Bayesian Dynamic Tensor Regression

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

[1]Ca' Foscari University of Venice

[2]Scuola Normale Superiore of Pisa

[3]Study Center Gerzensee, Foundation of the Swiss National Bank

Abstract

This appendix contains background results on tensors in Section S.1, and the derivation of the tensor forecast error variance decomposition in Section S.2. An example of MAR in Section S.3 and proofs of the remarks in Section 2 of the main paper in Section S.4. Details on the prior on tensor entries are given in Section S.5. Also, Section S.6 reports the details on posterior computation and Section S.7 describes the initialisation of the inferential algorithm. A summary of simulation results is provided in Section S.8. The data used in the empirical application is described in Section S.9.

S.1 Background Material on Tensor Calculus

This section provides additional results on tensor operators.

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

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

*e-mail: billio@unive.it

†e-mail: r.casarin@unive.it

‡e-mail: matteo.iacopini@unive.it

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

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

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

$$[\![\mathcal{Y}; A_1, \ldots, A_N]\!] = \mathcal{Y} \bar{\times}_1 A_1 \bar{\times}_1 A_2 \ldots \bar{\times}_1 A_N,$$

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

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

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

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

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

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

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

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

Definition S.1.2 (Kruskal operator)**.**
Let $\mathbf{N} = \{1, \ldots, N\}$ *and* $(A_n)_n$ *be a collection of* N *matrices such that* $A_n \in \mathbb{R}^{I_n \times R}$

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

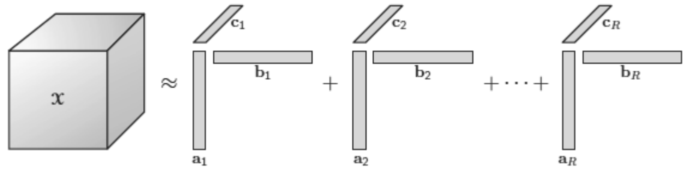

Figure 1: PARAFAC decomposition of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with $\mathbf{a}_r \in \mathbb{R}^{I_1}$, $\mathbf{b}_r \in \mathbb{R}^{I_2}$ and $\mathbf{c}_r \in \mathbb{R}^{I_3}$, $r = 1, \ldots, R$. Figure from Kolda and Bader (2009).

for $n \in \mathbf{N}$. Let $\mathcal{I}$ be the identity tensor of size $R \times \ldots \times R$, i.e. a tensor having ones along the superdiagonal and zeros elsewhere. The Kruskal operator is defined as

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

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

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

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 $A \odot^K B = (\mathbf{a}_{:,1} \otimes \mathbf{b}_{:,1}, \ldots, \mathbf{a}_{:,K} \otimes \mathbf{b}_{:,K})$) in combination with the mode-n matricization and the vectorization operators, we get the following additional representations of $\mathcal{X} = [\![A_1, \ldots, A_N]\!]$

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

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 $I_1 \times \ldots \times I_N$ and let $I^* = \prod_{i=1}^{N} I_i$. Then there exists a $I^* \times I^*$ vec-permutation (or commutation) matrix $K_{1 \to n}$ such that

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

where $\mathbf{X}_{(1)}^{T_\sigma} = (\mathcal{X}^{T_\sigma})_{(1)} = \mathbf{X}_{(n)}$ is the mode-1 matricization of the transposed tensor $\mathcal{X}^{T_\sigma}$ 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 $\mathscr{R}=\{r_1,\dots,r_m\}\subset\mathbf{D}$ and $\mathscr{C}=\{c_1,\dots,c_p\}\subset\mathbf{N}$, i.e. $\mathbf{N}=\mathscr{R}\cup\mathscr{C}$, $\mathscr{R}\cap\mathscr{C}=\emptyset$ and $N=m+p$. Then

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

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

Proof. We demonstrate the statement for $\mathscr{R}=\{n\}$, $n\in\mathbf{N}$, however the results follows from the same steps also in the general case $\#\mathscr{R}>1$. The strategy 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

$$\begin{aligned}
&(2\pi)^{-\frac{\prod_i I_i}{2}}\big|\Sigma_1\big|^{-\frac{I_{-1}}{2}}\cdots\big|\Sigma_n\big|^{-\frac{I_{-n}}{2}}\cdots\big|\Sigma_N\big|^{-\frac{I_{-N}}{2}}= && \text{(S4)}\\
&=(2\pi)^{-\frac{\prod_i I_i}{2}}\big|\Sigma_1\big|^{-\frac{I_{-1}}{2}}\cdots\big|\Sigma_{n-1}\big|^{-\frac{I_{-(n-1)}}{2}}\big|\Sigma_{n+1}\big|^{-\frac{I_{-(n+1)}}{2}}\cdots\big|\Sigma_N\big|^{-\frac{I_{-N}}{2}}\big|\Sigma_n\big|^{-\frac{I_{-n}}{2}}\\
&=(2\pi)^{-\frac{\prod_i I_i}{2}}\big|\Sigma_N\otimes\ldots\otimes\Sigma_{n-1}\otimes\Sigma_{n+1}\otimes\ldots\otimes\Sigma_N\big|^{-\frac{n}{2}}\big|\Sigma_n\big|^{-\frac{I_{-n}}{2}}. && \text{(S5)}
\end{aligned}$$

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

$$\mathcal{Y}\bar{\times}_N\mathcal{U}\bar{\times}_N\mathcal{Y}=\sum_{i_1,\dots,i_n,\dots,i_N}\sum_{i'_1,\dots,i'_N} y_{i_1,\dots,i_N}(u^{-1}_{i_1,i'_1}\cdot\ldots\cdot u_{i_n,i'_n}\cdot\ldots\cdot u^{-1}_{i_N,i'_N})y_{i'_1,\dots,i'_n,\dots,i'_N}.$$

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

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

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

$$\begin{aligned}
\mathcal{Y} \bar{\times}_N \mathcal{U} \bar{\times}_N \mathcal{Y} &= \mathrm{vec}\left(\mathcal{Y}\right)' \cdot \mathcal{U}_{(\mathscr{S}_1,\mathscr{S}_2)} \cdot \mathrm{vec}\left(\mathcal{Y}\right) \qquad \text{(S6)} \\
&= \mathrm{vec}\left(\mathcal{Y}\right)' \cdot \big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_n^{-1} \otimes \ldots \otimes \Sigma_1^{-1}\big) \cdot \mathrm{vec}\left(\mathcal{Y}\right) \\
&= \mathrm{vec}\left(\mathcal{Y}^\sigma\right)' \cdot \big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_1^{-1} \otimes \Sigma_n^{-1}\big) \cdot \mathrm{vec}\left(\mathcal{Y}^\sigma\right) \\
&= \mathrm{vec}\left(\mathbf{Y}_{(n)}\right)' \cdot \big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_1^{-1} \otimes \Sigma_n^{-1}\big) \cdot \mathrm{vec}\left(\mathbf{Y}_{(n)}\right) \\
&= \mathrm{vec}\left(\mathbf{Y}_{(n)}\right)' \cdot \mathrm{vec}\left(\Sigma_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot \big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_1^{-1}\big)\right) \\
&= \mathrm{tr}\left(\mathbf{Y}'_{(n)} \cdot \Sigma_n^{-1} \cdot \mathbf{Y}_{(n)} \cdot \big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_1^{-1}\big)\right) \\
&= \mathrm{tr}\left(\big(\Sigma_N^{-1} \otimes \ldots \otimes \Sigma_1^{-1}\big)\big(\mathbf{X}_{(n)} - \mathbf{M}_{(n)}\big)' \Sigma_n^{-1} \big(\mathbf{X}_{(n)} - \mathbf{M}_{(n)}\big)\right). \qquad \text{(S7)}
\end{aligned}$$

Since the term in (S4) and (S6) are the normalizing constant and the exponent of the tensor normal distribution, whereas (S5) and (S7) are the corresponding expressions for the desired matrix normal distribution, the result is proved for the case $\#\mathscr{R} = 1$. In the general case $\#\mathscr{R} = r > 1$ the proof follows from the same reasoning, by substituting the permutation σ with another permutation σ' which exchanges the modes of the tensor such that the first r modes of the transpose tensor $\mathcal{Y}^{\sigma'}$ correspond to the elements of $\mathscr{R}$. □

S.2 Forecast error variance decomposition

From the results in eqs. (14)-(15), 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,

$$\theta_{ij}(h) = \frac{\sum_{k=0}^{h} IRF^2_{ij}(k)}{\sum_{k=0}^{h}\sum_{j=0}^{I^*} IRF^2_{ij}(k)}, \qquad \theta^*_{i,j}(h) = \frac{\sum_{k=0}^{h}(\psi^G_{ij}(k;n))^2}{\sum_{k=0}^{h}\sum_{j=0}^{I^*}(\psi^G_{ij}(k;n))^2}.$$

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

$$\theta^O_{i,j}(h) = \frac{\sum_{k=0}^{h}(\psi^O_{ij}(k;n))^2}{\sum_{k=0}^{h}\sum_{j=0}^{I^*}(\psi^O_{ij}(k;n))^2} = \frac{\sum_{k=0}^{h}(\mathbf{e}_i'\Psi_k LP\mathbf{e}_j)^2}{\sum_{k=0}^{h}\mathbf{e}_i'(\Psi_k L)D(\Psi_k L)'\mathbf{e}_i}.$$

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

$$\theta^G_{i,j}(h) = \frac{\sum_{k=0}^{h}(\psi^G_{ij}(k;n))^2}{\sum_{k=0}^{h}\sum_{j=0}^{I^*}(\psi^G_{ij}(k;n))^2} = \frac{\sum_{k=0}^{h}(\mathbf{e}_i'\Psi_k LDD_{jj}^{-1/2}\mathbf{e}_j)^2}{\sum_{k=0}^{h}\mathbf{e}_i'(\Psi_k L)D(\Psi_k L)'\mathbf{e}_i} \tag{S8}$$

Finally, the modified tFEVD applied to the tensor GIRF (S8) yields

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

where $\Lambda = \mathrm{diag}(D_{11}^{-1}, \ldots, D_{I^*I^*}^{-1})$.

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

$\text{vec}\left(\mathcal{Y}_t\right) = \mathbf{y}_t$ and $\text{vec}\left(\mathcal{E}_t\right) = \boldsymbol{\epsilon}_t$, as follows

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

Therefore, model (5) becomes

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

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

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

where, for each $i = 1, \ldots, 4$, 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

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

S.4 Proofs

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

Proof of result in Remark 2.4. Consider model (5) when $I_j = 1$, for $j = 1, \ldots, N$. Note that a N-order tensor whose modes have all unit length is equivalent to a 1-order tensor, i.e. a scalar. As a consequence, the dependent variable becomes $y_t \in \mathbb{R}$ and the autoregressive coefficient tensors reduce to $\alpha_j \in \mathbb{R}$, $j = 0, \ldots, p$. The coefficient 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

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

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

$$\boldsymbol{\beta} = \sum_{r=1}^{R} \beta_1^{(r)} \circ \ldots \circ \beta_N^{(r)} \circ \boldsymbol{\beta}_{N+1}^{(r)} = \sum_{r=1}^{R} \beta_1^{(r)} \cdot \ldots \cdot \beta_N^{(r)} \cdot \boldsymbol{\beta}_{N+1}^{(r)}$$

since again the outer product reduces to the ordinary scalar multiplication and all

$\beta^{(r)}_{j,k}$, $k = 1, \ldots, N$, $r = 1, \ldots, R$ are scalars, while the marginal corresponding to the last mode $N+1$ is a vector of length J^*. $\square$

Proof of result in Remark 2.6. Consider model (5) with $I_1 = m$ and $I_j = 1$, for $j = 2, \ldots, N$. Denote by $\mathbf{x}_t = \text{vec}\left(\mathcal{X}_t\right)$ 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 $j = 1, \ldots, p$ 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

$$A_j = \sum_{r=1}^{R} \boldsymbol{\alpha}^{(r)}_{j,1} \circ \left(\alpha^{(r)}_{j,2} \cdot \ldots \cdot \alpha^{(r)}_{j,N-1}\right) \circ \boldsymbol{\alpha}^{(r)}_{N} = \sum_{r=1}^{R} A^{(r)}_j \cdot \left(\alpha^{(r)}_{j,2} \cdot \ldots \cdot \alpha^{(r)}_{j,N-1}\right).$$

Similarly, for the matrix B one gets

$$B = \sum_{r=1}^{R} \boldsymbol{\beta}^{(r)}_{1} \circ \left(\beta^{(r)}_{2} \cdot \ldots \cdot \beta^{(r)}_{N}\right) \circ \boldsymbol{\beta}^{(r)}_{N+1} = \sum_{r=1}^{R} B^{(r)} \cdot \left(\beta^{(r)}_{2} \cdot \ldots \cdot \beta^{(r)}_{N}\right).$$

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 $\mathcal{H} \in \{0,1\}^{I_1 \times \ldots \times I_N \times J}$, such that it has 0 only in those cells which are known to be null. Then $\bar{\mathcal{B}} = \mathcal{H} \odot \mathcal{B}$ 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.4.1 (follows from Remark 2.7)**.**
From the VECM in eq. (9) and denoting $\mathbf{y}_{t-1} = \mathrm{vec}\big(Y_{t-1}\big)$ *we can obtain an explicit form for the long run equilibrium (or cointegrating) relations, as follows*

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

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 thelong run cointegrating relationships, and the marginals* $(\boldsymbol{\gamma}_2^{(r)})_r$ *are the corresponding loadings.*

S.5 Prior distribution on tensor entries

Proof of Lemma 3.1. The distribution of each of these products has been characterised by Springer and Thompson (1970), who proved the 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* $x_h \sim \mathcal{N}(0, \sigma_h^2)$, $h = 1, \ldots, H$, *is proportional to a Meijer G-function*

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

where the normalising constant is

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

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

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

The integral is taken over a vertical line in the complex plane. Note that in the special case $H = 2$ we have $z \sim c_1P_1 - c_2P_2$, with $P_1, P_2 \sim \chi^2_1$ and $c_1 = Var(x_1+x_2)/4$, $c_2 = Var(x_1 - x_2)/4$. In this case, the resulting distribution is called product Normal distribution.

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

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

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

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

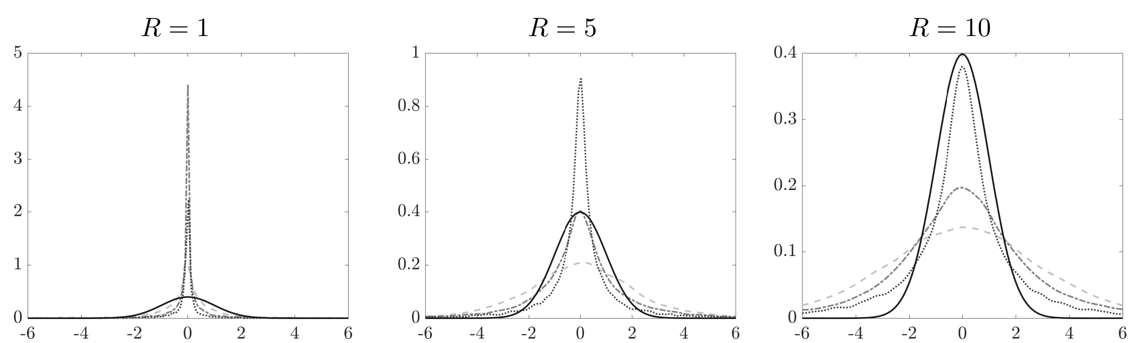

Figure 2: Monte Carlo simulation from the prior distribution of entry $b_{i_1,\ldots,i_N}$ of a generic N-order tensor, for varying rank R. In column: simulation with $R = 1$ (left), $R = 5$ (middle) and $R = 10$ (right). In all plots: standard Normal (continuous line) and prior for $b_{i_1,\ldots,i_N}$,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

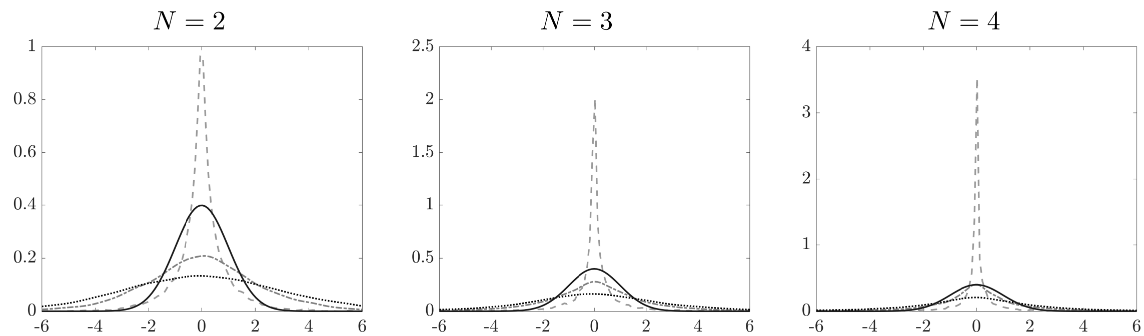

Figure 3: Monte Carlo simulation from the prior distribution of entry $b_{i_1,\dots,i_N}$ of a N-order tensor, with rank R, for varying N. In column: simulation with $N = 2$ (left), $N = 3$ (middle) and $N = 4$ (right). In all plots: standard Normal (continuous line) and prior for $b_{i_1,\dots,i_N}$,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

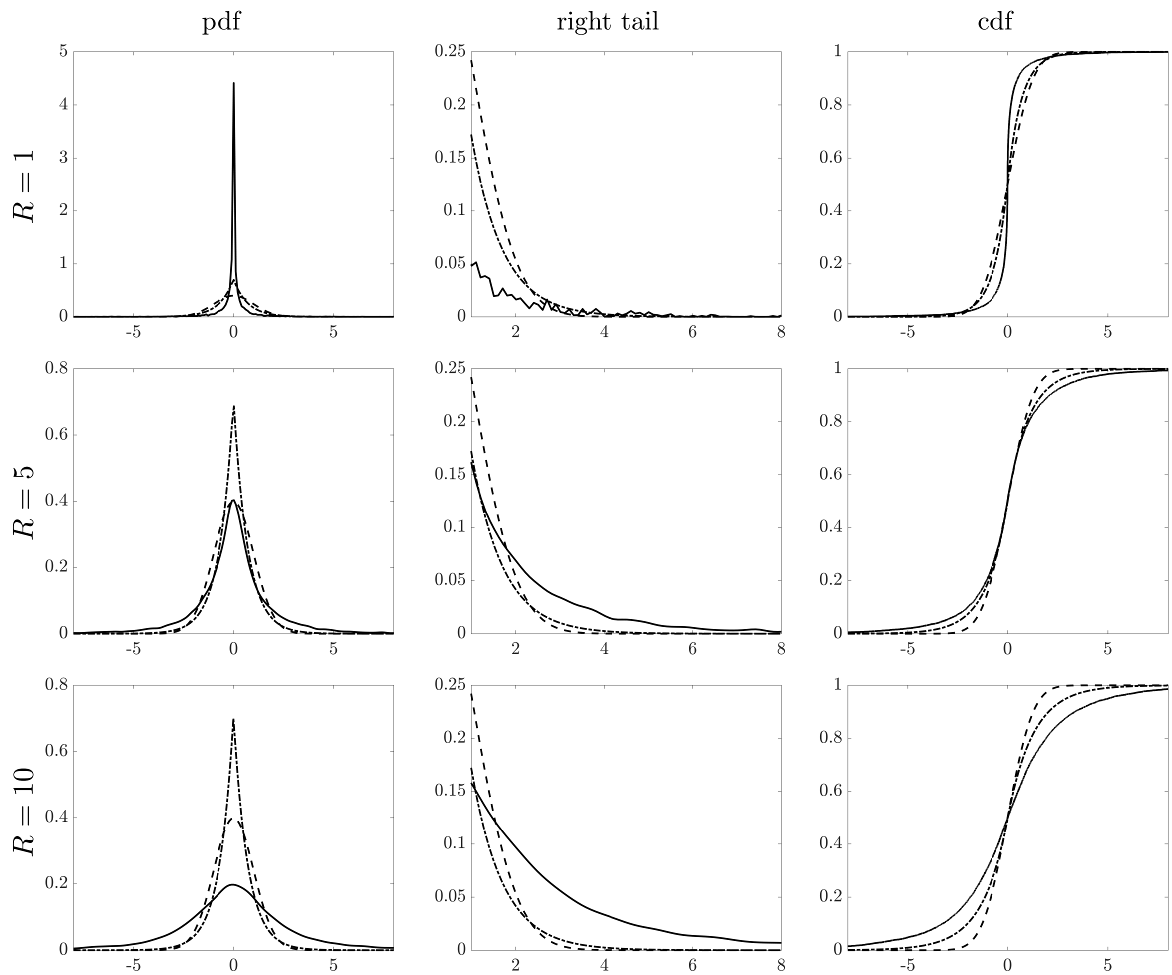

Figure 4: Monte Carlo simulation from the prior distribution of a generic 4-order tensor entry b_{ijkp} (continuous line), standard Normal distribution (dashed line) and standard Laplace distribution (dash-dotted line). In column: probability density function (*left*), right tail the probability density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

S.6 Computational details - TAR(1) model

In this section we will follow the convention of denoting the prior distributions with $\pi(\cdot)$. In addition, let $\mathbf{W} = (W_{j,r})_{j,r}$ be the collection of all (local variance) matrices $W_{j,r}$, for $j = 1, \dots, J$ and $r = 1, \dots, R$, let $I_0 = \sum_{j=1}^{J} I_j$ be the sum of the length of each mode of the tensor $\mathcal{B}$ and let $\mathbf{Y} = (\mathcal{Y}_t)_t$ the collection of observed variables. Recall that in the TAR(1) model in eq. (18), the variable $\mathcal{Y}_t$ is a 3-order tensor, thus we have $J = 4$.

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 $\boldsymbol{\phi}$ is

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

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

$$
\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{B},\mathbf{W}) &\propto \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau - 1 - \frac{RI_0}{2}} e^{-b_\tau \tau} \Big(\prod_{r=1}^R \exp\Big(-\frac{1}{2\tau\phi_r} C_r \Big)\Big) \mathrm{d}\tau \\
&= \prod_{r=1}^R \phi_r^{\alpha-1-\frac{I_0}{2}} \int_0^{+\infty} \tau^{a_\tau - 1 - \frac{Rd_0}{2}} \exp\Big(-b_\tau \tau - \sum_{r=1}^R \frac{C_r}{2\tau\phi_r} \Big) \mathrm{d}\tau. \qquad \text{(S9)}
\end{aligned}
$$

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

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

with $K_c(\cdot)$ a modified Bessel function of the second type. Our goal is to reconcile eq. (S9) to the kernel of this distribution. Since by definition $\sum_{r=1}^R \phi_r = 1$, it holds that $\sum_{r=1}^R (b_\tau \tau \phi_r) = (b_\tau \tau) \sum_{r=1}^R \phi_r = b_\tau \tau$. This allows to rewrite the exponential as

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

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

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

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

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

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

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

This is the kernel of a generalized inverse Gaussian

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

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 $\boldsymbol{\beta}_j^{(r)}$ defined in eq. (20) implies that each component follows a double exponential distribution

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

with probability density function given by

$$\pi(\beta_{j,p}^{(r)}|\lambda_{j,r},\phi_r,\tau) = \frac{\lambda_{j,r}}{2\sqrt{\tau\phi_r}} \exp\Big(- \frac{\big|\beta_{j,p}^{(r)}\big|}{(\lambda_{j,r}/\sqrt{\tau\phi_r})^{-1}}\Big). \tag{S10}$$

Then, exploiting the Gamma prior and eq. (S10)

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

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

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

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

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

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

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

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

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 $\mathcal{A} = \mathrm{vec}\left(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3\right)$. Then, from Lemma A.3 it holds

$$\begin{aligned}
\mathrm{vec}\left(\mathcal{A}\right) &= \mathrm{vec}\left(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3\right) = \boldsymbol{\alpha}_3 \otimes \mathrm{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right) \\
&= \boldsymbol{\alpha}_3 \otimes \left(\boldsymbol{\alpha}_2 \otimes \mathbf{I}_I\right) \mathrm{vec}\left(\boldsymbol{\alpha}_1\right) = \left(\boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \mathbf{I}_I\right)\boldsymbol{\alpha}_1 &\text{(S11)} \\
&= \boldsymbol{\alpha}_3 \otimes \left(\left(\mathbf{I}_J \otimes \boldsymbol{\alpha}_1\right) \mathrm{vec}\left(\boldsymbol{\alpha}_2'\right)\right) = \left(\boldsymbol{\alpha}_3 \otimes \mathbf{I}_J \otimes \boldsymbol{\alpha}_1\right)\boldsymbol{\alpha}_2 &\text{(S12)} \\
&= \mathrm{vec}\left(\mathrm{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right)\boldsymbol{\alpha}_3'\right) = \left(\mathbf{I}_K \otimes \mathrm{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right)\right) \mathrm{vec}\left(\boldsymbol{\alpha}_3'\right) \\
&= \left(\mathbf{I}_K \otimes \mathrm{vec}\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\right)\right)\boldsymbol{\alpha}_3 = \left(\mathbf{I}_K \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1\right)\boldsymbol{\alpha}_3. &\text{(S13)}
\end{aligned}$$

Consider the model in eq. (18), it holds

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

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

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

It is then possible to make explicit the dependence on each PARAFAC marginal by exploiting the results in eq. (S11)-(S13), as follows

$$\begin{aligned}
\mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} &= \mathrm{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \cdot \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} = \mathbf{b}_4 \boldsymbol{\beta}_4^{(r)} &\text{(S14)} \\
&= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_I\right)\boldsymbol{\beta}_1^{(r)} = \mathbf{b}_1 \boldsymbol{\beta}_1^{(r)} &\text{(S15)} \\
&= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\beta}_1^{(r)}\right)\boldsymbol{\beta}_2^{(r)} = \mathbf{b}_2 \boldsymbol{\beta}_2^{(r)} &\text{(S16)} \\
&= \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\mathbf{I}_K \otimes \boldsymbol{\beta}_2^{(r)} \otimes \boldsymbol{\beta}_1^{(r)}\right)\boldsymbol{\beta}_3^{(r)} = \mathbf{b}_3 \boldsymbol{\beta}_3^{(r)}. &\text{(S17)}
\end{aligned}$$

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

$$L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) = \prod_{t=1}^{T} (2\pi)^{-\frac{I_1 I_2 I_3}{2}} \left|\Sigma_3\right|^{-\frac{I_1 I_2}{2}} \left|\Sigma_2\right|^{-\frac{I_1 I_3}{2}} \left|\Sigma_1\right|^{-\frac{I_2 I_3}{2}}$$

$$\cdot \exp\Big(-\frac{1}{2}(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t)\bar{\times}_3\big(\circ_{j=1}^3 \Sigma_j^{-1}\big)\bar{\times}_3(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t)\Big)$$

$$\propto \exp\Big(-\frac{1}{2}\sum_{t=1}^T \tilde{\mathcal{E}}_t\bar{\times}_3(\Sigma_1^{-1}\circ\Sigma_2^{-1}\circ\Sigma_3^{-1})\bar{\times}_3\tilde{\mathcal{E}}_t\Big),$$

with

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

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

$$\begin{aligned}
L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2,\Sigma_3) &= \prod_{t=1}^T (2\pi)^{-\frac{I_1I_2I_3}{2}}\big|\Sigma_3\otimes\Sigma_2\otimes\Sigma_1\big|^{-\frac{1}{2}} \\
&\quad\cdot \exp\Big(-\frac{1}{2}\operatorname{vec}\big(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\big)'\big(\Sigma_3^{-1}\otimes\Sigma_2^{-1}\otimes\Sigma_1^{-1}\big)\operatorname{vec}\big(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\big)\Big) \\
&\propto \exp\Big(-\frac{1}{2}\sum_{t=1}^T \operatorname{vec}\big(\tilde{\mathcal{E}}_t\big)'\big(\Sigma_3^{-1}\otimes\Sigma_2^{-1}\otimes\Sigma_1^{-1}\big)\operatorname{vec}\big(\tilde{\mathcal{E}}_t\big)\Big).
\end{aligned}$$

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

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

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

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

$$+ \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \Bigg). \tag{S18}$$

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. (S15) we get

$$\begin{aligned}
&L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) \propto \\
&\propto \exp\Bigg(-\frac{1}{2} \sum_{t=1}^{T} -2\left(\mathbf{y}_t' - \operatorname{vec}\left(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t\right)'\right) \boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \mathbf{x}_t' \boldsymbol{\beta}_4^{(r)} \\
&\quad + \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right)' \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \boldsymbol{\Sigma}^{-1} \operatorname{vec}\left(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\right) \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \Bigg) \\
&= \exp\Bigg(-\frac{1}{2} \sum_{t=1}^{T} -2\left(\mathbf{y}_t' - \operatorname{vec}\left(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t\right)'\right) \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \\
&\quad + \left(\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)}\right)' \boldsymbol{\Sigma}^{-1} \left(\langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)}\right) \Bigg) \\
&= \exp\Bigg(-\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\beta}_1^{(r)\prime} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle^2 \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \\
&\quad - 2\left(\mathbf{y}_t' - \operatorname{vec}\left(\mathcal{B}_{-r} \bar{\times}_1 \mathbf{x}_t\right)'\right) \boldsymbol{\Sigma}^{-1} \langle \boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t \rangle \left(\boldsymbol{\beta}_3^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \otimes \mathbf{I}_{I_1}\right) \boldsymbol{\beta}_1^{(r)} \Bigg) \\
&= \exp\left(-\frac{1}{2} \boldsymbol{\beta}_1^{(r)\prime} \mathbf{S}_1^L \boldsymbol{\beta}_1^{(r)} - 2\mathbf{m}_1^L \boldsymbol{\beta}_1^{(r)} \right),
\end{aligned} \tag{S19}$$

with

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

Consider the case $j = 2$. From eq. (S16) we get

$$
\begin{aligned}
&L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3) \propto \\
&\propto \exp\Big(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \operatorname{vec}\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\boldsymbol{\Sigma}^{-1}\operatorname{vec}\big(\boldsymbol{\beta}_1^{(r)} \otimes \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\big)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)} \\
&\quad + \operatorname{vec}\big(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\big)'\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\boldsymbol{\Sigma}^{-1}\operatorname{vec}\big(\boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \boldsymbol{\beta}_3^{(r)}\big)\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\Big) \\
&= \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \operatorname{vec}\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\boldsymbol{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \circ \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)} \\
&\quad + \big(\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)}\big)'\boldsymbol{\Sigma}^{-1}\big(\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)}\big)\Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_2^{(r)\prime}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle^2\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)} \\
&\quad - 2\big(\mathbf{y}_t' - \operatorname{vec}\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\boldsymbol{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)}, \mathbf{x}_t\rangle\big(\boldsymbol{\beta}_3^{(r)} \otimes \mathbf{I}_{I_2} \otimes \boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_2^{(r)}\Bigg) \\
&= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_2^{(r)\prime}\mathbf{S}_2^L\boldsymbol{\beta}_2^{(r)} - 2\mathbf{m}_2^L\boldsymbol{\beta}_2^{(r)}\Big),
\end{aligned}
\tag{S20}
$$

with

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

Consider the case $j = 3$, by exploiting eq. (S17) we get

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

$$
\begin{aligned}
&= \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \mathrm{vec}\,\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\mathbf{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)},\mathbf{x}_t\rangle\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)} \\
&\quad + \big(\langle\boldsymbol{\beta}_4^{(r)},\mathbf{x}_t\rangle\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)}\big)'\mathbf{\Sigma}^{-1}\big(\langle\boldsymbol{\beta}_4^{(r)},\mathbf{x}_t\rangle\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)}\big)\Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_3^{(r)\prime}\langle\boldsymbol{\beta}_4^{(r)},\mathbf{x}_t\rangle^2\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\mathbf{\Sigma}^{-1}\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)} \\
&\quad - 2\big(\mathbf{y}_t' - \mathrm{vec}\,\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\mathbf{\Sigma}^{-1}\langle\boldsymbol{\beta}_4^{(r)},\mathbf{x}_t\rangle\big(\mathbf{I}_{I_3}\otimes\boldsymbol{\beta}_2^{(r)}\otimes\boldsymbol{\beta}_1^{(r)}\big)\boldsymbol{\beta}_3^{(r)}\Bigg) \\
&= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_3^{(r)\prime}\mathbf{S}_3^L\boldsymbol{\beta}_3^{(r)} - 2\mathbf{m}_3^L\boldsymbol{\beta}_3^{(r)}\Big),
\end{aligned} \tag{S21}
$$

with

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

Finally, in the case $j = 4$. From eq. (S18) we get

$$
\begin{aligned}
&L(\mathbf{Y}|\mathcal{B},\Sigma_1,\Sigma_2,\Sigma_3) \propto \\
&\propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} -2\big(\mathbf{y}_t' - \mathrm{vec}\,\big(\mathcal{B}_{-r}\bar{\times}_1\mathbf{x}_t\big)'\big)\mathbf{\Sigma}^{-1}\,\mathrm{vec}\,\big(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\circ\boldsymbol{\beta}_3^{(r)}\big)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)} \\
&\quad + \boldsymbol{\beta}_4^{(r)\prime}\mathbf{x}_t\,\mathrm{vec}\,\big(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\circ\boldsymbol{\beta}_3^{(r)}\big)'\mathbf{\Sigma}^{-1}\,\mathrm{vec}\,\big(\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\circ\boldsymbol{\beta}_3^{(r)}\big)\mathbf{x}_t'\boldsymbol{\beta}_4^{(r)}\Bigg) \\
&= \exp\Big(-\frac{1}{2}\boldsymbol{\beta}_4^{(r)\prime}\mathbf{S}_4^L\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)}\Big),
\end{aligned} \tag{S22}
$$

with

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

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

S.6.5.1 Full conditional distribution of $\beta_1^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_1^{(r)}$ is obtained by combining the prior distribution in eq. (20) and the likelihood in eq. (S19) as follows

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

where

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

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

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

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

The posterior full conditional distribution of $\boldsymbol{\beta}_2^{(r)}$ is obtained by combining the prior distribution in eq. (20) and the likelihood in eq. (S20) as follows

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

where

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

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

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

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

The posterior full conditional distribution of $\boldsymbol{\beta}_3^{(r)}$ is obtained by combining the prior distribution in eq. (20) and the likelihood in eq. (S21) as follows

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

where

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

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

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

S.6.5.4 Full conditional distribution of $\beta_4^{(r)}$

The posterior full conditional distribution of $\boldsymbol{\beta}_4^{(r)}$ is obtained by combining the prior distribution in eq. (20) and the likelihood in eq. (S22) as follows

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

$$
\begin{aligned}
&= \exp\Big(-\frac{1}{2}\big(\boldsymbol{\beta}_4^{(r)\prime}\mathbf{S}_4^L\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)} + \boldsymbol{\beta}_4^{(r)\prime}(W_{4,r}\phi_r\tau)^{-1}\boldsymbol{\beta}_4^{(r)}\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\big(\boldsymbol{\beta}_4^{(r)\prime}\big(\mathbf{S}_4^L + (W_{4,r}\phi_r\tau)^{-1}\big)\boldsymbol{\beta}_4^{(r)} - 2\mathbf{m}_4^L\boldsymbol{\beta}_4^{(r)}\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\big(\boldsymbol{\beta}_4^{(r)\prime}\bar{\Sigma}_{\boldsymbol{\beta}_4^r}^{-1}\boldsymbol{\beta}_4^{(r)} - 2\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_4^r}\boldsymbol{\beta}_4^{(r)}\big)\Big),
\end{aligned}
$$

where

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

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

$$
p(\boldsymbol{\beta}_4^{(r)}|\boldsymbol{\beta}_{-4}^{(r)}, \mathcal{B}_{-r}, W_{4,r}, \phi_r, \tau, \Sigma_1, \Sigma_2, \Sigma_3, \mathbf{Y}) \sim \mathcal{N}_{I_1I_2I_3}(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}_4^r}, \bar{\Sigma}_{\boldsymbol{\beta}_4^r}).
$$

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

$$
\begin{aligned}
&p(\Sigma_1|\mathcal{B}, \mathbf{Y}, \Sigma_2, \Sigma_3, \gamma) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\Sigma_1|\gamma) \\
&\propto \big|\Sigma_1\big|^{-\frac{TI_2I_3}{2}} \exp\Big(-\frac{1}{2}\sum_{t=1}^T \operatorname{vec}\big(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\big)'(\Sigma_3^{-1}\otimes\Sigma_2^{-1}\otimes\Sigma_1^{-1}) \\
&\quad \cdot \operatorname{vec}\big(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\big)\Big) \cdot \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+1}{2}} \exp\Big(-\frac{1}{2}\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big)\Big) \\
&\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\Big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big) + \sum_{t=1}^T \operatorname{vec}\big(\tilde{\mathcal{E}}_t\big)'(\mathbf{Z}_1\otimes\Sigma_1^{-1})\operatorname{vec}\big(\tilde{\mathcal{E}}_t\big)\Big)\Big) \\
&\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\Big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big) \\
&\quad + \sum_{t=1}^T \operatorname{vec}\big(\tilde{\mathbf{E}}_{(1),t}\big)'(\mathbf{Z}_1\otimes\Sigma_1^{-1})\operatorname{vec}\big(\tilde{\mathbf{E}}_{(1),t}\big)\Big)\Big) \\
&\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\Big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big)
\end{aligned}
$$

$$+\sum_{t=1}^{T} \operatorname{tr}\Big(\operatorname{vec}\big(\tilde{\mathbf{E}}_{(1),t}\big)' \operatorname{vec}\big(\Sigma_1^{-1}\tilde{\mathbf{E}}_{(1),t}\mathbf{Z}_1\big)\Big)\Big)$$

$$\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big) + \sum_{t=1}^{T} \operatorname{tr}\big(\tilde{\mathbf{E}}'_{(1),t}\Sigma_1^{-1}\tilde{\mathbf{E}}_{(1),t}\mathbf{Z}_1\big)\big)\Big)$$

$$\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big) + \sum_{t=1}^{T} \operatorname{tr}\big(\tilde{\mathbf{E}}_{(1),t}\mathbf{Z}_1\tilde{\mathbf{E}}'_{(1),t}\Sigma_1^{-1}\big)\big)\Big).$$

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

$$\begin{aligned} p(\Sigma_1|\mathcal{B},\mathbf{Y},\Sigma_2,\Sigma_3) &\propto \big|\Sigma_1\big|^{-\frac{\nu_1+I_1+TI_2I_3+1}{2}} \exp\Big(-\frac{1}{2}\big(\operatorname{tr}\big(\gamma\Psi_1\Sigma_1^{-1}\big) + \operatorname{tr}\big(S_1\Sigma_1^{-1}\big)\big)\Big) \\ &\propto \big|\Sigma_1\big|^{-\frac{(\nu_1+TI_2I_3)+I_1+1}{2}} \exp\Big(-\frac{1}{2}\operatorname{tr}\big((\gamma\Psi_1+S_1)\Sigma_1^{-1}\big)\Big), \end{aligned}$$

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

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

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 $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t$ 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

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

$$\propto \big|\Sigma_2\big|^{-\frac{TI_1I_3}{2}} \exp\Big(-\frac{1}{2}\sum_{t=1}^{T}(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t)\bar{\times}_3(\Sigma_1^{-1}\circ\Sigma_2^{-1}\circ\Sigma_3^{-1})$$

$$\bar{\times}_3(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t)\Big) \cdot \big|\Sigma_2\big|^{-\frac{\nu_2+I_2+1}{2}} \exp\Big(-\frac{1}{2}\operatorname{tr}\big(\Psi_2\Sigma_2^{-1}\big)\Big)$$

$$\propto \big|\Sigma_2\big|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\Big(-\frac{1}{2}\big(\operatorname{tr}\big(\gamma\Psi_2\Sigma_2^{-1}\big)$$

$$+\sum_{t=1}^{T} \tilde{\mathcal{E}}_t\bar{\times}_3(\Sigma_1^{-1}\circ\Sigma_2^{-1}\circ\Sigma_3^{-1})\bar{\times}_3\tilde{\mathcal{E}}_t\big)\Big)$$

$$\propto \big|\Sigma_2\big|^{-\frac{\nu_2+I_2+TI_1I_3+1}{2}} \exp\Big(-\frac{1}{2}\big(\operatorname{tr}\big(\gamma\Psi_2\Sigma_2^{-1}\big)$$

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

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

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

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 $\tilde{\mathcal{E}}_t = \mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t$, $\tilde{\mathbf{E}}_{(1),t}$ 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

$$
\begin{aligned}
&p(\Sigma_3|\mathcal{B}, \mathbf{Y}, \Sigma_1, \Sigma_2, \gamma) \propto L(\mathbf{Y}|\mathcal{B}, \Sigma_1, \Sigma_2, \Sigma_3)\pi(\Sigma_3|\gamma) \\
&\propto \left|\Sigma_3\right|^{-\frac{TI_1I_2}{2}} \exp\Big(-\frac{1}{2}\sum_{t=1}^{T}\operatorname{vec}\left(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\right)'(\Sigma_3^{-1} \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}) \\
&\quad \cdot \operatorname{vec}\left(\mathcal{Y}_t - \mathcal{B}\bar{\times}_1\mathbf{x}_t\right)\Big) \cdot \left|\Sigma_3\right|^{-\frac{\nu_3+I_3+1}{2}} \exp\Big(-\frac{1}{2}\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right)\Big) \\
&\propto \left|\Sigma_3\right|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\Big(-\frac{1}{2}\Big(\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) + \sum_{t=1}^{T}\operatorname{vec}\left(\tilde{\mathcal{E}}_t\right)'(\Sigma_3^{-1} \otimes \mathbf{Z}_3)\operatorname{vec}\left(\tilde{\mathcal{E}}_t\right)\Big)\Big) \\
&\propto \left|\Sigma_3\right|^{-\frac{\nu_3+I_3+TI_1I_2+1}{2}} \exp\Big(-\frac{1}{2}\Big(\operatorname{tr}\left(\gamma\Psi_3\Sigma_3^{-1}\right) \\
&\quad +\sum_{t=1}^{T}\operatorname{vec}\left(\tilde{\mathbf{E}}_{(1),t}\right)'(\Sigma_3^{-1} \otimes \mathbf{Z}_3)\operatorname{vec}\left(\tilde{\mathbf{E}}_{(1),t}\right)\Big)\Big)
\end{aligned}
$$

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

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

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

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

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

S.6.9 Full conditional distribution of γ

Using a gamma prior distribution we have

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

thus

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

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: $(\boldsymbol{\beta}_1^{(r)}, \ldots, \boldsymbol{\beta}_J^{(r)})_{r=1}^R$. 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 Bolzmann-like distribution

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

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

We 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:

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

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

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

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[2].

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

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

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

$$Y_t = \mathcal{B} \bar{\times}_1 \operatorname{vec}\left(X_t\right) + E_t, \qquad E_t \sim \mathcal{N}_{I_1,I_2}(\mathbf{0}, \Sigma_1, \mathbf{I}_{I_2}).$$

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. (20). 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 $\boldsymbol{\beta}_1^{(r)}, \boldsymbol{\beta}_2^{(r)}, \boldsymbol{\beta}_3^{(r)}$, $r = 1, \ldots, R$ (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.

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

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 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.

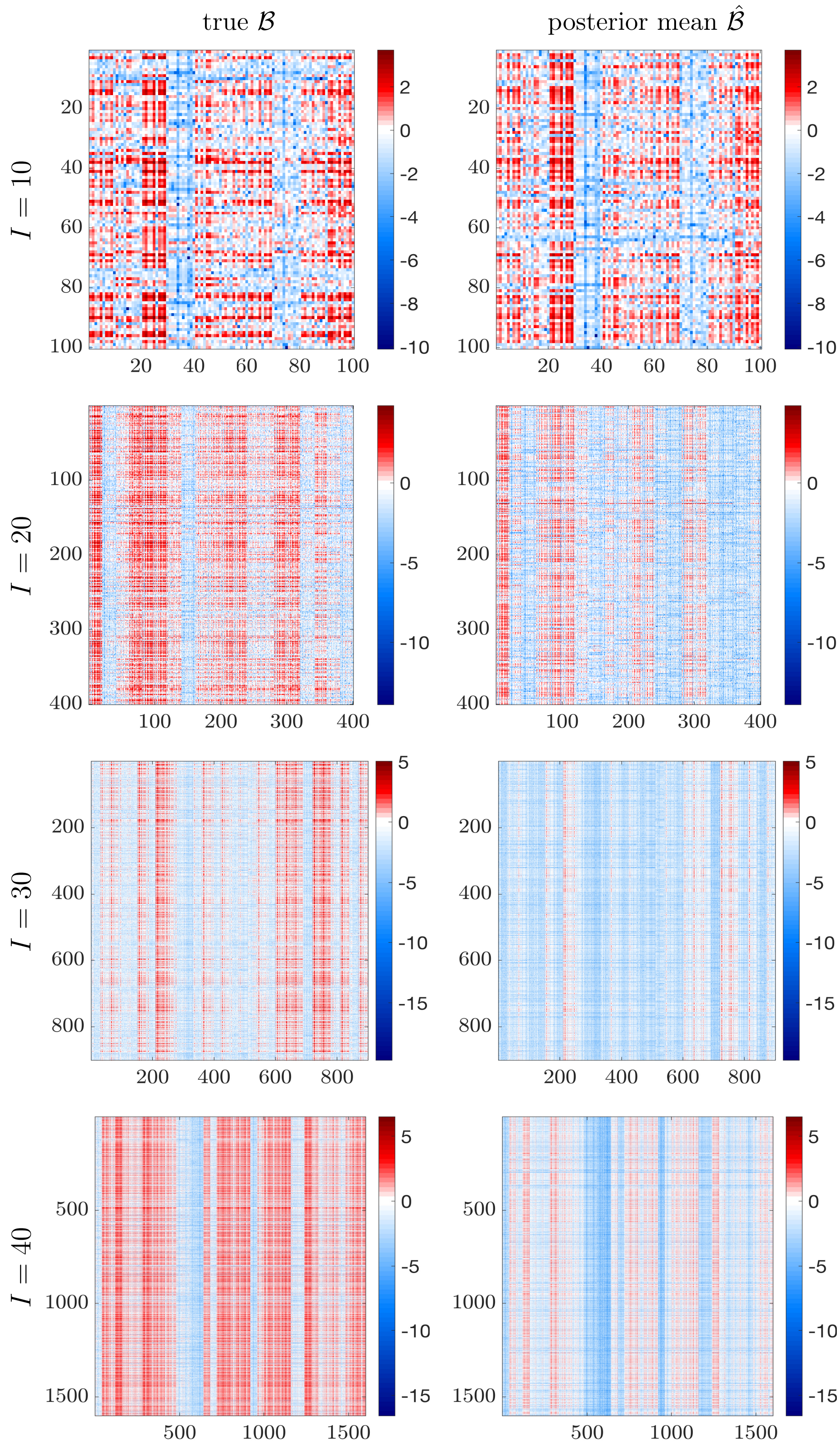

Figure 5: Logarithm of the absolute value of the coefficient tensors (in matricized form): true $\mathcal{B}$ (*left*) and posterior mean estimate $\hat{\mathcal{B}}$ (*right*), for four experiments with different size I (*in row*).

S.8.1 Experiment: I=10

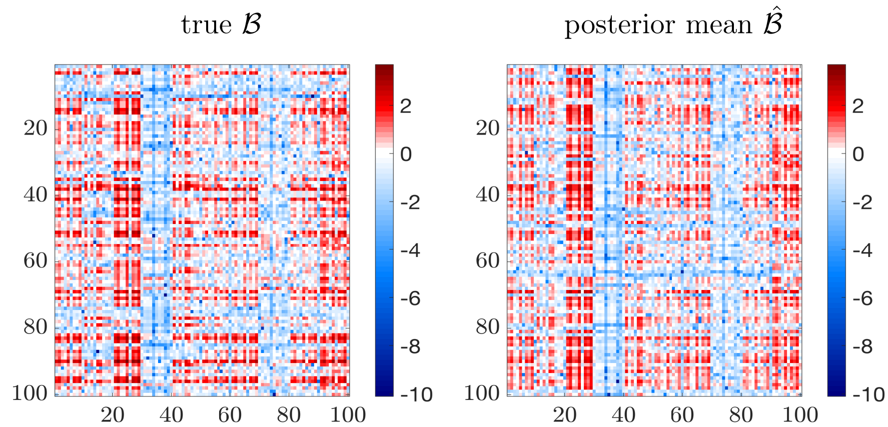

Figure 6: Experiment $I = 10$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true $\mathcal{B}$ (*left*) and posterior mean estimate $\hat{\mathcal{B}}$ (*right*).

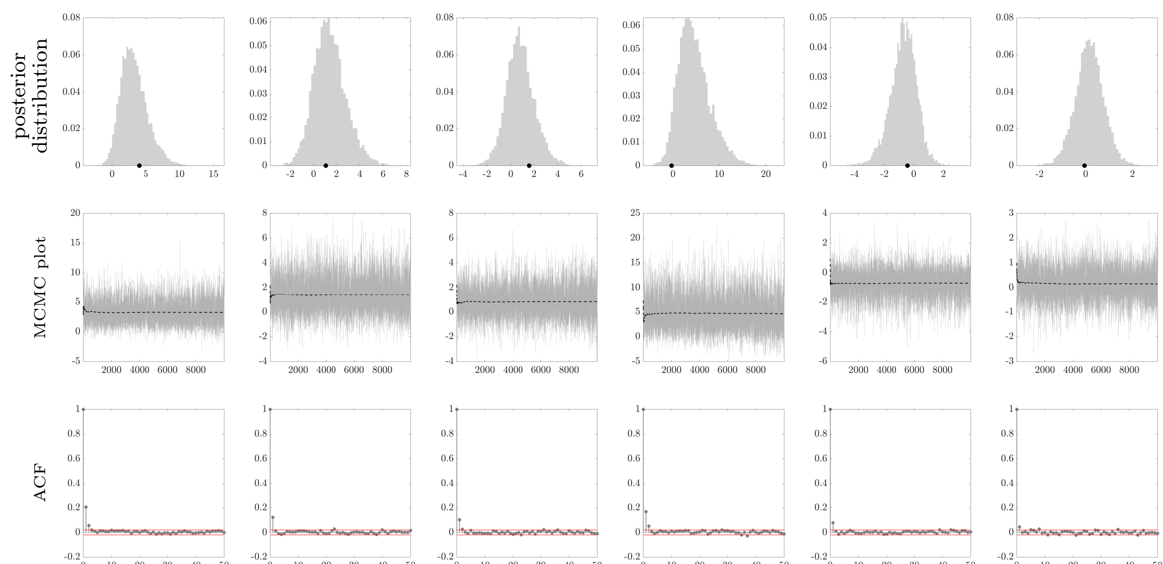

Figure 7: Experiment $I = 10$. Posterior distribution (*first row*, the black dot is the true value), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$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

S.8.2 Experiment: I=20

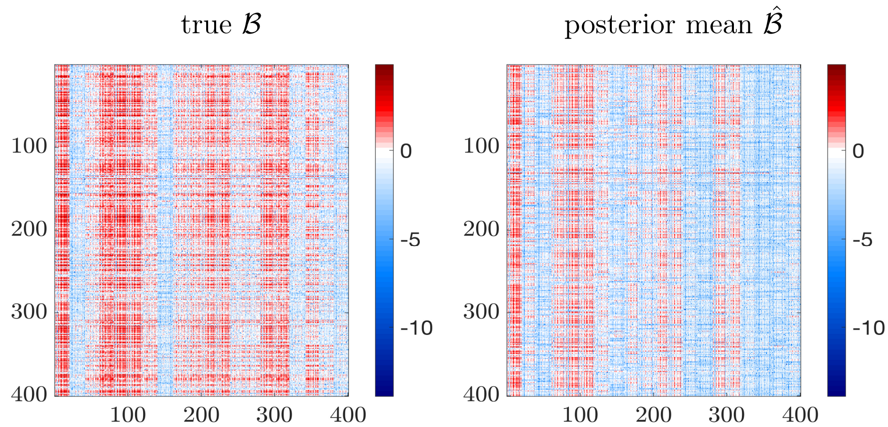

Figure 8: Experiment $I = 20$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true $\mathcal{B}$ (*left*) and posterior mean estimate $\hat{\mathcal{B}}$ (*right*).

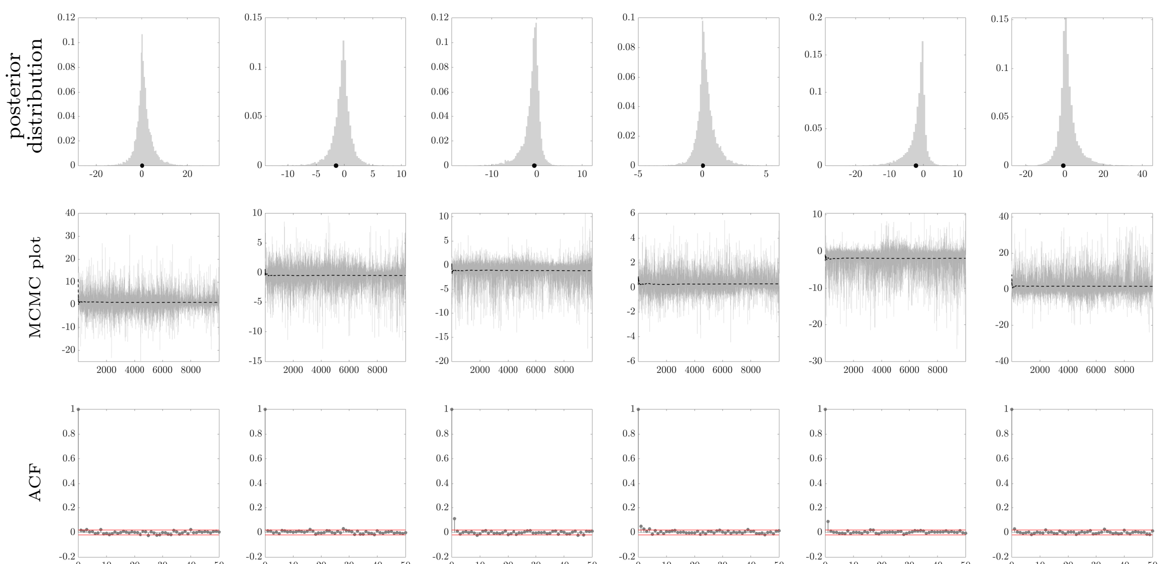

Figure 9: Experiment $I = 20$. Posterior distribution (*first row*, the black dot is the true value), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$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

S.8.3 Experiment: I=30

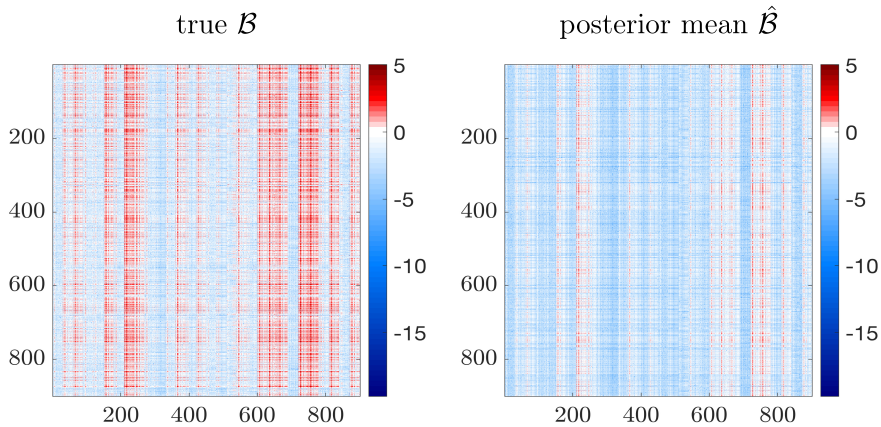

Figure 10: Experiment $I = 30$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true $\mathcal{B}$ (*left*) and posterior mean estimate $\hat{\mathcal{B}}$ (*right*).

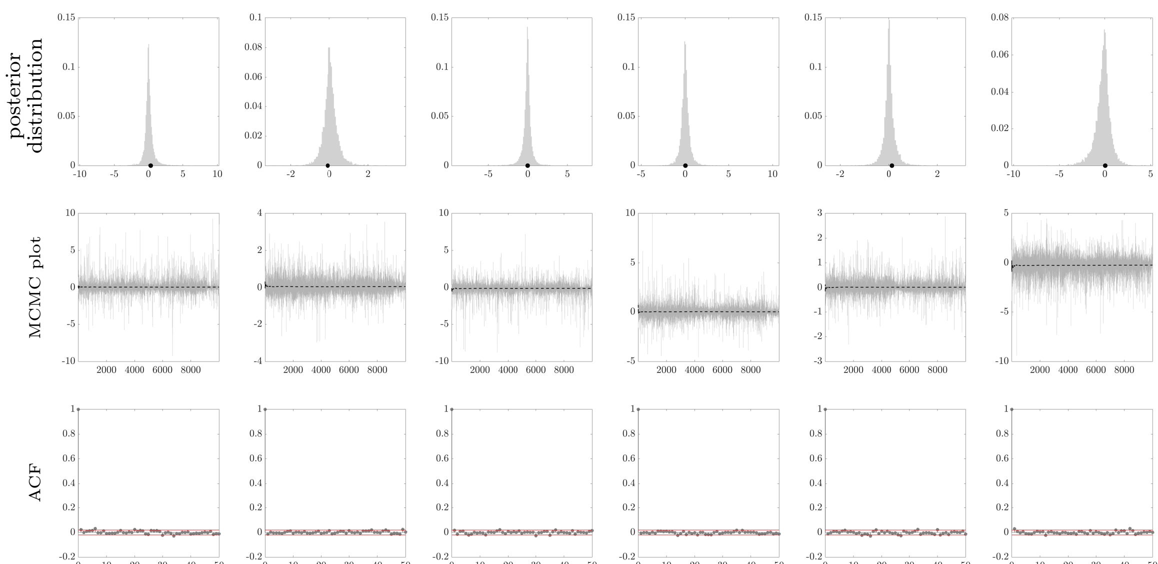

Figure 11: Experiment $I = 30$. Posterior distribution (*first row*, the black dot is the true value), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$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

S.8.4 Experiment: I=40

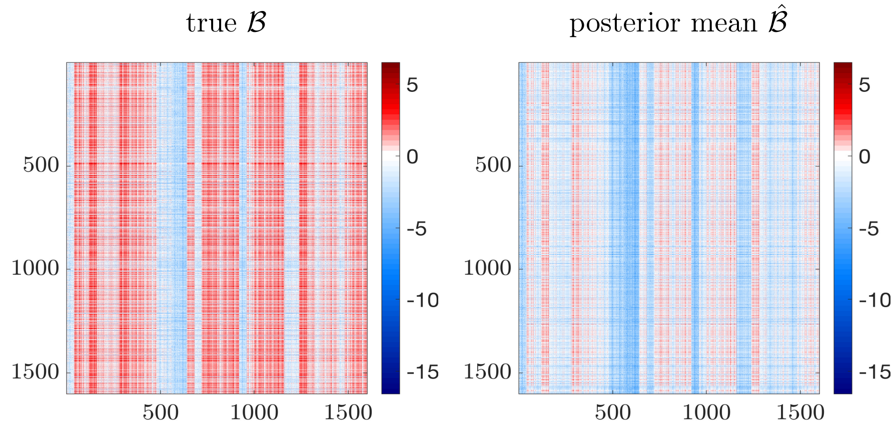

Figure 12: Experiment $I = 40$. Logarithm of the absolute value of the coefficient tensors (in matricized form): true $\mathcal{B}$ (*left*) and posterior mean estimate $\hat{\mathcal{B}}$ (*right*).

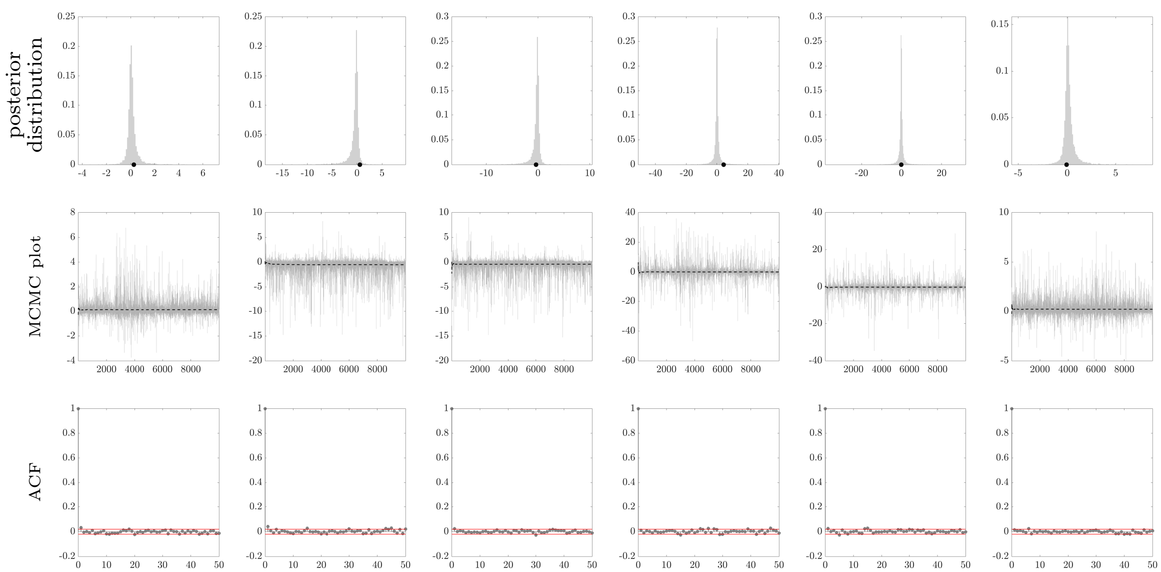

Figure 13: Experiment $I = 40$. Posterior distribution (*first row*, the black dot is the true value), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for some randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$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

S.9 Data Description

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

The bilateral trade data come from the United Nations COMTRADE database[3], whereas the data on bilateral outstanding capital come from the Bank of International Settlements database[4], both are publicly available resources. For each couple (i, j) of countries, the international trade data from COMTRADE report total exports from country i to country j occurred during year t, while the BIS dataset gives the total amount of claims (i.e., credit) of country i vis-á-vis country j in year t. We use a subset of the COMTRADE database. Our sample of yearly observations for 10 countries ($I_1 = I_2 = I = 10$) runs from 2003 to 2016. In order to remove potential non-linearities in the data, we take the logarithm all variables of interest. We thus consider the international trade and financial network in each period as one observation from a real-valued tensor-valued stochastic process. To sum up, our dataset consists in a 3-order tensor-valued time series of length $T = 13$. At each time t, the 3-order tensor $\mathcal{Y}_t$ has dimension (I_1, I_2, I_3), with $I_1 = I_2 = I = 10$ and $I_3 = 2$, and it represents a 2-layer node-aligned network (or multiplex) with 10 vertices (countries), where each edge is given by a bilateral trade flow or financial stock. The entry $(i, j, 1, t)$ of $\mathcal{Y}_t$ reports the total exports of country i vis-à-vis country j, in year t, whereas entry $(i, j, 2, t)$ contains the total outstanding credit from country i towards country j, in year t. The series $(\mathcal{Y}_t)_t$, $t = 1, \ldots, T$, has been standardized (over the temporal dimension).

S.10 Additional plots for TAR(1) model

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

[4] http://stats.bis.org/statx/toc/LBS.html

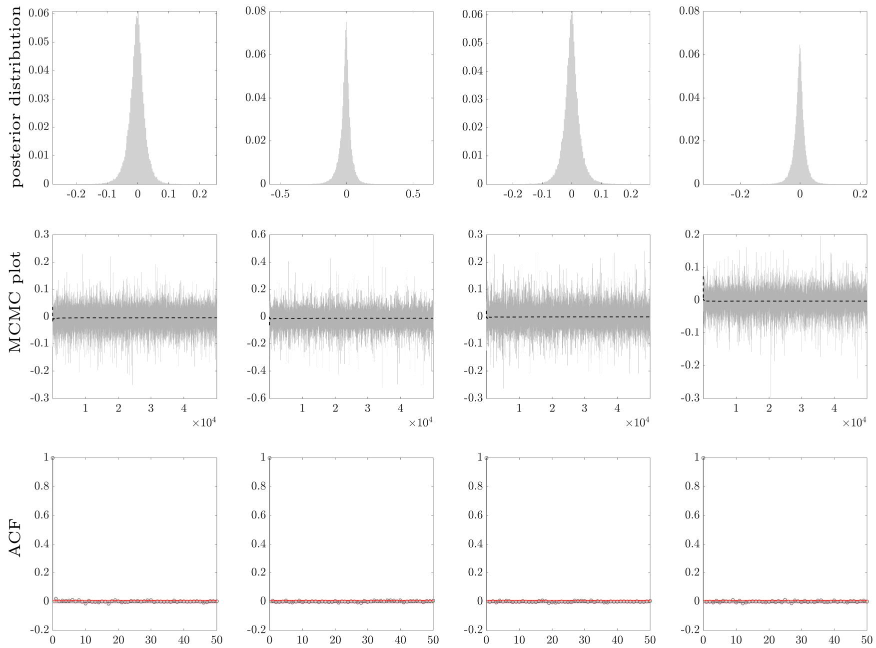

Figure 14: Posterior distribution (*first row*), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for four randomly chosen cells of the estimated coefficient tensor $\hat{\mathcal{B}}$ (in each *column*).

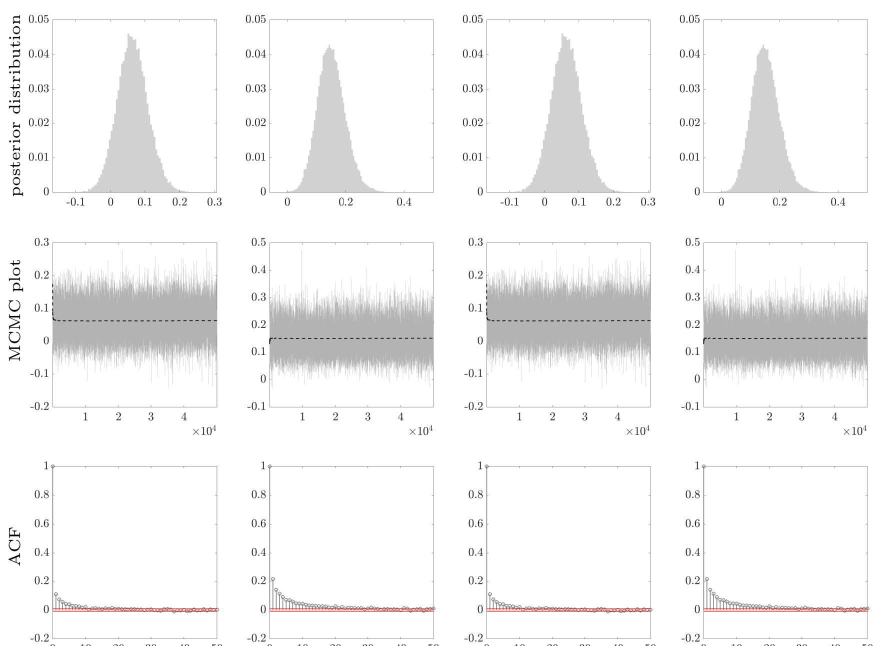

Figure 15: Posterior distribution (*first row*), MCMC plot (*second row*, dashed line represents the progressive mean) and autocorrelation function (*third row*) for two randomly chosen cells of the estimated covariance matrix $\hat{\Sigma}_1$ (*first and second column*) and $\hat{\Sigma}_2$ (*third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

References

Anundsen, A. K., K. Gerdrup, F. Hansen, and K. Kragh-Sørensen (2016): "Bubbles and crises: The role of house prices and credit," *Journal of Applied Econometrics*, 31, 1291–1311.

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

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

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

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

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

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

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

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

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

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

Kolda, T. G. (2006): "Multilinear operators for higher-order decompositions," Tech. rep., Sandia National Laboratories.

KOLDA, T. G. AND B. W. BADER (2009): "Tensor decompositions and applications," *SIAM Review*, 51, 455–500.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74, 119–147.

KRUIJER, W., J. ROUSSEAU, AND A. VAN DER VAART (2010): "Adaptive Bayesian density estimation with location-scale mixtures," *Electronic Journal of Statistics*, 4, 1225–1257.

LANNE, M. AND H. NYBERG (2016): "Generalized forecast error variance decomposition for linear and nonlinear multivariate models," *Oxford Bulletin of Economics and Statistics*, 78, 595–603.

MEYFROIDT, P., T. K. RUDEL, AND E. F. LAMBIN (2010): "Forest transitions, trade, and the global displacement of land use," *Proceedings of the National Academy of Sciences*, 107, 20917–20922.

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

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY (2007): *Numerical recipes: the art of scientic computing*, Cambridge University Press.

ROBERT, C. P. AND G. CASELLA (2004): *Monte Carlo statistical methods*, Springer.

SCHWEITZER, F., G. FAGIOLO, D. SORNETTE, F. VEGA-REDONDO, A. VESPIGNANI, AND D. R. WHITE (2009): "Economic networks: the new challenges," *Science*, 325, 422–425.

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

SQUARTINI, T., G. FAGIOLO, AND D. GARLASCHELLI (2011): "Randomizing world trade. I. A binary network analysis," *Physical Review E*, 84, 046117.

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