. K. van Dijk (2017). Bayesian analysis of
303 boundary and near-boundary evidence in econometric models with reduced
304 rank. *Bayesian Analysis* 12(3), 879–917.
- 305 Canova, F. and M. Ciccarelli (2004). Forecasting and turning point predictions
306 in a Bayesian panel VAR model. *Journal of Econometrics* 120(2), 327–359.
- 307 Canova, F. and M. Ciccarelli (2009). Estimating multicountry VAR models.
308 *International Economic Review* 50(3), 929–959.
- 309 Canova, F., M. Ciccarelli, and E. Ortega (2007). Similarities and convergence
310 in G-7 cycles. *Journal of Monetary Economics* 54(3), 850–878.
- 311 Carriero, A., G. Kapetanios, and M. Marcellino (2016). Structural analysis with
312 multivariate autoregressive index models. *Journal of Econometrics* 192(2),
313 332 – 348.
- 314 Carroll, J. D. and J.-J. Chang (1970). Analysis of individual differences
315 in multidimensional scaling via an N-way generalization of “Eckart-Young”
316 decomposition. *Psychometrika* 35, 283–319.
- 317 Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator
318 for sparse signals. *Biometrika* 97(2), 465–480.
- 319 Eaton, J. and S. Kortum (2002). Technology, geography, and trade.
320 *Econometrica* 70(5), 1741–1779.
- 321 Engle, R. F. and C. W. Granger (1987). Co-integration and error correction:
322 representation, estimation, and testing. *Econometrica*, 251–276.
- 323 Fagiolo, G., J. Reyes, and S. Schiavo (2009). World-trade web: topological
324 properties, dynamics, and evolution. *Physical Review E* 79(3), 036115.
- 325 Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a
326 quantitative explanation. *Econometrica* 79(4), 1069–1101.
- 327 Guhaniyogi, R., S. Qamar, and D. B. Dunson (2017). Bayesian tensor regression.
328 *Journal of Machine Learning Research* 18(79), 1–31.
- 329 Harshman, R. A. (1970). Foundations of the PARAFAC procedure: models and
330 conditions for an “explanatory” multi-modal factor analysis. *UCLA Working
331 Papers in Phonetics*, 16, 1- 84.

- 332 Hidalgo, C. A. and R. Hausmann (2009). The building blocks of economic
333 complexity. *Proceedings of the national academy of sciences* 106(26), 10570–
334 10575.
- 335 Kharrazi, A., E. Rovenskaya, and B. D. Fath (2017). Network structure impacts
336 global commodity trade growth and resilience. *PloS one* 12(2), e0171184.
- 337 Kiers, H. A. (2000). Towards a standardized notation and terminology in
338 multiway analysis. *Journal of Chameometrics* 14(3), 105–122.
- 339 Kolda, T. G. (2006). Multilinear operators for higher-order decompositions.
340 Technical report, Sandia National Laboratories.
- 341 Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications.
342 *SIAM Review* 51(3), 455–500.
- 343 Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in
344 nonlinear multivariate models. *Journal of Econometrics* 74(1), 119–147.
- 345 Kruijer, W., J. Rousseau, and A. Van Der Vaart (2010). Adaptive Bayesian
346 density estimation with location-scale mixtures. *Electronic Journal of
347 Statistics* 4, 1225–1257.
- 348 Lanne, M. and H. Nyberg (2016). Generalized forecast error variance
349 decomposition for linear and nonlinear multivariate models. *Oxford Bulletin
350 of Economics and Statistics* 78(4), 595–603.
- 351 Meyfroidt, P., T. K. Rudel, and E. F. Lambin (2010). Forest transitions, trade,
352 and the global displacement of land use. *Proceedings of the National Academy
353 of Sciences* 107(49), 20917–20922.
- 354 Pan, R. (2014). Tensor transpose and its properties. *arXiv preprint
355 arXiv:1411.1503*.
- 356 Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007).
357 *Numerical recipes: the art of scientific computing*. Cambridge University Press.
- 358 Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods*. Springer.
- 359 Schotman, P. and H. K. Van Dijk (1991). A Bayesian analysis of the unit root
360 in real exchange rates. *Journal of Econometrics* 49(1-2), 195–238.
- 361 Schweitzer, F., G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani,
362 and D. R. White (2009). Economic networks: the new challenges.
363 *Science* 325(5939), 422–425.
- 364 Springer, M. and W. Thompson (1970). The distribution of products of
365 beta, gamma and Gaussian random variables. *SIAM Journal on Applied
366 Mathematics* 18(4), 721–737.

- 367 Squartini, T., G. Fagiolo, and D. Garlaschelli (2011). Randomizing world trade.
368 i. a binary network analysis. *Physical Review E* 84(4), 046117.
- 369 Zellner, A. (1962). An efficient method of estimating seemingly unrelated
370 regressions and tests for aggregation bias. *Journal of the American Statistical
371 Association* 57(298), 348–368.
- 372 Zhu, Z., F. Cerina, A. Chessa, G. Caldarelli, and M. Riccaboni (2014). The
373 rise of China in the international trade network: a community core detection
374 approach. *PloS one* 9(8), e105496.