Supplementary material for "Bayesian Markov Switching Tensor Regression for Time-varying Networks"

Monica Billio
Department of Economics, Ca' Foscari University of Venice
and
Roberto Casarin
Department of Economics, Ca' Foscari University of Venice
and
Matteo Iacopini
Scuola Normale Superiore of Pisa

July 3, 2019

S.1 Tensor Calculus and Decompositions

In this section we introduce some notation for multilinear arrays (i.e. tensors) and some operations on tensors and between tensors and lower dimensional objects (such as matrices and vectors) and the representation results for tensors (or tensor decomposition/approximation). A noteworthy introduction to tensors and corresponding operations is Cichocki et al. (2016), while a remarkable reference for tensor decomposition methods is Kolda and Bader (2009).

A multidimensional array is an object which generalises the concept of matrix. It may have an arbitrary number of dimensions (or modes), whose number is the *order* of the tensor. Consequently, a matrix is a second order tensor. We denote a whole column (or row) of a matrix by the symbol ":" and the same is used for tensors, where this symbol denotes that we are considering the corresponding whole dimension. The mode-k fiber of a tensor is the vector obtained along the dimension k by fixing all the other dimensions. Differently from the bidimensional case, however, with higher order arrays it is possible to identify also slices (i.e. matrices) or generalizations of them, by fixing all but two or more dimensions (or modes) of the tensor. For example, the mode-k fiber of the tensor $\mathcal{X}$ is denoted by

$$\mathcal{X}_{(i_1,\dots,i_{k-1},:,i_{k+2},\dots,i_D)}\,. \tag{S.1.1}$$

The operation of transforming a D-array $\mathcal{X}$ into a matrix is called *mode-n matricization* and is denoted by $\mathbf{X}_{(n)}$ It consists in rearranging all the mode-n fibers to be the columns of a matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{d_n \times \bar{d}_{(-n)}}$ with $\bar{d}_{(-n)} = \Pi_{i\neq n} d_i$. For some examples, see Kolda and Bader (2009). Analogously to the matrix version, the *vectorization* of a tensor consists in stacking all the elements in a unique vector of dimension $\bar{d} = \Pi_i d_i$. Notice that, the ordering of the elements is not important as long as it is consistent across the calculations.

Many product operations have been defined for tensors, but here we constrain ourselves to the operators used in this work and we point to Cichocki et al. (2016) for a summary of other operators. The *mode-n product* between a tensor $\mathcal{X}$ and a vector $\mathbf{v} \in \mathbb{R}^{d_n}$ can be interpreted as the standard Euclidean inner product between the vector and each mode-n fiber of the tensor. Consequently, this operator suppresses one dimension of the tensor and results in a lower order tensor. It is defined element-wise by

$$\mathcal{Y}_{(i_1,\dots,i_{n-1},i_{n+1},\dots,i_D)} = (\mathcal{X} \times_n \mathbf{v})_{(i_1,\dots,i_{n-1},i_{n+1},\dots,i_D)} = \sum_{i_n=1}^{d_n} \mathcal{X}_{i_1,\dots,i_D} \mathbf{v}_{i_n},$$

with $\mathcal{Y} \in \mathbb{R}^{d_1 \times \dots, d_{n-i}, d_{n+1}, \dots \times d_D}$. This product is not commutative, since the order of the elements in the multiplication is relevant. The *outer product* $\circ$ of two tensors $\mathcal{Y} \in \mathbb{R}^{d_1^Y \times \dots \times d_M^Y}$ and $\mathcal{X} \in \mathbb{R}^{d_1^X \times \dots \times d_N^X}$ is the tensor $\mathcal{Z} \in \mathbb{R}^{d_1^Y \times \dots \times d_M^Y \times d_1^X \times \dots \times d_N^X}$ whose entries are

$$\mathcal{Z}_{i_1,\dots,i_M,j_1,\dots,j_N} = (\mathcal{Y} \circ \mathcal{X})_{i_1,\dots,i_M,j_1,\dots,j_N} = \mathcal{Y}_{i_1,\dots,i_M} \mathcal{X}_{j_1,\dots,j_N}\,.$$

For example, the outer product of two vectors is a matrix, while the outer product of two matrices is a tensor of order 4. As a special case, the outer product of two column vectors $\mathbf{a}$ and $\mathbf{b}$ can be equivalently represented by means of the Kronocker product $\otimes$:

$$\mathbf{a} \circ \mathbf{b} = \mathbf{b} \otimes \mathbf{a} = \mathbf{a} \cdot \mathbf{b}' .$$

We now define two tensor representations, or decompositions, which are useful in two respects: (i) the algebraic objects that form the decomposition are generally low dimensional and more tractable than the tensor; (ii) they can be used to provide a good approximation of the original array. The Tucker decomposition can be thought of as a higher-order generalization of Principal Component Analysis (PCA): a tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \ldots \times d_D}$ is decomposed into the product (along the corresponding modes) of a lower dimensional tensor $\mathcal{Y} \in \mathbb{R}^{y_1 \times \ldots \times y_D}$ and D factor matrices $A^{(l)} \in \mathbb{R}^{d_l \times y_l}$, $1 \leq l \leq D$. Following the notation in Kolda and Bader (2009)

$$\mathcal{X} = \mathcal{Y} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \ldots \times_D A^{(D)} = \sum_{i_1=1}^{y_1} \sum_{i_2=1}^{y_2} \ldots \sum_{i_D=1}^{y_D} y_{i_1,i_2,\ldots,i_D} \mathbf{a}_{i_1}^{(1)} \circ \mathbf{a}_{i_2}^{(2)} \circ \ldots \circ \mathbf{a}_{i_D}^{(D)} .$$

Here $\mathbf{a}_{i_l}^{(l)} \in \mathbb{R}^{g_l \times 1}$ is the l-th column of the matrix $A^{(l)}$. As a result, each entry of the tensor is obtained as

$$\mathcal{X}_{j_1,\ldots,j_D} = \sum_{i_1=1}^{y_1} \sum_{i_2=1}^{y_2} \ldots \sum_{i_D=1}^{y_D} y_{i_1,i_2,\ldots,i_D} a_{i_1,j_1}^{(1)} a_{i_2,j_2}^{(2)} \ldots a_{i_D,j_D}^{(D)} \quad 1 \leq j_l \leq d_l,\ 1 \leq l \leq D.$$

A special case of the Tucker decomposition is obtained when the core tensor collapses to a scalar and the factor matrices reduce to a single column vector each one is called PARAFAC(R)[1]. More precisely, the PARAFAC(R) decomposition allows to represent a D-order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \ldots \times d_D}$ as the sum of R rank one tensors, that is, of outer products

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

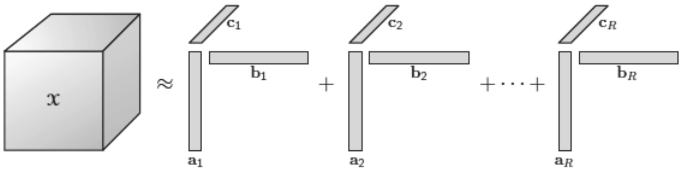


Figure 1: PARAFAC decomposition of $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, with $\mathbf{a}_r \in \mathbb{R}^{d_!}$, $\mathbf{b}_r \in \mathbb{R}^{d_2}$ and $\mathbf{c}_r \in \mathbb{R}^{d_3}$, $1 \leq r \leq R$. Figure from Kolda and Bader (2009).

(denoted by $\circ$) of vectors (also called marginals in this case)[2]

$$\mathcal{X} = \sum_{r=1}^{R} \mathcal{X}_r = \sum_{r=1}^{R} \mathbf{x}_1^{(r)} \circ \ldots \circ \mathbf{x}_D^{(r)},$$

with $\mathbf{x}_j^{(r)} \in \mathbb{R}^{d_j}$.

S.2 Computational Details

S.2.1 Data augmentation

To derive the likelihood function of the model in eq. (6) and develop an efficient inferential process, it is useful to start from eq. (2), which describes the statistical model for each edge as a a zero-inflated logit model. Starting from the seminal work of Lambert (1992), who proposed a modelling framework for count data with a great proportion of zeros, zero-inflated models have been applied to settings where the response variable is not integer valued. Binary responses have been considered by Harris and Zhao (2007), who dealt with

[2]An alternative representation may be used, if all the vectors $\mathbf{x}_j^r$ are normalized to have unitary length. In this case the weight of each component r is captured by the $r-th$ component of the vector $\boldsymbol{\lambda} \in \mathbb{R}^R$:

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \big(\mathbf{x}_1^{(r)} \circ \ldots \circ \mathbf{x}_D^{(r)}\big)$$

.

an ordered probit model. This is the closest approach to ours, though the specification in eq. (2) substantially differs in two aspects. First, we use of a logistic link function, which is known to have slightly fatter tails than the cumulative normal distribution used in probit models. Second, differently from the majority of the literature which assumes a constant mixing probability, the parameter $\rho(t)$ is evolving according to a latent process.

The likelihood function is:

$$L(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}) = \sum_{s_1,\ldots,s_T} \prod_{t=1}^{T} p(\mathcal{X}_t|s_t,\boldsymbol{\theta})p(s_t|s_{t-1}) \,.$$

Through the introduction of a latent variables $\mathbf{s} = \{s_t\}_{t=1}^{T}$, we obtain the data augmented likelihood

$$L(\boldsymbol{\mathcal{X}},\mathrm{s}|\boldsymbol{\theta}) = \prod_{t=1}^{T}\prod_{l=1}^{L}\prod_{h=1}^{L} \big(p(\mathcal{X}_t|s_t = l,\boldsymbol{\theta})p(s_t = l|s_{t-1} = h, \boldsymbol{\Xi})\big)^{\zeta_{t,l}\zeta_{t-1,h}} \,. \tag{S.2.1}$$

Considering the observation model in eq. (2) and defining $\mathcal{T}_l = \{t : \zeta_{t,l} = 1\}$ for each l, we can rewrite eq. (S.2.1) as

$$\begin{aligned} L(\boldsymbol{\mathcal{X}},\mathrm{s}|\boldsymbol{\theta}) = &\prod_{l=1}^{L}\prod_{t\in\mathcal{T}_l}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\left(\frac{(1-\rho_l)\exp(\mathbf{z}_t'\mathbf{g}_{ijk,l})}{1+\exp(\mathbf{z}_t'\mathbf{g}_{ijk,l})}\right)^{x_{ijk,t}}\left(\rho_l + \frac{1-\rho_l}{1+\exp(\mathbf{z}_t'\mathbf{g}_{ijk,l})}\right)^{1-x_{ijk,t}} \\ &\cdot \prod_{t=1}^{T}\prod_{l=1}^{L}\prod_{h=1}^{L}\xi_{h,l}^{\zeta_{t,l}\zeta_{t-1,h}} . \end{aligned}$$

Since the function cannot be expressed as a series of products due to the sum in the rightmost term, we choose to further augment the data through the introduction of latent allocation variables $\mathcal{D} = \{\mathcal{D}_l\}_{l=1}^{L}$, with $\mathcal{D}_l = (d_{ijk,l})_{ijk}$. Finally, we perform the augmentation for dealing with the logistic part of the model (Polson et al. (2013)). The complete data likelihood is

$$\begin{aligned} L(\boldsymbol{\mathcal{X}},\mathcal{D},\boldsymbol{\Omega},\mathrm{s}|\boldsymbol{\theta}) &= p(\boldsymbol{\mathcal{X}},\mathcal{D},\boldsymbol{\Omega}|\mathrm{s},\boldsymbol{\theta})p(\mathrm{s}|\boldsymbol{\theta}) \\ &= \prod_{t=1}^{T} p(\mathcal{X}_t,\mathcal{D}_t,\boldsymbol{\Omega}_t|s_t,\boldsymbol{\theta})p(s_t|\boldsymbol{\theta}) \end{aligned}$$

$$= \left(\prod_{l=1}^{L} \prod_{t \in \mathcal{T}_l} \prod_{i=1}^{I} \prod_{j=i}^{J} \prod_{k=1}^{K} p(x_{ijk,t}, d_{ijk,t}, \omega_{ijk,t} | s_t = l, \rho_l, \mathcal{G}_l) \right) p(\mathbf{s}|\boldsymbol{\Xi}). \quad \text{(S.2.2)}$$

We start by analysing in detail the first term. The joint distribution of the observation $x_{ijk,t}$ and the latent variables $(d_{ijk,t}, \omega_{ijk,t})$ is obtained from the marginal distribution of the observation in two steps. First, we augment the model by introducing the latent allocation $d_{ijk,l} \in \{0, 1\}$ for each l. Via this data augmentation step we are able to factorise the summation in eq. (2) for each regime l. In words, the allocation latent variable is used to identify the component of the mixture in eq. (2) from which the observation $x_{ijk,t}$ is drawn. Secondly, we use a further data augmentation step via the introduction of the latent variables $\omega_{ijk,t}$ following Polson et al. (2013), for dealing with the logistic regression. Introducing the allocation variable $d_{ijk,t}$ in eq. (2), for each i, j, k, t, one gets

$$\begin{aligned}
&p(x_{ijk,t} | d_{ijk,t}, s_t = l, \rho_l, \mathcal{G}_l) \\
&= \left(\delta_{\{0\}}(x_{ijk,t}) \right)^{d_{ijk,t}} \left(\left(\frac{\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})} \right)^{x_{ijk,t}} \left(1 - \frac{\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})} \right)^{1 - x_{ijk,t}} \right)^{1 - d_{ijk,t}} \\
&= \left(\delta_{\{0\}}(x_{ijk,t}) \right)^{d_{ijk,t}} \frac{\left(\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{x_{ijk,t}(1 - d_{ijk,t})}}{\left(1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{(1 - d_{ijk,t})}} .
\end{aligned}$$

$$\begin{aligned}
&p(x_{ijk,t}, d_{ijk,t} | s_t = l, \rho_l, \mathcal{G}_l) \\
&= \rho_l^{\mathbb{1}_{\{1\}}(d_{ijk,t})} \left(\delta_{\{0\}}(x_{ijk,t}) \right)^{\mathbb{1}\{d_{ijk,t} = 1\}} (1 - \rho_l)^{\mathbb{1}_{\{0\}}(d_{ijk,t})} \left(\mathcal{B}ern(x_{ijk,t} | \eta_{ijk,t}) \right)^{\mathbb{1}_{\{0\}}(d_{ijk,t})} \\
&= \rho_l^{d_{ijk,t}} \left(\delta_{\{0\}}(x_{ijk,t}) \right)^{d_{ijk,t}} (1 - \rho_l)^{1 - d_{ijk,t}} \frac{\left(\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{x_{ijk,t}(1 - d_{ijk,t})}}{\left(1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{(1 - d_{ijk,t})}} . \quad \text{(S.2.3)}
\end{aligned}$$

The marginal distribution of the allocation variable, for each for i, j, k, t, is

$$d_{ijk,t} | s_t, \boldsymbol{\rho} \sim \mathcal{B}ern(\rho_{s_t}),$$

By (Polson et al., 2013, Theorem 1), it is possible to decompose the ratio in the right hand side of eq. (S.2.3) as follows

$$\frac{\left(\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{x_{ijk,t}(1 - d_{ijk,t})}}{\left(1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{(1 - d_{ijk,t})}} =$$

$$= 2^{-(1-d_{ijk,t})} \int_0^\infty \exp\Big(-\frac{\omega_{ijk,t}}{2}(\mathbf{z}_t'\mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t}(\mathbf{z}_t'\mathbf{g}_{ijk,l})\Big) p(\omega_{ijk,t})\, \mathrm{d}\omega_{ijk,t}\,,$$

where, for every i, j, k, t,

$$\kappa_{ijk,t} = x_{ijk,t}(1-d_{ijk,t}) - \frac{1-d_{ijk,t}}{2} = (1-d_{ijk,t})(x_{ijk,t}-1/2).$$

Therefore we get the following conditional and joint distributions, respectively

$$\begin{aligned} p(x_{ijk,t},&d_{ijk,t}|\omega_{ijk,t}, s_t = l, \rho_l, \mathcal{G}_l) = \\ &= \rho_l^{d_{ijk,t}} \Big(\frac{1-\rho_l}{2}\Big)^{1-d_{ijk,t}} \exp\Big(-\frac{\omega_{ijk,t}}{2}(\mathbf{z}_t'\mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t}(\mathbf{z}_t'\mathbf{g}_{ijk,l})\Big), \\ p(x_{ijk,t}, d_{ijk,t},&\omega_{ijk,t}|s_t = l, \rho_l, \mathcal{G}_l) = p(x_{ijk,t}, d_{ijk,t}|\omega_{ijk,t}, s_t = l, \rho_l, \mathcal{G}_l) \cdot p(\omega_{ijk,t}). \end{aligned} \tag{S.2.4}$$

Finally, the marginal distribution of each latent variable $\omega_{ijk,t}$ from the data augmentation scheme follows a Pólya-Gamma distribution:

$$\omega_{ijk,t} \sim PG(1,0).$$

A continuous random variable $x \in [0, +\infty)$ has a Pólya-Gamma distribution with parameters $b > 0$, $c \in \mathbb{R}$ if the following stochastic representation holds

$$x \sim PG(b,c) \quad \Longleftrightarrow \quad x \overset{d}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-1/2)^2 + c^2/(4\pi^2)}$$

where $g_k \overset{iid}{\sim} \mathcal{G}a(b,1)$ and $\overset{d}{=}$ stands for equality in distribution. See Polson et al. (2013) for further details. Following the assumption of first order time homogeneous Markov chain, the last term in eq. (S.2.2) factorizes as

$$p(s_t|\boldsymbol{\theta}) = \prod_{v=1}^{t} p(s_v|s_{v-1}, \Xi) = p(s_0|\Xi) \prod_{v=1}^{t} \xi_{s_{v-1},s_v} = p(s_0|\Xi) \prod_{g=1}^{L}\prod_{l=1}^{L} \xi_{g,l}^{N_{gl}(\mathbf{s}^t)} \tag{S.2.5}$$

where $N_{gl}(\mathbf{s})$ is a function counting the number of transitions from state g to state l in the vector $\mathbf{s}$, that is $N_{gl}(\mathbf{s}) = \#\{\zeta_{t-1,g} = 1, \zeta_{t,l} = 1,\ t = 1, \ldots, T\}$, $g, l = 1, \ldots, L$, where #

denotes the cardinality of a set. The complete-data likelihood for $\mathcal{X}$ is thus obtained by plugging eq. (S.2.4) and eq. (S.2.5) in eq. (S.2.2)

$$
\begin{aligned}
&L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathrm{s}|\boldsymbol{\theta}) = \\
&= \prod_{l=1}^{L} \prod_{t\in\mathcal{T}_l} \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \Big(\frac{2\rho_l \delta_{\{0\}}(x_{ijk,t})}{1-\rho_l}\Big)^{d_{ijk,t}} \frac{1-\rho_l}{2} \exp\Big(-\frac{\omega_{ijk,t}}{2}(\mathbf{z}_t'\mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t}(\mathbf{z}_t'\mathbf{g}_{ijk,l})\Big) \\
&\quad \cdot \Big(\prod_{t=1}^{T} \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} p(\omega_{ijk,t}) \Big) \cdot \Big(\prod_{g=1}^{L} \prod_{l=1}^{L} \xi_{g,l}^{N_{gl}(\mathbf{s})} \Big). \qquad \text{(S.2.6)}
\end{aligned}
$$

S.2.2 Gibbs sampler

Let $\mathbf{W}_l$ be the l-th slice of $\mathcal{W}$, i.e. the $(4 \times R)$ matrix $\mathbf{W}_l = \{w_{h,r,l}\}_{h,r}$, similarly, let $\mathbf{W}^{(r)} = \{w_{h,r,l}\}_{h,l}$. The structure of the Gibbs sampler is summarised in the following steps:

Step 1. sample latent variables from

$$
p(\mathbf{s}, \mathbf{D}, \boldsymbol{\Omega}|\boldsymbol{\mathcal{X}}, \mathcal{G}, \boldsymbol{\rho}, \boldsymbol{\Xi}) = p(\mathbf{s}|\boldsymbol{\mathcal{X}}, \mathcal{G}, \boldsymbol{\rho}, \boldsymbol{\Xi}) p(\mathbf{D}|\mathbf{s}, \mathcal{G}, \boldsymbol{\rho}) p(\boldsymbol{\Omega}|\boldsymbol{\mathcal{X}}, \mathbf{s}, \mathcal{G})
$$

where

- $p(\mathbf{s}|\boldsymbol{\mathcal{X}}, \mathcal{G}, \boldsymbol{\rho}, \boldsymbol{\Xi})$ using the Forward Filter Backward Sampler
- $p(d_{ijk,t}|s_t, \mathcal{G}_t, \rho_t) \sim \mathcal{B}ern(\tilde{p}_{d_{ijk,t}})$
- $p(\omega_{ijkv,t}|x_{ijk,t}, s_t, \mathcal{G}_t) \sim PG(1,\, \mathbf{z}_t'\mathbf{g}_{ijk,s_t})$

Step 2. sample variance hyperparameters from

$$
p(\boldsymbol{\phi}, \tau, \mathcal{W}|\mathcal{G}) = p(\boldsymbol{\phi}|\mathcal{G}, \mathcal{W}) p(\tau|\mathcal{G}, \boldsymbol{\phi}, \mathcal{W}) p(\mathcal{W}|\mathcal{G}, \boldsymbol{\lambda}, \boldsymbol{\phi}, \tau) p(\boldsymbol{\lambda}|\mathcal{W})
$$

where

- $p(\psi_r|\mathcal{G}^{(r)}, \mathbf{W}_r) \sim \mathrm{GiG}\Big(2\bar{b}^{\tau}, \sum_{h=1}^{4}\sum_{l=1}^{L} \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{w_{h,r,l}}, \bar{\alpha} - n\Big)$ then $\phi_r = \psi_r / \sum_i \psi_i$

- $p(\tau|\mathcal{G},\mathcal{W},\boldsymbol{\phi}) \sim \mathrm{GiG}\bigg(2\bar{b}^{\tau},\, \sum_{r=1}^{R}\sum_{h=1}^{4}\sum_{l=1}^{L}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\phi_r w_{h,r,l}},\, (\bar{\alpha}-n)R\bigg)$
- $p(w_{h,r,l}|\boldsymbol{\gamma}_{h,l}^{(r)},\phi_r,\tau,\lambda_l) \sim \mathrm{GiG}\bigg(\lambda_l^2,\, \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\tau\phi_r},\, 1-\frac{n_h}{2}\bigg)$
- $p(\lambda_l|\mathbf{W}_l) \propto \lambda_l^{\bar{a}_l^{\lambda}+8R-1}\exp\bigg(-\lambda_l\bar{b}_l^{\lambda}-\frac{\lambda_l^2}{2}\sum_{r=1}^{R}\sum_{h=1}^{4}w_{h,r,l}\bigg)$

Step 3. sample tensor marginals from

$$p\left(\mathcal{G}|\boldsymbol{\mathcal{X}},\mathbf{s},\boldsymbol{\phi},\tau,\mathcal{W}\right) = \prod_{l=1}^{L}\prod_{r=1}^{R}\prod_{h=1}^{4} p(\boldsymbol{\gamma}_{h,l}^{(r)}|\boldsymbol{\gamma}_{-h,l}^{(r)},\mathcal{G}_{-r,l},\boldsymbol{\phi},\tau,\mathcal{W},\boldsymbol{\mathcal{X}},\mathbf{s})$$

where

- $p(\boldsymbol{\gamma}_{h,l}^{(r)}|\boldsymbol{\gamma}_{-h,l}^{(r)},\mathcal{G}_{-r,l},\boldsymbol{\phi},\tau,\mathcal{W},\boldsymbol{\mathcal{X}},\mathbf{s}) \sim \mathcal{N}_{d_h}(\boldsymbol{\mu}_{\boldsymbol{\gamma}_{h,l}},\Sigma_{\boldsymbol{\gamma}_{h,l}})$

Step 4. sample switching parameters and transition matrix from

$$p(\rho_l,\boldsymbol{\xi}_l|\mathbf{s},\mathcal{D}) = p(\rho_l|\mathbf{s},\mathcal{D})p(\boldsymbol{\xi}_l|\mathbf{s})$$

where

- $p(\rho_l|\mathbf{s},\mathcal{D}) \sim \mathcal{B}e(\tilde{a}_l,\tilde{b}_l)$
- $p(\boldsymbol{\xi}_l|\mathbf{s}) \sim \mathcal{D}ir(\tilde{\mathbf{c}})$

S.2.3 Full conditional distribution of ϕ_r

The full conditional of the common component ϕ_r of the variance of the marginals from the PARAFAC, for each $r = 1,\dots,R$, can be obtained in closed form collapsing τ. This can be done by exploiting a result in Guhaniyogi et al. (2017), which states that the posterior full conditional of each ϕ_r can be obtained by normalising Generalised Inverse Gaussian distributed random variables ψ_r

$$\phi_r = \frac{\psi_r}{\sum_{i=1}^{R}\psi_i} \qquad \forall\, r \tag{S.2.7}$$

where

$$\psi_r \sim \mathrm{GiG}\bigg(2\bar{b}^\tau,\ \sum_{h=1}^{4}\sum_{l=1}^{L}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{w_{h,r,l}},\ \bar{\alpha}-n\bigg). \tag{S.2.8}$$

In the previous notation, $GiG(\cdot)$ stands for the Generalized Inverse Gaussian distribution. The Generalized Inverse Gaussian probability density function with three parameters $a > 0$, $b > 0$, $p \in \mathbb{R}$, for the random variable $x \in (0, +\infty)$, is given by:

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

with $K_p(\cdot)$ a modified Bessel function of the second type. The result follows from:

$$\begin{aligned}
p(\boldsymbol{\phi}|\mathcal{G},\mathcal{W}) &\propto p(\boldsymbol{\phi})\int_0^\infty p(\mathcal{G}|\mathcal{W},\boldsymbol{\phi},\tau)p(\tau)\,\mathrm{d}\tau \\
&\propto \prod_{r=1}^{R}\phi_r^{\bar{\alpha}-1}\int_0^\infty\prod_{r=1}^{R}\prod_{h=1}^{4}\prod_{l=1}^{L}\big|\tau\phi_r w_{h,r,l}\mathbf{I}_{n_h}\big|^{-1/2}\exp\Big(-\frac{1}{2}\boldsymbol{\gamma}_{h,l}^{(r)\prime}(\tau\phi_r w_{h,r,l}\mathbf{I}_{n_h})^{-1}\boldsymbol{\gamma}_{h,l}^{(r)}\Big) \\
&\cdot\tau^{a_\tau-1}\exp(-\bar{b}^\tau\tau)\,\mathrm{d}\tau \\
&= \int_0^\infty\prod_{r=1}^{R}\phi_r^{\bar{\alpha}-1}\prod_{h=1}^{4}\big|\tau\phi_r\mathbf{I}_{n_h}\big|^{-1}\exp\Big(-\frac{1}{2}\sum_{l=1}^{L}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\tau\phi_r w_{h,r,l}}\Big)\tau^{\bar{a}^\tau-1}\exp(-\bar{b}^\tau\tau)\,\mathrm{d}\tau.
\end{aligned}$$

We define $n = n_1+n_2+n_3+n_4 = I+J+K+Q$ and exploit the property $\det(kA) = k^n\det(A)$, for a square matrix A of size n and a scalar k. From the assumption $\bar{a}^\tau = \bar{\alpha}R$, we obtain

$$\begin{aligned}
&\propto \int_0^\infty\prod_{r=1}^{R}\phi_r^{\bar{\alpha}-1}\prod_{h=1}^{4}\big|\tau\phi_r\mathbf{I}_{n_h}\big|^{-1}\exp\Big(-\frac{1}{2}\sum_{l=1}^{L}(\tau\phi_r w_{h,r,l})^{-1}\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}\Big)\tau^{\bar{a}^\tau-1}\exp(-\bar{b}^\tau\tau)\,\mathrm{d}\tau \\
&\propto \int_0^\infty\prod_{r=1}^{R}(\tau\phi_r)^{\bar{\alpha}-1}(\tau\phi_r)^{-n}\exp\Big(-\frac{1}{2}\Big(2\bar{b}^\tau\tau+\sum_{h=1}^{4}\sum_{l=1}^{L}(\tau\phi_r w_{h,r,l})^{-1}\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}\Big)\Big)\,\mathrm{d}\tau \\
&= \int_0^\infty\Big(\prod_{r=1}^{R}(\tau\phi_r)^{\bar{\alpha}-n-1}\Big)\exp\Big(-\frac{1}{2}\sum_{r=1}^{R}\Big(2\bar{b}^\tau\tau\phi_r+\frac{1}{\tau\phi_r}\sum_{h=1}^{4}\sum_{l=1}^{L}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{w_{h,r,l}}\Big)\Big)\,\mathrm{d}\tau,
\end{aligned} \tag{S.2.9}$$

where in the last line we used $\sum_{r=1}^{R}\phi_r = 1$. It can be seen that the integrand is the kernel of a GiG with respect to the random variable $\psi_r = \tau\phi_r$. Following Guhaniyogi et al. (2017) and Kruijer et al. (2010), it is possible to sample from the posterior of ϕ_r, for each $r = 1, \dots, R$ by first sampling ψ_r from a GiG with kernel given in eq. (S.2.9), then normalising over r, as reported in eq. (S.2.8)-(S.2.7), respectively.

S.2.4 Full conditional distribution of τ

The full conditional of the global component of the variance of the PARAFAC marginals is

$$p(\tau|\mathcal{G}, \mathcal{W}, \boldsymbol{\phi}) \sim \mathrm{GiG}\bigg(2\bar{b}^\tau, \sum_{r=1}^{R}\sum_{h=1}^{4}\sum_{l=1}^{L} \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\phi_r w_{h,r,l}}, (\bar{\alpha} - n)R\bigg),$$

The posterior full conditional distribution is derived from

$$\begin{aligned}
p(\tau|\mathcal{G}, \mathcal{W}, \boldsymbol{\phi}) &\propto p(\tau)p(\mathcal{G}|\mathcal{W}, \boldsymbol{\phi}, \tau) \\
&\propto \tau^{\bar{a}^\tau - 1}\exp(-\bar{b}^\tau\tau)\prod_{r=1}^{R}\prod_{h=1}^{4}\prod_{l=1}^{L}\big|\tau\phi_r w_{h,r,l}\mathbf{I}_{n_h}\big|^{-1/2}\exp\bigg(-\frac{1}{2}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\phi_r w_{h,r,l}}\bigg) \\
&\propto \tau^{\bar{a}^\tau - nR - 1}\exp\bigg(-\frac{1}{2}\bigg(2\bar{b}^\tau\tau + \frac{1}{\tau}\sum_{r=1}^{R}\sum_{h=1}^{4}\sum_{l=1}^{L}\frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\phi_r w_{h,r,l}}\bigg)\bigg),
\end{aligned}$$

which is the kernel of a GiG. We sample from this distribution by a Hamiltonian Monte Carlo (Neal (2011)).

S.2.5 Full conditional distribution of $w_{h,r,l}$

The full conditional distribution of the local component of the variance of each PARAFAC marginal, for each h, r, l, is

$$p(w_{h,r,l}|\boldsymbol{\gamma}_{h,l}^{(r)}, \phi_r, \tau, \lambda_l) \sim \mathrm{GiG}\bigg(\lambda_l^2, \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime}\boldsymbol{\gamma}_{h,l}^{(r)}}{\tau\phi_r}, 1 - \frac{n_h}{2}\bigg),$$

which follows from

$$\begin{aligned}
p(w_{h,r,l}|\boldsymbol{\gamma}_{h,l}^{(r)}, \phi_r, \tau, \lambda_l) &\propto p(w_{h,r,l}|\lambda_l)p(\boldsymbol{\gamma}_{h,l}^{(r)}|w_{h,r,l}, \phi_r, \tau) \\
&\propto \exp\bigg(-\frac{\lambda_l^2}{2}w_{h,r,l}\bigg)\big|\tau\phi_r w_{h,r,l}\mathbf{I}_{n_h}\big|^{-1/2}\exp\bigg(-\frac{1}{2}\boldsymbol{\gamma}_{h,l}^{(r)\prime}(\tau\phi_r w_{h,r,l}\mathbf{I}_{n_h})^{-1}\boldsymbol{\gamma}_{h,l}^{(r)}\bigg)
\end{aligned}$$

$$\propto \exp\Big(-\frac{\lambda_l^2}{2} w_{h,r,l} \Big) w_{h,r,l}^{-n_h/2} \exp\Big(-\frac{1}{2} \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime} \boldsymbol{\gamma}_{h,l}^{(r)}}{\tau \phi_r w_{h,r,l}} \Big)$$
$$= w_{h,r,l}^{-n_h/2} \exp\Big(-\frac{1}{2} \Big(\lambda_l^2 w_{h,r,l} + \frac{1}{w_{h,r,l}} \frac{\boldsymbol{\gamma}_{h,l}^{(r)\prime} \boldsymbol{\gamma}_{h,l}^{(r)}}{\tau \phi_r} \Big) \Big).$$

S.2.6 Full conditional distribution of λ_l

The posterior distribution of λ_l is given by

$$p(\lambda_l | \mathbf{W}_l) \propto \lambda_l^{\bar{a}_l^\lambda + 8R - 1} \exp\Big(-\lambda_l \bar{b}_l^\lambda - \frac{\lambda_l^2}{2} \sum_{r=1}^{R} \sum_{h=1}^{4} w_{h,r,l} \Big).$$

It is obtained from

$$\begin{aligned} p(\lambda_l | \mathbf{W}_l) &\propto p(\lambda_l) p(\mathbf{W}_l | \lambda_l) \\ &\propto \lambda_l^{\bar{a}_l^\lambda - 1} \exp(-\bar{b}_l^\lambda \lambda_l) \prod_{r=1}^{R} \prod_{h=1}^{4} \frac{\lambda_l^2}{2} \exp\Big(-\frac{\lambda_l^2}{2} w_{h,r,l} \Big) \\ &\propto \lambda^{\bar{a}_l^\lambda + 8R - 1} \exp(-\lambda_l \bar{b}_l^\lambda) \exp\Big(-\frac{\lambda_l^2}{2} \sum_{r=1}^{R} \sum_{h=1}^{4} w_{h,r,l} \Big). \end{aligned}$$

Since the second exponential is always smaller than one due to the positiveness of all the parameters $\lambda_l, w_{h,r,l}$, we can sample from this distribution by means of an accept/reject algorithm using as proposal density a Gamma distribution $\mathcal{G}a(\tilde{a}, \tilde{b})$ with parameters:

$$\tilde{a} = \bar{a}_l^\lambda + 8R, \qquad \tilde{b} = \bar{b}_l^\lambda.$$

We sample from this distribution by a Hamiltonian Monte Carlo (Neal (2011)).

S.2.7 Full conditional distribution of $\boldsymbol{\gamma}_{h,l}^{(r)}$

For deriving the full conditional distribution of each PARAFAC marginal, $\boldsymbol{\gamma}_{h,l}^{(r)}$, of the tensor $\mathcal{G}_l$, $l = 1, \ldots, L$, we start by manipulating the complete data likelihood in eq. (S.2.6) with the aim of singling out $\boldsymbol{\gamma}_{h,l}^{(r)}$. From eq. (S.2.4), considering all the entries of $\mathcal{X}_t$ at a given $t \in \{1, \ldots, T\}$ and denoting with $p(\mathcal{G}_l)$ the prior distribution induced on $\mathcal{G}_l$ by the hierarchical prior on the PARAFAC marginals, the following proportionality relation holds

$$\begin{aligned}
p(\mathcal{G}_l|\mathcal{X}_t, \mathcal{D}_t, \mathbf{\Omega}_t, s_t = l, \rho_l) &\propto p(\mathcal{G}_l) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \exp\Big(-\frac{\omega_{ijk,t}}{2} (\mathbf{z}_t' \mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t} (\mathbf{z}_t' \mathbf{g}_{ijk,l}) \Big) p(\omega_{ijk,t}) \\
&= p(\mathcal{G}_l) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \exp\Big(-\frac{\omega_{ijk,t}}{2} \Big((\mathbf{z}_t' \mathbf{g}_{ijk,l})^2 - 2\frac{\kappa_{ijk,t}}{\omega_{ijk,t}} (\mathbf{z}_t' \mathbf{g}_{ijk,l}) \Big) \Big) p(\omega_{ijk,t}) \\
&= p(\mathcal{G}_l) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \exp\Big(-\frac{\omega_{ijk,t}}{2} \Big(\mathbf{z}_t' \mathbf{g}_{ijk,l} - \frac{\kappa_{ijk,t}}{\omega_{ijk,t}} \Big)^2 \Big) p(\omega_{ijk,t}). \qquad \text{(S.2.10)}
\end{aligned}$$

Define $u_{ijk,t} = \kappa_{ijk,t}/\omega_{ijk,t}$, then we rewrite eq. (S.2.10) in compact form as

$$\begin{aligned}
&p(\mathcal{G}_l|\mathcal{X}_t, \mathcal{D}_t, \mathbf{\Omega}_t, s_t = l, \rho_l) \propto \\
&\propto p(\mathcal{G}_l) \exp\Big(-\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \omega_{ijk,t} (\mathbf{z}_t' \mathbf{g}_{ijk,l} - u_{ijk,t})^2 \Big) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} p(\omega_{ijk,t}) \\
&= p(\mathcal{G}_l) \exp\Big(-\frac{1}{2} \sum_{i=1}^{I} (\mathcal{G}_l \times_4 \mathbf{z}_t - \mathcal{U}_t)_i' \operatorname{diag}\big(\boldsymbol{\omega}_{i:,t}\big) (\mathcal{G}_l \times_4 \mathbf{z}_t - \mathcal{U}_t)_i \Big) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} p(\omega_{ijk,t}) \\
&= p(\mathcal{G}_l) \exp\Big(-\frac{1}{2} (\operatorname{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big) - \operatorname{vec}\big(\mathcal{U}_t\big))' \operatorname{diag}\big(\operatorname{vec}\big(\mathbf{\Omega}_t\big)\big) (\operatorname{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big) - \operatorname{vec}\big(\mathcal{U}_t\big)) \Big) \\
&\quad \cdot \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} p(\omega_{ijk,t}) \\
&= p(\mathcal{G}_l) f(\mathcal{G}_l, \mathbf{z}_t, \mathcal{U}_t, \mathbf{\Omega}_t) \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} p(\omega_{ijk,t}), \qquad \text{(S.2.11)}
\end{aligned}$$

where $f(\cdot)$ is a function which contains the kernel of a multivariate normal distribution with respect to the variable $\operatorname{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big)$.

Considering eq. (S.2.6) and (S.2.11) we obtain the proportionality relation

$$L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \mathbf{\Omega}, \mathbf{s}|\boldsymbol{\theta}) = \prod_{l=1}^{L} \prod_{t \in \mathcal{T}_l} p(\mathcal{X}_t, \mathcal{D}_t, \mathbf{\Omega}_t, s_t|\boldsymbol{\theta}) \propto \prod_{l=1}^{L} \prod_{t \in \mathcal{T}_l} f(\mathcal{G}_l, \mathbf{z}_t, \mathcal{U}_t, \mathbf{\Omega}_t). \qquad \text{(S.2.12)}$$

We are now ready to compute the full conditional distributions of each vector $\boldsymbol{\gamma}_{h,l}^{(r)}$, $h = 1,\dots,4$, $l = 1,\dots,L$ and $r = 1,\dots,R$. To this aim, notice that

$$\mathcal{G}_l = \sum_{r=1}^{R} \boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)} \circ \boldsymbol{\gamma}_{4,l}^{(r)} = \mathcal{G}_l^{(r)} + \mathcal{G}_l^{(-r)} \tag{S.2.13}$$

where we have defined

$$\mathcal{G}_l^{(r)} = \boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)} \circ \boldsymbol{\gamma}_{4,l}^{(r)}, \qquad \mathcal{G}_l^{(-r)} = \sum_{\substack{v=1 \\ v\neq r}}^{R} \mathcal{G}_l^{(v)} .$$

By exploiting the definitions of mode-n product and PARAFAC decomposition, we obtain

$$\mathcal{G}_{l,t} = \mathcal{G}_l \times_4 \mathbf{z}_t = \sum_{r=1}^{R} \big(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\big)\langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle = \sum_{r=1}^{R} \mathcal{G}_{l,t}^{(r)} .$$

Here $\langle \cdot,\cdot \rangle$ denotes the standard inner product in the Euclidean space $\mathbb{R}^n$. Since the latter is a scalar, we have

$$\bar{\mathbf{g}}_{l,t} = \mathrm{vec}\big(\mathcal{G}_{l,t}\big) = \mathrm{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big) = \sum_{r=1}^{R} \mathrm{vec}\big(\mathcal{G}_{l,t}^{(r)}\big) = \sum_{r=1}^{R} \bar{\mathbf{g}}_{l,t}^{(r)} . \tag{S.2.14}$$

The vectorisation of a tensor can be expressed in the following way, which is a generalisation of a well known property holding for matrices: it consists in stacking in a column vector all the vectorised slices of the tensor. For the sake of clarity, let $\boldsymbol{\alpha}_1 \in \mathbb{R}^I$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^J$ and $\boldsymbol{\alpha}_3 \in \mathbb{R}^K$ and let the tensor $\mathcal{A} = \boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3$. Denote $\mathcal{A}_{::k} \in \mathbb{R}^{I\times J}$ the k-th frontal slice of the tensor $\mathcal{A}$. Then, by applying the properties of Kronecker product and vectorization operator we obtain[3]

$$\mathrm{vec}\big(\mathcal{A}\big) = \mathrm{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2 \circ \boldsymbol{\alpha}_3\big) = \big(\,\mathrm{vec}\big(\mathcal{A}_{::1}\big)', \dots, \mathrm{vec}\big(\mathcal{A}_{::K}\big)'\big)'$$

[3]The outer product and Kronecker products are two operators acting on

$$\begin{aligned} \circ : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_K} \quad &\to \mathbb{R}^{n_1 \times \ldots \times n_K} \\ \otimes : \mathbb{R}^{n_1 \times m_1} \times \mathbb{R}^{n_2 \times m_2} \quad &\to \mathbb{R}^{n_1 n_2 \times m_1 m_2}. \end{aligned}$$

Note that the Kronecker product is defined on the space of matrices (and vectors, as a particular case), while the outer product is defined on arrays of possible different number of dimensions (e.g. it is defined

$$= \big(\text{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2\big)' \alpha_{3,1}, \dots, \text{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2\big)' \alpha_{3,K} \big)'$$
$$= \boldsymbol{\alpha}_3 \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \circ \boldsymbol{\alpha}_2\big) = \boldsymbol{\alpha}_3 \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big). \tag{S.2.15}$$

The use of the same property allows to rewrite eq. (S.2.15) in three equivalent ways, each one written as a product of a matrix and one of the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$, respectively. In fact, we have

$$\text{vec}\big(\mathcal{A}\big) = \boldsymbol{\alpha}_3 \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big) = \boldsymbol{\alpha}_3 \otimes \big(\boldsymbol{\alpha}_2 \otimes \mathbf{I}_I\big) \,\text{vec}\big(\boldsymbol{\alpha}_1\big) = \big(\boldsymbol{\alpha}_3 \otimes \boldsymbol{\alpha}_2 \otimes \mathbf{I}_I\big)\boldsymbol{\alpha}_1 \tag{S.2.16}$$
$$\text{vec}\big(\mathcal{A}\big) = \boldsymbol{\alpha}_3 \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big) = \boldsymbol{\alpha}_3 \otimes \big[(\mathbf{I}_J \otimes \boldsymbol{\alpha}_1) \,\text{vec}\big(\boldsymbol{\alpha}_2'\big)\big] = (\boldsymbol{\alpha}_3 \otimes \mathbf{I}_J \otimes \boldsymbol{\alpha}_1)\boldsymbol{\alpha}_2 \tag{S.2.17}$$
$$\begin{aligned} \text{vec}\big(\mathcal{A}\big) &= \boldsymbol{\alpha}_3 \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big) = \text{vec}\big(\,\text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big)\boldsymbol{\alpha}_3'\big) = (\mathbf{I}_K \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big)) \,\text{vec}\big(\boldsymbol{\alpha}_3'\big) \\ &= (\mathbf{I}_K \otimes \text{vec}\big(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2'\big))\boldsymbol{\alpha}_3 = (\mathbf{I}_K \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_1)\boldsymbol{\alpha}_3. \end{aligned} \tag{S.2.18}$$

Starting from eq. (S.2.14), we can apply for $\boldsymbol{\gamma}_{1,l}^{(r)}, \dots, \boldsymbol{\gamma}_{3,l}^{(r)}$ the same argument as for $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_3$, with the aim of manipulating the likelihood function and obtain three different expressions where the dependence on $\boldsymbol{\gamma}_{1,l}^{(r)}, \boldsymbol{\gamma}_{2,l}^{(r)}, \boldsymbol{\gamma}_{3,l}^{(r)}$, respectively, is made explicit. This will be used for deriving the posterior full conditional distributions of the PARAFAC marginals. Thus,

between two vectors, and returns a matrix, as well as between a vector and a matrix, yielding a third order tensor). When dealing with two vectors $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$, their outer product and Kronecker product are related and given by, respectively

$$\mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}' \in \mathbb{R}^{n \times m}$$
$$\mathbf{u} \otimes \mathbf{v} = \text{vec}\big(\mathbf{v}\mathbf{u}'\big) = \text{vec}\big(\mathbf{v} \circ \mathbf{u}\big) \in \mathbb{R}^{nm}.$$

For two matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times k}$ it holds

$$\text{vec}\big(\mathbf{AB}\big) = (\mathbf{I}_k \otimes \mathbf{A}) \,\text{vec}\big(\mathbf{B}\big) = (\mathbf{B}' \otimes \mathbf{I}_m) \,\text{vec}\big(\mathbf{A}\big) \in \mathbb{R}^{mk \times 1}.$$

Moreover, if $n = 1$ then $\mathbf{B}$ is a row vector of length k, as a consequence $\mathbf{B}' = \text{vec}\big(\mathbf{B}\big) \in \mathbb{R}^{k \times 1}$. See (Cichocki et al., 2009, p.31).

from eq. (S.2.14) we have

$$\bar{\mathbf{g}}_{l,t}^{(r)} = \text{vec}\left(\mathcal{G}_{l,t}^{(r)}\right) = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \,\text{vec}\left(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\right) = \text{vec}\left(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\right)\mathbf{z}_t' \boldsymbol{\gamma}_{4,l}^{(r)} = \mathbf{A}_4 \boldsymbol{\gamma}_{4,l}^{(r)}, \tag{S.2.19}$$

where

$$\mathbf{A}_4 = \text{vec}\left(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\right)\mathbf{z}_t'.$$

Exploiting eq. (S.2.16) we have

$$\bar{\mathbf{g}}_{l,t}^{(r)} = \text{vec}\left(\mathcal{G}_{l,t}^{(r)}\right) = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I\right)\boldsymbol{\gamma}_{1,l}^{(r)} = \mathbf{A}_1 \boldsymbol{\gamma}_{1,l}^{(r)}, \tag{S.2.20}$$

with

$$\mathbf{A}_1 = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I\right).$$

Exploiting eq. (S.2.17) we have

$$\bar{\mathbf{g}}_{l,t}^{(r)} = \text{vec}\left(\mathcal{G}_{l,t}^{(r)}\right) = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\right)\boldsymbol{\gamma}_{2,l}^{(r)} = \mathbf{A}_2 \boldsymbol{\gamma}_{2,l}^{(r)}, \tag{S.2.21}$$

with

$$\mathbf{A}_2 = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\right).$$

Finally, using eq. (S.2.18) we obtain

$$\bar{\mathbf{g}}_{l,t}^{(r)} = \text{vec}\left(\mathcal{G}_{l,t}^{(r)}\right) = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\mathbf{I}_K \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\right)\boldsymbol{\gamma}_{3,l}^{(r)} = \mathbf{A}_3 \boldsymbol{\gamma}_{3,l}^{(r)}, \tag{S.2.22}$$

with

$$\mathbf{A}_3 = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \left(\mathbf{I}_K \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\right).$$

Using the definition of $f(\mathcal{G}_l, \mathbf{z}_t, \mathcal{U}_t^{(l)}, \boldsymbol{\Omega}_t)$, eq. (S.2.14) and eq. (S.2.13) we can thus write

$$\text{vec}\left(\mathcal{G}_l \times_4 \mathbf{z}_t\right) = \bar{\mathbf{g}}_{l,t}^{(r)} + \sum_{\substack{v=1 \\ v\neq r}}^{R} \bar{\mathbf{g}}_{l,t}^{(v)} = \bar{\mathbf{g}}_{l,t}^{(r)} + \bar{\mathbf{g}}_{l,t}^{(-r)}.$$

From eq. (S.2.12), by focusing on regime $l \in \{1, \ldots, L\}$, we get

$$L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}|\boldsymbol{\theta}) \propto$$

$$\propto \exp\Big(-\frac{1}{2}\big(\operatorname{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big) - \operatorname{vec}\big(\mathcal{U}_t\big)\big)' \operatorname{diag}\big(\operatorname{vec}\big(\boldsymbol{\Omega}_t\big)\big)\big(\operatorname{vec}\big(\mathcal{G}_l \times_4 \mathbf{z}_t\big) - \operatorname{vec}\big(\mathcal{U}_t\big)\big)\Big)$$
$$= \exp\Big(-\frac{1}{2}\big(\bar{\mathbf{g}}_{l,t}^{(r)} + \bar{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t\big)'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\bar{\mathbf{g}}_{l,t}^{(r)} + \bar{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t\big)\Big) \tag{S.2.23}$$

where, for reducing the burden of notation, we have defined $\mathbf{u}_t = \operatorname{vec}\big(\mathcal{U}_t\big)$ and $\bar{\bar{\boldsymbol{\Omega}}}_t = \operatorname{diag}\big(\operatorname{vec}\big(\boldsymbol{\Omega}_t\big)\big)$. We can now single out a specific component $\mathcal{G}_l^{(r)}$ of the PARAFAC decomposition of the tensor $\mathcal{G}$, which is incorporated in $\bar{\mathbf{g}}_{l,t}^{(r)}$. In fact, we can manipulate the function in eq. (S.2.23) to obtain, for each $l = 1, \ldots, L$

$$L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}|\boldsymbol{\theta}) \propto \prod_{t\in\mathcal{T}_l} \exp\left(-\frac{1}{2}\left(\bar{\mathbf{g}}_{l,t}^{(r)\prime}\,\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} + 2\mathbf{g}_{l,t}^{(r)\prime}\,\bar{\bar{\boldsymbol{\Omega}}}_t(\bar{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t) + (\bar{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t)'\,\bar{\bar{\boldsymbol{\Omega}}}_t(\bar{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t)\right)\right)$$
$$\propto \prod_{t\in\mathcal{T}_l} \exp\Big(-\frac{1}{2}\big(\bar{\mathbf{g}}_{l,t}^{(r)\prime}\,\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} - 2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\,\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)}\big)\Big). \tag{S.2.24}$$

S.2.7.1 Full conditional distribution of $\boldsymbol{\gamma}_{1,l}^{(r)}$

The full conditional distribution of $\boldsymbol{\gamma}_{1,l}^{(r)}$ is

$$p(\boldsymbol{\gamma}_{1,l}^{(r)}|\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}, \boldsymbol{\gamma}_{2,l}^{(r)}, \boldsymbol{\gamma}_{3,l}^{(r)}, \boldsymbol{\gamma}_{4,l}^{(r)}, \mathcal{G}_l^{(-r)}, w_{1,r}, \phi_r, \tau) \propto \mathcal{N}_I(\tilde{\boldsymbol{\zeta}}_{1,l}^r, \tilde{\boldsymbol{\Lambda}}_{1,l}^r)$$

where

$$\tilde{\boldsymbol{\Lambda}}_{1,l}^r = \Big((\tau\phi_r w_{1,r}\mathbf{I}_I)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)^{-1}$$
$$\tilde{\boldsymbol{\zeta}}_{1,l}^r = \tilde{\boldsymbol{\Lambda}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}(\tau\phi_r w_{1,r}\mathbf{I}_I)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)'.$$

By exploiting the rightmost term in the equality chain in eq. (S.2.20), we can simplify the two addenda in eq. (S.2.24) as

$$\begin{aligned}
\bar{\mathbf{g}}_{l,t}^{(r)\prime}\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= \Big(\mathbf{A}_1\boldsymbol{\gamma}_{1,l}^{(r)}\Big)'\bar{\bar{\boldsymbol{\Omega}}}_t\Big(\mathbf{A}_1\boldsymbol{\gamma}_{1,l}^{(r)}\Big) \\
&= \langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I\Big)'\bar{\bar{\boldsymbol{\Omega}}}_t\Big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I\Big)\boldsymbol{\gamma}_{1,l}^{(r)}\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle \\
&= \boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime} \otimes \boldsymbol{\gamma}_{2,l}^{(r)\prime} \otimes \mathbf{I}_I'\Big)\bar{\bar{\boldsymbol{\Omega}}}_t\Big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I\Big)(\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle)^2\Big)\boldsymbol{\gamma}_{1,l}^{(r)}
\end{aligned}$$

$$= \boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} \tag{S.2.25}$$

and

$$\begin{aligned}
-2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= -2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\boldsymbol{\gamma}_{3,l}^{(r)}\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\mathbf{I}_I\big)\boldsymbol{\gamma}_{1,l}^{(r)}\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle \\
&= -2\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\boldsymbol{\gamma}_{3,l}^{(r)}\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\mathbf{I}_I\big)\boldsymbol{\gamma}_{1,l}^{(r)} \\
&= -2\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}.
\end{aligned} \tag{S.2.26}$$

By applying Bayes' rule and plugging eq. (S.2.25) and eq. (S.2.26) into eq. (S.2.24) we get

$$\begin{aligned}
p(\boldsymbol{\gamma}_{1,l}^{(r)}|-) &\propto L(\boldsymbol{\mathcal{X}},\mathcal{D},\boldsymbol{\Omega},\mathbf{s}|\boldsymbol{\theta})p(\boldsymbol{\gamma}_{1,l}^{(r)}|\mathcal{W},\boldsymbol{\phi},\tau) \\
&\propto \prod_{t\in\mathcal{T}_l}\exp\Big(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}\Big)\Big) \\
&\cdot\exp\Big(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}\Big)\Big) \\
&= \exp\Big(-\frac{1}{2}\Big(\sum_{t\in\mathcal{T}_l}\Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}\Big) \\
&\quad + \Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}\Big)\Big)\Big) \\
&= \exp\Big(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{1,l}^{(r)} - 2\Big(\sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{1,l}^{(r)} \\
&\quad + \boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{1,l}^{(r)}\Big)\Big) \\
&= \exp\Big(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{1,l}^{(r)\prime}\Big(\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{1,l}^{(r)} \\
&\quad - 2\Big(\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{1,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{1,l}^{(r)}\Big)\Big).
\end{aligned}$$

This is the kernel of a multivariate normal distribution with parameters

$$\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{1,l}^{r} &= \Big((\tau\phi_r w_{1,r}\mathbf{I}_I)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)^{-1} \\
\tilde{\boldsymbol{\zeta}}_{1,l}^{r} &= \tilde{\boldsymbol{\Lambda}}_{1,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{1,l}^{r\prime}(\tau\phi_r w_{1,r}\mathbf{I}_I)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{1,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{1,l,t}^{(r)}\Big)^{-1}\Big)'.
\end{aligned}$$

S.2.7.2 Full conditional distribution of $\boldsymbol{\gamma}_{2,l}^{(r)}$

The full conditional distribution of $\boldsymbol{\gamma}_{2,l}^{(r)}$ is

$$p(\boldsymbol{\gamma}_{2,l}^{(r)}|\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}, \boldsymbol{\gamma}_{1,l}^{(r)}, \boldsymbol{\gamma}_{3,l}^{(r)}, \boldsymbol{\gamma}_{4,l}^{(r)}, \mathcal{G}_l^{(-r)}, w_{2,r}, \phi_r, \tau) \propto \mathcal{N}_J(\tilde{\boldsymbol{\zeta}}_{2,l}^r, \tilde{\boldsymbol{\Lambda}}_{2,l}^r)$$

where

$$\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{2,l}^r &= \Big((\tau\phi_r w_{2,r}\mathbf{I}_J)^{-1} + \sum_{t\in\mathcal{T}_l} \Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)^{-1} \\
\tilde{\boldsymbol{\zeta}}_{2,l}^r &= \tilde{\boldsymbol{\Lambda}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}(\tau\phi_r w_{2,r}\mathbf{I}_J)^{-1} + \sum_{t\in\mathcal{T}_l} \bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)'.
\end{aligned}$$

By exploiting the central term in the equality chain in eq. (S.2.21), we can simplify the two addenda in eq. (S.2.24) as

$$\begin{aligned}
\bar{\mathbf{g}}_{l,t}^{(r)\prime}\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= \Big(\mathbf{A}_2\boldsymbol{\gamma}_{2,l}^{(r)}\Big)'\bar{\bar{\boldsymbol{\Omega}}}_t\Big(\mathbf{A}_2\boldsymbol{\gamma}_{2,l}^{(r)}\Big) \\
&= \langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle\boldsymbol{\gamma}_{2,l}^{(r)\prime}\big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\big)'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{2,l}^{(r)}\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle \\
&= \boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\big(\boldsymbol{\gamma}_{3,l}^{(r)\prime} \otimes \mathbf{I}_J' \otimes \boldsymbol{\gamma}_{1,l}^{(r)\prime}\big)\bar{\bar{\boldsymbol{\Omega}}}_t\big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\big)(\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle)^2\Big)\boldsymbol{\gamma}_{2,l}^{(r)} \\
&= \boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} \qquad \text{(S.2.27)}
\end{aligned}$$

and

$$\begin{aligned}
-2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= -2(\mathbf{u}_t - \mathbf{g}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle\big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{2,l}^{(r)} \\
&= -2\langle\boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t\rangle(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{2,l}^{(r)} \\
&= -2\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}. \qquad \text{(S.2.28)}
\end{aligned}$$

By applying Bayes' rule and plugging eq. (S.2.27) and eq. (S.2.28) into eq. (S.2.24) we get

$$\begin{aligned}
p(\boldsymbol{\gamma}_{2,l}^{(r)}|-) &\propto L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}|\boldsymbol{\theta})p(\boldsymbol{\gamma}_{2,l}^{(r)}|\mathcal{W}, \boldsymbol{\phi}, \tau) \\
&\propto \prod_{t\in\mathcal{T}_l} \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}\Big)\bigg)
\end{aligned}$$

$$
\begin{aligned}
&\cdot \exp\Bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}\Big)\Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\Big(\sum_{t\in\mathcal{T}_l}\Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}\Big) \\
&\quad + \Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}\Big)\Big)\Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{2,l}^{(r)} - 2\Big(\sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{2,l}^{(r)} \\
&\quad + \boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{2,l}^{(r)}\Big)\Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{2,l}^{(r)\prime}\Big(\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{2,l}^{(r)} \\
&\quad - 2\Big(\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{2,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{2,l}^{(r)}\Big)\Bigg).
\end{aligned}
$$

This is the kernel of a multivariate normal distribution with parameters

$$
\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{2,l}^{r} &= \Big((\tau\phi_r w_{2,r}\mathbf{I}_J)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)^{-1} \\
\tilde{\boldsymbol{\zeta}}_{2,l}^{r} &= \tilde{\boldsymbol{\Lambda}}_{2,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{2,l}^{r\prime}(\tau\phi_r w_{2,r}\mathbf{I}_J)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{2,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{2,l,t}^{(r)}\Big)^{-1}\Big)'.
\end{aligned}
$$

S.2.7.3 Full conditional distribution of $\boldsymbol{\gamma}_{3,l}^{(r)}$

The full conditional distribution of $\boldsymbol{\gamma}_{3,l}^{(r)}$ is

$$
p(\boldsymbol{\gamma}_{3,l}^{(r)}|\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}, \boldsymbol{\gamma}_{1,l}^{(r)}, \boldsymbol{\gamma}_{2,l}^{(r)}, \boldsymbol{\gamma}_{4,l}^{(r)}, \mathcal{G}_l^{(-r)}, w_{3,r}, \phi_r, \tau) \propto \mathcal{N}_K(\tilde{\boldsymbol{\zeta}}_{3,l}^{r}, \tilde{\boldsymbol{\Lambda}}_{3,l}^{r})
$$

where

$$
\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{3,l}^{r} &= \Big((\tau\phi_r w_{3,r}\mathbf{I}_K)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)^{-1} \\
\tilde{\boldsymbol{\zeta}}_{3,l}^{r} &= \tilde{\boldsymbol{\Lambda}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}(\tau\phi_r w_{3,r}\mathbf{I}_K)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)'.
\end{aligned}
$$

By exploiting the rightmost term in the equality chain in eq. (S.2.22), we can simplify the two addenda in eq. (S.2.24) as

$$
\begin{aligned}
\bar{\mathbf{g}}_{l,t}^{(r)\prime}\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= \Big(\mathbf{A}_3\boldsymbol{\gamma}_{3,l}^{(r)}\Big)'\bar{\bar{\boldsymbol{\Omega}}}_t\Big(\mathbf{A}_3\boldsymbol{\gamma}_{3,l}^{(r)}\Big) \\
&= \langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle\boldsymbol{\gamma}_{3,l}^{(r)\prime}\big(\mathbf{I}_K\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\boldsymbol{\gamma}_{1,l}^{(r)}\big)'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\mathbf{I}_K\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{3,l}^{(r)}\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle \\
&= \boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\big(\mathbf{I}_K'\otimes\boldsymbol{\gamma}_{2,l}^{(r)\prime}\otimes\boldsymbol{\gamma}_{1,l}^{(r)\prime}\big)\bar{\bar{\boldsymbol{\Omega}}}_t\big(\mathbf{I}_K\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\boldsymbol{\gamma}_{1,l}^{(r)}\big)(\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle)^2\Big)\boldsymbol{\gamma}_{3,l}^{(r)} \\
&= \boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}
\end{aligned} \tag{S.2.29}
$$

and

$$
\begin{aligned}
-2(\mathbf{u}_t-\bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= -2(\mathbf{u}_t-\bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\mathbf{I}_K\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{3,l}^{(r)}\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle \\
&= -2\langle\boldsymbol{\gamma}_{4,l}^{(r)},\mathbf{z}_t\rangle(\mathbf{u}_t-\bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\mathbf{I}_K\otimes\boldsymbol{\gamma}_{2,l}^{(r)}\otimes\boldsymbol{\gamma}_{1,l}^{(r)}\big)\boldsymbol{\gamma}_{3,l}^{(r)} \\
&= -2\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}.
\end{aligned} \tag{S.2.30}
$$

By applying Bayes' rule and plugging eq. (S.2.29) and eq. (S.2.30) into eq. (S.2.24) we get

$$
\begin{aligned}
p(\boldsymbol{\gamma}_{3,l}^{(r)}|-) &\propto L(\boldsymbol{\mathcal{X}},\mathcal{D},\boldsymbol{\Omega},\mathbf{s}|\boldsymbol{\theta})p(\boldsymbol{\gamma}_{3,l}^{(r)}|\mathcal{W},\boldsymbol{\phi},\tau) \\
&\propto \prod_{t\in\mathcal{T}_l}\exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}-2\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}\Big)\bigg) \\
&\quad\cdot\exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}-2\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\sum_{t\in\mathcal{T}_l}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}-2\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}\Big) \\
&\quad+\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}-2\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}\Big)\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{3,l}^{(r)}-2\Big(\sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{3,l}^{(r)} \\
&\quad+\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}-2\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{3,l}^{(r)}\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1}+\sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{3,l}^{(r)}
\end{aligned}
$$

$$- 2\Big(\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{3,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l} \bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{3,l}^{(r)}\Big)\Big).$$

This is the kernel of a multivariate normal distribution with parameters

$$\tilde{\boldsymbol{\Lambda}}_{3,l}^{r} = \Big((\tau\phi_r w_{3,r}\mathbf{I}_K)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)^{-1}$$
$$\tilde{\boldsymbol{\zeta}}_{3,l}^{r} = \tilde{\boldsymbol{\Lambda}}_{3,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{3,l}^{r\prime}(\tau\phi_r w_{3,r}\mathbf{I}_K)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{3,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{3,l,t}^{(r)}\Big)^{-1}\Big)'.$$

S.2.7.4 Full conditional distribution of $\boldsymbol{\gamma}_{4,l}^{(r)}$

The full conditional distribution of $\boldsymbol{\gamma}_{4,l}^{(r)}$ is

$$p(\boldsymbol{\gamma}_{4,l}^{(r)}|\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}, \boldsymbol{\gamma}_{1,l}^{(r)}, \boldsymbol{\gamma}_{2,l}^{(r)}, \boldsymbol{\gamma}_{3,l}^{(r)}, \mathcal{G}_l^{(-r)}, w_{4,r}, \phi_r, \tau) \propto \mathcal{N}_Q(\tilde{\boldsymbol{\zeta}}_{4,l}^{r}, \tilde{\boldsymbol{\Lambda}}_{4,l}^{r})$$

where

$$\tilde{\boldsymbol{\Lambda}}_{4,l}^{r} = \Big((\tau\phi_r w_{4,r}\mathbf{I}_Q)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)^{-1}$$
$$\tilde{\boldsymbol{\zeta}}_{4,l}^{r} = \tilde{\boldsymbol{\Lambda}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}(\tau\phi_r w_{4,r}\mathbf{I}_Q)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)'.$$

By exploiting the central term in the equality chain in eq. (S.2.19), we can simplify the two addenda in eq. (S.2.24) as

$$\begin{aligned}
\bar{\mathbf{g}}_{l,t}^{(r)\prime}\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= \big(\mathbf{A}_4\boldsymbol{\gamma}_{4,l}^{(r)}\big)'\bar{\bar{\boldsymbol{\Omega}}}_t\big(\mathbf{A}_4\boldsymbol{\gamma}_{4,l}^{(r)}\big) \\
&= \boldsymbol{\gamma}_{4,l}^{(r)\prime}\mathbf{z}_t \operatorname{vec}\big(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\big)\bar{\bar{\boldsymbol{\Omega}}}_t \operatorname{vec}\big(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\big)\mathbf{z}_t'\boldsymbol{\gamma}_{4,l}^{(r)} \\
&= \boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} \qquad \text{(S.2.31)}
\end{aligned}$$

and

$$\begin{aligned}
-2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t\bar{\mathbf{g}}_{l,t}^{(r)} &= -2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})'\bar{\bar{\boldsymbol{\Omega}}}_t \operatorname{vec}\big(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}\big)\mathbf{z}_t'\boldsymbol{\gamma}_{4,l}^{(r)} \\
&= -2\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}. \qquad \text{(S.2.32)}
\end{aligned}$$

By applying Bayes' rule and plugging eq. (S.2.31) and eq. (S.2.32) into eq. (S.2.24) we get

$$
\begin{aligned}
p(\boldsymbol{\gamma}_{4,l}^{(r)}|-) &\propto L(\boldsymbol{\mathcal{X}}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s}|\boldsymbol{\theta}) p(\boldsymbol{\gamma}_{4,l}^{(r)}|\mathcal{W}, \boldsymbol{\phi}, \tau) \\
&\propto \prod_{t\in\mathcal{T}_l} \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}\Big)\bigg) \\
&\quad \cdot \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\sum_{t\in\mathcal{T}_l}\Big(\boldsymbol{\gamma}_{3,l}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} - 2\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}\Big) \\
&\quad + \Big(\boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}\Big)\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{4,l}^{(r)} - 2\Big(\sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{4,l}^{(r)} \\
&\quad + \boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)} - 2\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1}\boldsymbol{\gamma}_{4,l}^{(r)}\Big)\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big(\boldsymbol{\gamma}_{4,l}^{(r)\prime}\Big(\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{4,l}^{(r)} \\
&\quad - 2\Big(\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\Lambda}}_{4,l}^{r}\Big)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)\boldsymbol{\gamma}_{4,l}^{(r)}\Big)\bigg).
\end{aligned}
$$

This is the kernel of a multivariate normal distribution with parameters

$$
\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{4,l}^{r} &= \Big((\tau\phi_r w_{4,r}\mathbf{I}_Q)^{-1} + \sum_{t\in\mathcal{T}_l}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)^{-1} \\
\tilde{\boldsymbol{\zeta}}_{4,l}^{r} &= \tilde{\boldsymbol{\Lambda}}_{4,l}^{r\prime}\Big(\bar{\boldsymbol{\zeta}}_{4,l}^{r\prime}(\tau\phi_r w_{4,r}\mathbf{I}_Q)^{-1} + \sum_{t\in\mathcal{T}_l}\bar{\boldsymbol{\mu}}_{4,l,t}^{(r)\prime}\Big(\bar{\bar{\boldsymbol{\Sigma}}}_{4,l,t}^{(r)}\Big)^{-1}\Big)'.
\end{aligned}
$$

S.2.8 Full conditional distribution of $\omega_{ijk,t}$

The full conditional distribution for the latent variable $\omega_{ijk,t}$ for every i, j, k, t is

$$
p(\omega_{ijk,t}|x_{ijk,t}, s_t, \mathcal{G}_{s_t}) \sim PG(1,\, \mathbf{z}_t'\mathbf{g}_{ijk,s_t}).
$$

To shorten the notation, define $\psi_{ijk,t} = \mathbf{z}_t'\mathbf{g}_{ijk,s_t}$. The full conditional is derived by integrating out the latent allocation variable $d_{ijk,t}$, as follows

$$
\begin{aligned}
&p(\omega_{ijk,t}|x_{ijk,t}, s_t, \mathcal{G}_{s_t}) \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int p(\omega_{ijk,t}, d_{ijk,t}|x_{ijk,t}, s_t, \mathcal{G}_{s_t}, \rho_{s_t}) p(\rho_{s_t}) \,\mathrm{d}\rho_{s_t} \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int \frac{p(x_{ijk,t}, d_{ijk,t}|\omega_{ijk,t}, s_t, \mathcal{G}_{s_t}, \rho_{s_t}) p(\omega_{ijk,t}) p(\rho_{s_t})}{\int_\Omega p(x_{ijk,t}, \omega_{ijk,t}, d_{ijk,t}|s_t, \mathcal{G}_{s_t}, \rho_{s_t}) \,\mathrm{d}\omega_{ijk,t}} \,\mathrm{d}\rho_{s_t} \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int \frac{p(x_{ijk,t}, \omega_{ijk,t}, d_{ijk,t}|s_t, \mathcal{G}_{s_t}, \rho_{s_t})}{p(x_{ijk,t}, d_{ijk,t}|s_t, \mathcal{G}_{s_t}, \rho_{s_t})} p(\rho_{s_t}) \,\mathrm{d}\rho_{s_t} \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int \frac{(\rho_{s_t}\delta_{\{0\}}(x_{ijk,t}))^{d_{ijk,t}} \left(\frac{1-\rho_{s_t}}{2}\right)^{d_{ijk,t}} \exp\Big(-\frac{\omega_{ijk,t}}{2}\psi_{ijk,t}^2 + \kappa_{ijk,t}\psi_{ijk,t}\Big) p(\omega_{ijk,t}) p(\rho_{s_t})}{(\rho_{s_t}\delta_{\{0\}}(x_{ijk,t}))^{d_{ijk,t}} \left(\frac{1-\rho_{s_t}}{2}\right)^{d_{ijk,t}} (\exp(\psi_{ijk,t}x_{ijk,t})/(1+\exp(\psi_{ijk,t})))^{1-d_{ijk,t}}} \,\mathrm{d}\rho_{s_t} \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int \exp(\kappa_{ijk,t}^{(s_t)}\psi_{ijk,t}) \frac{\exp(\psi_{ijk,t}x_{ijk,t}(1-d_{ijk,t}))}{(1+\exp(\psi_{ijk,t}))^{1-d_{ijk,t}}} \exp\Big(-\frac{\omega_{ijk,t}}{2}\psi_{ijk,t}^2\Big) p(\omega_{ijk,t}) p(\rho_{s_t}) \,\mathrm{d}\rho_{s_t} \\
&= \sum_{d_{ijk,t}\in\{0,1\}} \int \left(\frac{1+\exp(\psi_{ijk,t})}{\exp(\psi_{ijk,t}/2)}\right)^{1-d_{ijk,t}} \Big(\exp(-\psi_{ijk,t}^2\omega_{ijk,t}/2) p(\omega_{ijk,t})\Big) p(\rho_{s_t}) \,\mathrm{d}\rho_{s_t} \\
&= \Big(1 + \frac{1+\exp(\psi_{ijk,t})}{\exp(\psi_{ijk,t}/2)}\Big)\Big(\exp(-\psi_{ijk,t}^2\omega_{ijk,t}/2) p(\omega_{ijk,t})\Big) \\
&\propto \exp(-\psi_{ijk,t}^2\omega_{ijk,t}/2) p(\omega_{ijk,t}).
\end{aligned}
$$

Since $p(\omega_{ijk,t}) \sim PG(1, 0)$, by in (Polson et al., 2013, Theorem 1) the result follows.

S.2.9 Full conditional distribution of $d_{ijk,t}$

The full conditional posterior probabilities for the latent allocation variables $d_{ijk,t}$, which select the component of the mixture in eq. (2), for each i, j, k, t, are given by

$$
\begin{aligned}
p(d_{ijk,t} = 1|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) &= \frac{\tilde{p}(d_{ijk,t} = 1|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t})}{\tilde{p}(d_{ijk,t} = 1|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) + \tilde{p}(d_{ijk,t} = 0|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t})} \\
p(d_{ijk,t} = 0|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) &= \frac{\tilde{p}(d_{ijk,t} = 0|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t})}{\tilde{p}(d_{ijk,t} = 1|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) + \tilde{p}(d_{ijk,t} = 0|\boldsymbol{\mathcal{X}}, \mathrm{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t})}.
\end{aligned}
$$

The unnormalised posterior probabilities are given by

$$
\begin{aligned}
\tilde{p}(d_{ijk,t}=1|\boldsymbol{\mathcal{X}},\mathrm{s},\mathcal{G}_{s_t},\boldsymbol{\rho}_{s_t}) &= \rho_{s_t}\delta_{\{0\}}(x_{ijk,t}) \\
\tilde{p}(d_{ijk,t}=0|\boldsymbol{\mathcal{X}},\mathrm{s},\mathcal{G}_{s_t},\boldsymbol{\rho}_{s_t}) &= (1-\rho_{s_t})\frac{\exp\big((\mathbf{z}_t'\mathbf{g}_{ijk,s_t})x_{ijk,t}\big)}{1+\exp(\mathbf{z}_t'\mathbf{g}_{ijk,s_t})}.
\end{aligned}
$$

The result starting from eq. (S.2.6) by integrating out the latent variables $\boldsymbol{\Omega}$, as follows

$$
\begin{aligned}
\tilde{p}(d_{ijk,t}|\boldsymbol{\mathcal{X}},\mathrm{s},\mathcal{G}_{s_t},\boldsymbol{\rho}_{s_t}) &\propto p(\boldsymbol{\mathcal{X}},\mathrm{s}|\mathcal{G}_{s_t},\boldsymbol{\rho}_{s_t},d_{ijk,t})p(d_{ijk,t}) \\
&\propto \rho_{s_t}^{d_{ijk,t}}\delta_{\{0\}}(x_{ijk,t})^{d_{ijk,t}}(1-\rho_{s_t})^{1-d_{ijk,t}}\frac{(\exp(\mathbf{z}_t'\mathbf{g}_{ijk,s_t}))^{x_{ijk,t}(1-d_{ijk,t})}}{(1+\exp(\mathbf{z}_t'\mathbf{g}_{ijk,s_t}))^{(1-d_{ijk,t})}} \\
&= \big(\rho_{s_t}\delta_{\{0\}}(x_{ijk,t})\big)^{d_{ijk,t}}\left((1-\rho_{s_t})\frac{(\exp(\mathbf{z}_t'\mathbf{g}_{ijk,s_t}))^{x_{ijk,t}}}{1+\exp(\mathbf{z}_t'\mathbf{g}_{ijk,s_t})}\right)^{1-d_{ijk,t}}.
\end{aligned}
$$

S.2.10 Full conditional distribution of ρ_l

For each regime $l=1,\dots,L$, the full conditional distribution for the mixing probability ρ_l of the observation model in eq (2) is

$$
p(\rho_l|\boldsymbol{\mathcal{X}},\mathcal{D},\mathrm{s}) = p(\rho_l|\mathcal{D},\mathrm{s}) \propto \mathcal{B}e(\tilde{a}_l^\rho,\tilde{b}_l^\rho)
$$

with

$$
\tilde{a}_l = N_1^l + \bar{a}_l^\rho, \qquad \tilde{b}_l = N_0^l + \bar{b}_l^\rho.
$$

We get this result starting from eq. (S.2.6) and integrating out the latent variables $\boldsymbol{\Omega}$, as follows

$$
\begin{aligned}
p(\rho_l|\boldsymbol{\mathcal{X}},\mathcal{D},\mathrm{s}) &\propto p(\rho_l)\int L(\boldsymbol{\mathcal{X}},\mathcal{D},\mathrm{s}|\rho_l,\mathcal{G}_l)p(\mathcal{G}_l)\,\mathrm{d}\mathcal{G}_l \\
&\propto \Bigg(\int\prod_{t\in\mathcal{T}_l}\prod_{i=1}^I\prod_{j=1}^J\prod_{k=1}^K \rho_l^{d_{ijk,t}}\big(\delta_{\{0\}}(x_{ijk,t})\big)^{d_{ijk,t}}(1-\rho_l)^{1-d_{ijk,t}} \\
&\quad\cdot\frac{\big(\exp(\mathbf{z}_{ijk,t}'\mathbf{g}_{ijk,l})\big)^{x_{ijk,t}(1-d_{ijk,t})}}{\big(1+\exp(\mathbf{z}_{ijk,t}'\mathbf{g}_{ijk,l})\big)^{(1-d_{ijk,t})}}\,\mathrm{d}\mathcal{G}_l\Bigg)\rho_l^{\bar{a}_l^\rho-1}(1-\rho_l)^{\bar{b}_l^\rho-1}
\end{aligned}
$$

$$
\begin{aligned}
&\propto \left(\prod_{t\in\mathcal{T}_l} \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \rho_l^{d_{ijk,t}} (1-\rho_l)^{1-d_{ijk,t}} \right) \rho_l^{\bar{a}_l^\rho - 1} (1-\rho_l)^{\bar{b}_l^\rho - 1} \\
&= \rho_l^{N_1^l} (1-\rho_l)^{N_0^l} \rho_l^{\bar{a}_l^\rho - 1} (1-\rho_l)^{\bar{b}_l^\rho - 1} \\
&= \rho_l^{N_1^l + \bar{a}_l^\rho - 1} (1-\rho_l)^{N_0^l + \bar{b}_l^\rho - 1},
\end{aligned}
$$

where we have defined the counting variables

$$
N_1^l = \sum_{t\in\mathcal{T}_l} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \mathbb{1}_{\{1\}}(d_{ijk,t}), \qquad N_0^l = \sum_{t\in\mathcal{T}_l} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \mathbb{1}_{\{0\}}(d_{ijk,t}).
$$

S.2.11 Full conditional distribution of $\boldsymbol{\xi}_l$

The full conditional distribution of each row $l = 1, \dots, L$ of the transition matrix of the hidden Markov chain is given by

$$
p(\boldsymbol{\xi}_l | \mathbf{s}) \propto \mathcal{D}ir(\tilde{\mathbf{c}}),
$$

where

$$
\tilde{\mathbf{c}} = \left(\bar{c}_1 + N_{l,1}(\mathbf{s}), \dots, \bar{c}_L + N_{l,L}(\mathbf{s}) \right).
$$

It follows from

$$
\begin{aligned}
p(\boldsymbol{\xi}_l | \mathbf{s}) &\propto p(\boldsymbol{\xi}_l) p(\mathbf{s} | \boldsymbol{\xi}_l) \\
&\propto \left(\prod_{k=1}^{L} \xi_{l,k}^{\bar{c}_k - 1} \right) \left(\prod_{g=1}^{L} \prod_{k=1}^{L} \xi_{g,k}^{N_{g,k}(\mathbf{s})} \right) \\
&\propto \left(\prod_{k=1}^{L} \xi_{l,k}^{\bar{c}_k - 1} \right) \left(\prod_{k=1}^{L} \xi_{l,k}^{N_{l,k}(\mathbf{s})} \right) \\
&= \prod_{k=1}^{L} \xi_{l,k}^{\bar{c}_k + N_{l,k}(\mathbf{s}) - 1}.
\end{aligned}
$$

We denoted the collection of all hidden states by $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_T)$ and defined with $N_{i,j}(\mathbf{s}) = \sum_t \zeta_{t-1,i} \zeta_{t,j}$ the number of transitions from state i to state j up to time T.

S.2.12 Full conditional distribution of s_t

For sampling the trajectory $\mathbf{s} = (s_1, \ldots, s_T)$, we update the whole path from the full joint conditional distribution in a multi-move Gibbs sampler step, also called the Forward Filtering Backward Sampling algorithm (FFBS, see Frühwirth-Schnatter (2006)). This consists in sampling the path from the joint full conditional distribution $p(\mathbf{s}|-)$. It is based on the factorisation of the full joint conditional distribution as the product of the entries of the transition matrix $\boldsymbol{\Xi}$ and the filtered probabilities. Starting from the complete data likelihood, we integrate the latent variables $(\mathcal{D}, \boldsymbol{\Omega})$ and sample the trajectory from the full joint conditional distribution

$$p(\mathbf{s}|\boldsymbol{\mathcal{X}}, \mathcal{G}, \boldsymbol{\rho}, \boldsymbol{\Xi}) \propto p(\boldsymbol{\mathcal{X}}|\mathbf{s}, \mathcal{G}, \boldsymbol{\rho}) p(\mathbf{s}|\boldsymbol{\Xi}).$$

Consequently, at each iteration of the Gibbs sampler we firstly compute the filtered state probabilities using $p(\boldsymbol{\mathcal{X}}|\mathbf{s}, \mathcal{G}, \boldsymbol{\rho})$ as likelihood function. Define $\mathcal{X}^{t-1} = \{\mathcal{X}_1, \ldots, \mathcal{X}_{t-1}\}$. The predictive probability correspond to the conditional distribution of the observations given the state

$$\begin{aligned} p(\mathcal{X}_t|s_t = l, \mathcal{X}^{t-1}, \mathcal{G}, \boldsymbol{\rho}) &= p(\mathcal{X}_t|s_t = l, \mathcal{G}, \boldsymbol{\rho}) \\ &= \prod_{i=1}^{I} \prod_{j=i}^{J} \prod_{k=1}^{K} p(x_{ijk,t}|s_t = l, \rho_l, \mathcal{G}_l). \end{aligned}$$

From eqq. (16)-(17)-(18), the logarithm of the predictive probability is

$$\log p(\mathcal{X}_t|s_t = l, \mathcal{X}^{t-1}, \mathcal{G}, \boldsymbol{\rho}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \log p(x_{ijk,t}|s_t = l, \rho_l, \mathcal{G}_l)$$

where

$$\begin{aligned} p(x_{ijk,t} = 1|s_t = l, \rho_l, \mathcal{G}_l) &= (1 - \rho_l) \frac{\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})} \\ p(x_{ijk,t} = 0|s_t = l, \rho_l, \mathcal{G}_l) &= \rho_l + (1 - \rho_l) \frac{1}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})}. \end{aligned}$$

S.3 Simulation Results

We consider a simulation experiment for the model in eq. (6) where $I = J = 30$ and $K = 1$, corresponding to a single-layer binary network with adjacency matrices $\mathbf{X}_t$. The simulated sample size is $T = 60$. We consider $Q = 3$ common covariates, generated from a stationary Gaussian VAR(1) process with entries of the coefficient matrix drawn independently from a truncated standard normal distribution. We assume $L = 2$ regimes and generate the trajectory of the Markov chain assuming a transition matrix with rows $\boldsymbol{\xi}_1 = (0.8, 0.2)$ and $\boldsymbol{\xi}_2 = (0.3, 0.7)$. We set the sparsity parameters to $\rho_1 = 0.7$ and $\rho_2 = 0.4$. For each regime $l = 1, 2$ and $h = 1, 2, 3$, we generate the PARAFAC marginals of the tensor $\mathcal{G}_l$

$$\boldsymbol{\gamma}_{h,1}^{(r)} \overset{iid}{\sim} \mathcal{N}_{n_h}(-0.4 \cdot \boldsymbol{\iota}_{n_h}, \mathbf{I}_{n_h}) \qquad \boldsymbol{\gamma}_{h,2}^{(r)} \overset{iid}{\sim} \mathcal{N}_{n_h}(-0.4 \cdot \boldsymbol{\iota}_{n_h}, \mathbf{I}_{n_h}),$$

where with set the rank $R = 5$. All simulations have been performed using MATLAB with the Tensor Toolbox v.2.6[4].

The left panel of Fig. 2 compares the true and estimated regimes, showing that the sampler is correctly classifying the hidden states, with few exceptions. The right panel reports the posterior distribution of the sparsity parameters: we correctly estimate the value in regime 1, and slightly underestimate the value in regime 2. The results for the coefficient tensor are reported in Fig. 3 and Fig. 4 for regime 1 and 2, respectively. Overall, there is a slight tendency to over-shrink to, which is a well known drawback of global-local hierarchical prior distributions (e.g., see Carvalho et al. (2010)). Finally, Fig. 5 shows the results for 6 randomly chosen entries of the estimated coefficient tensor, in the two regimes. We find evidence of good mixing and convergence of the chain. The true value is in a high posterior density interval, except in few cases in regime 1.

[4] Available at: `http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html`

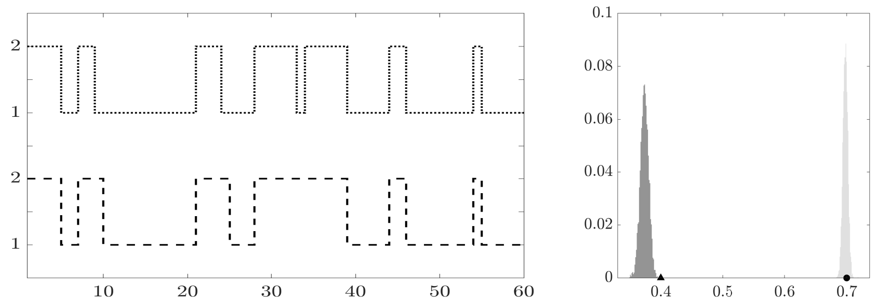


Figure 2: *Left:* true regimes (dashed line) and estimated regimes (dotted line). *Right:* posterior distribution of the sparsity parameters. Light (dark) grey denotes regimes 1 (regime 2). The black markers denote true values: the circle is ρ_1 and the triangle ρ_2.

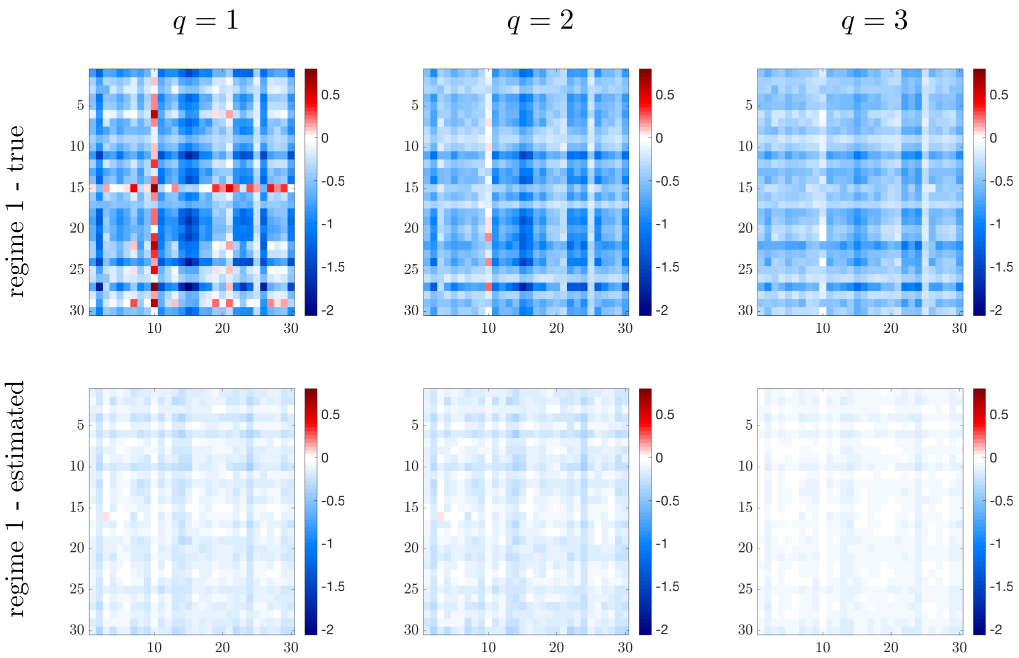


Figure 3: True and estimated coefficient tensors, in matricised form, in regime 1. In each plot, entry (i, j) represents the effect of the covariate reported in column on the probability of observing the edge between node i and node j. Same color scale, with red, blue and white colors indicating positive, negative and zero valued coefficients, respectively.

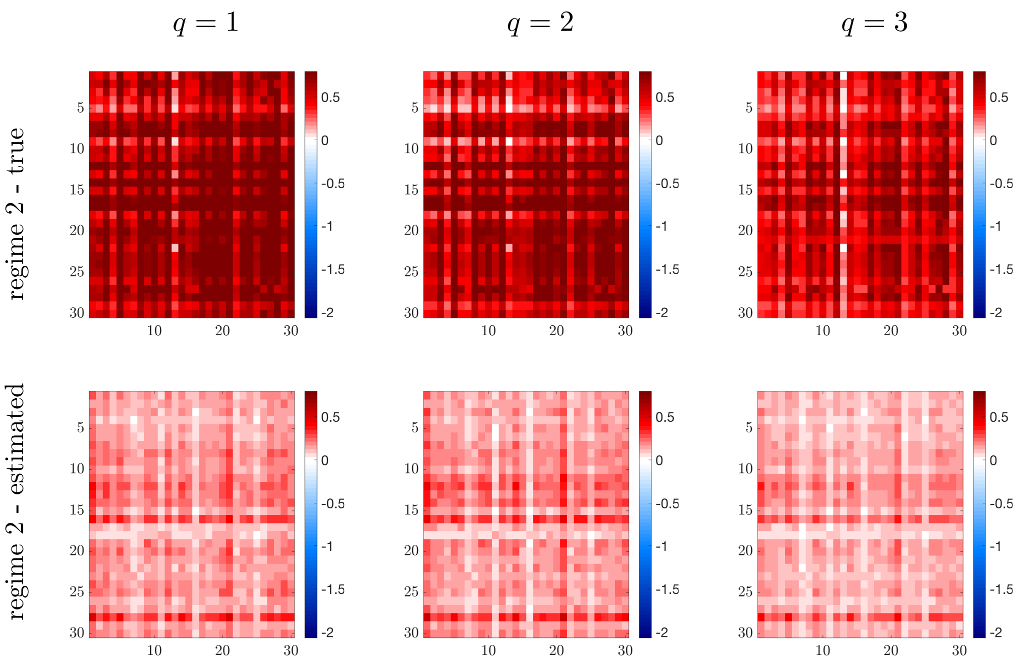


Figure 4: True and estimated coefficient tensors, in matricised form, in regime 2. In each plot, entry (i, j) represents the effect of the covariate reported in column on the probability of observing the edge between node i and node j. Same color scale, with red, blue and white colors indicating positive, negative and zero valued coefficients, respectively.

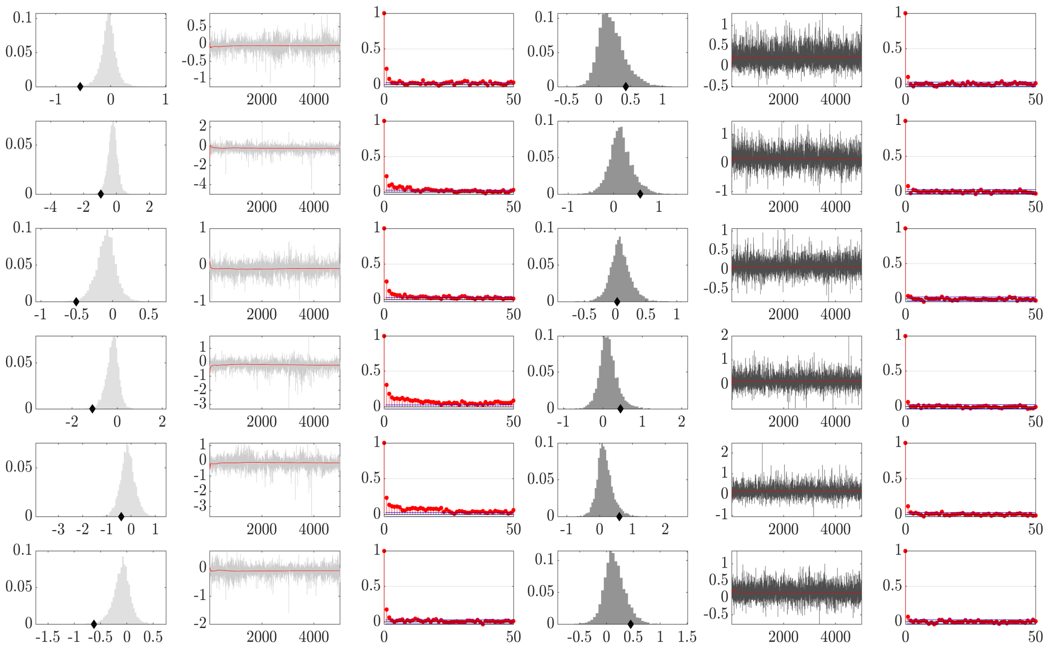


Figure 5: Plots for 6 randomly chosen entries of the estimated coefficient tensor. Light (dark) grey denotes regime 1 (regime 2). Posterior distribution (column 1 and 4), MCMC plot (columns 2 and 5) and autocorrelation function (columns 3 and 6). Black squares denote true values and the continuous line in the MCMC plots is the progressive mean.

S.4 Computational Details for the Pooled Model

S.4.1 Posterior distributions

In the pooled model case, we assume that the coefficient tensor in each regime $l = 1, \ldots, L$ has entries

$$\mathcal{G}_{ijkq,l} = \mathbf{g}_{q,l} \, ,$$

that is, $\mathbf{g}_{ijk,l} = \mathbf{g}_l \in \mathbb{R}^Q$ for each ijk and each regime l. Defining by $\mathcal{H}$ a tensor of size $(I \times J \times K \times Q \times Q)$ with entries

$$\mathcal{H}_{ijkqr} = \begin{cases} 1 & \text{if } q = r \\ 0 & \text{if } q \neq r \end{cases}$$

we have the representation

$$\mathcal{G}_l = \mathcal{H} \times_5 \mathbf{g}_l.$$

We assume the following prior distributions

$$\begin{aligned} \mathbf{g}_l | \tau, w_l &\sim \mathcal{N}_Q(\bar{\boldsymbol{\zeta}}_l, \tau w_l \mathbf{I}_Q) \\ p(\tau) &\sim \mathcal{G}a(\bar{a}^\tau, \bar{b}^\tau) \\ p(w_l|\lambda_l) &\sim \mathcal{E}xp(\lambda_l^2/2) \\ p(\lambda_l) &\sim \mathcal{G}a(\bar{a}_l^\lambda, \bar{b}_l^\lambda). \end{aligned}$$

We can rewrite the complete data likelihood as

$$\begin{aligned} L(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}) &\propto \prod_{t \in \mathcal{T}_l} \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \exp\Big(- \frac{\omega_{ijk,t}}{2} (\mathbf{z}_t' \mathbf{g}_l)^2 + \kappa_{ijk,t} (\mathbf{z}_t' \mathbf{g}_l) \Big) \\ &\propto \exp\Bigg(- \frac{1}{2} \sum_{t \in \mathcal{T}_l} \sum_{i,j,k} \mathbf{g}_l' \mathbf{z}_t \omega_{ijk,t} \mathbf{z}_t' \mathbf{g}_l - 2 \mathbf{g}_l' \mathbf{z}_t \kappa_{ijk,t} \Bigg). \end{aligned}$$

This yields the posterior distribution for $\mathbf{g}_l$

$$\begin{aligned} p(\mathbf{g}_l|\boldsymbol{\Omega}_t, \tau, w_l) &\propto \exp\Bigg(- \frac{1}{2} \sum_{t \in \mathcal{T}_l} \sum_{i,j,k} \mathbf{g}_l' \mathbf{z}_t \omega_{ijk,t} \mathbf{z}_t' \mathbf{g}_l - 2 \mathbf{z}_t' \mathbf{g}_l \kappa_{ijk,t} \Bigg) \exp\Bigg(- \frac{1}{2} \frac{(\mathbf{g}_l - \bar{\boldsymbol{\zeta}}_l)'(\mathbf{g}_l - \bar{\boldsymbol{\zeta}}_l)}{\tau w_l} \Bigg) \\ &= \exp\Bigg(- \frac{1}{2} \Bigg(\frac{(\mathbf{g}_l - \bar{\boldsymbol{\zeta}}_l)'(\mathbf{g}_l - \bar{\boldsymbol{\zeta}}_l)}{\tau w_l} + \sum_{t \in \mathcal{T}_l} \sum_{i,j,k} \mathbf{z}_t' \mathbf{g}_l \omega_{ijk,t} \mathbf{z}_t' \mathbf{g}_l - 2 \mathbf{g}_l' \mathbf{z}_t \kappa_{ijk,t} \Bigg) \Bigg) \\ &= \exp\Bigg(- \frac{1}{2} \Bigg(\mathbf{g}_l' \Bigg(\frac{1}{\tau w_l} + \sum_{t \in \mathcal{T}_l} \sum_{i,j,k} \mathbf{z}_t \omega_{ijk,t} \mathbf{z}_t' \Bigg) \mathbf{g}_l' - 2 \Bigg(\frac{\bar{\boldsymbol{\zeta}}_l'}{\tau w_l} + \sum_{t \in \mathcal{T}_l} \sum_{i,j,k} \kappa_{ijk,t} \mathbf{z}_t' \Bigg) \mathbf{g}_l \Bigg) \Bigg). \end{aligned}$$

Therefore, for each $l = 1, \ldots, L$

$$p(\mathbf{g}_l|\boldsymbol{\Omega}_t, \tau, w_l) \propto \mathcal{N}_Q(\mathbf{m}_l, \mathbf{S}_l),$$

with

$$\mathbf{S}_l = \left(\frac{1}{\tau w_l} + \sum_{t\in\mathcal{T}_l}\sum_{i,j,k} \mathbf{z}_t'\omega_{ijk,t}\mathbf{z}_t\right)^{-1}, \qquad \mathbf{m}_l = \mathbf{S}_l\left(\frac{\bar{\boldsymbol{\zeta}}_l'}{\tau w_l} + \sum_{t\in\mathcal{T}_l}\sum_{i,j,k} \mathbf{z}_t'\kappa_{ijk,t}\right).$$

The posterior distribution of τ is

$$\begin{aligned}
p(\tau|\mathbf{g},\mathbf{w}) &\propto p(\tau)p(\mathbf{g}|\mathbf{w},\tau) \\
&\propto \tau^{\bar{a}^\tau-1}\exp(-\bar{b}^\tau\tau)\prod_{l=1}^{L}\exp\Big(-\frac{\mathbf{g}_l'\mathbf{g}_l}{2\tau w_l}\Big) \\
&= \tau^{\bar{a}^\tau-1}\exp\Big(-\frac{1}{2}\Big(2\bar{b}^\tau\tau + \sum_{l=1}^{L}\frac{\mathbf{g}_l'\mathbf{g}_l}{w_l}\frac{1}{\tau}\Big)\Big) \\
&\propto \mathrm{GiG}\Big(\bar{a}^\tau - 1, 2\bar{b}^\tau, \sum_{l=1}^{L}\frac{\mathbf{g}_l'\mathbf{g}_l}{w_l}\Big).
\end{aligned}$$

The posterior distribution of w_l is

$$\begin{aligned}
p(w_l|g_l,\tau,\lambda_l) &\propto p(w_l|\lambda_l)p(g_l|w_l,\tau) \\
&\propto \exp\Big(-\frac{\lambda_l^2}{2}w_l\Big)\exp\Big(-\frac{\mathbf{g}_l'\mathbf{g}_l}{2\tau w_l}\Big) \\
&= \exp\Big(-\frac{1}{2}\Big(\lambda_l^2 w_l + \frac{\mathbf{g}_l'\mathbf{g}_l}{\tau}\frac{1}{w_l}\Big)\Big) \\
&\propto \mathrm{GiG}\Big(1, \lambda_l^2, \frac{\mathbf{g}_l'\mathbf{g}_l}{\tau}\Big).
\end{aligned}$$

The posterior distribution of λ_l (integrating out w_l) is

$$\begin{aligned}
p(\lambda_l|\tau,\mathbf{g}_l) &\propto p(\lambda_l)\int p(\mathbf{g}_l|\tau,w_l)p(w_l|\lambda_l)\,\mathrm{d}w_l \\
&\propto p(\lambda_l)p(\mathbf{g}_l|\tau,\lambda_l) \\
&\propto \lambda_l^{\bar{a}_l^\lambda-1}\exp(-\bar{b}_l^\lambda\lambda_l)\frac{\sqrt{\tau}}{2\lambda_l}\exp\Big(-\frac{\|\mathbf{g}_l\|_1\sqrt{\tau}}{\lambda_l}\Big) \\
&\propto \lambda_l^{\bar{a}_l^\lambda-2}\exp\Big(-\frac{1}{2}\Big(2\bar{b}_l^\lambda\lambda_l + \|\mathbf{g}_l\|_1\sqrt{\tau}\frac{1}{\lambda_l}\Big)\Big)
\end{aligned}$$

$$\propto \mathrm{GiG}\left(\bar{a}_l^{\lambda} - 1, 2\bar{b}_l^{\lambda}, \left\|\mathbf{g}_l\right\|_1 \sqrt{\tau}\right).$$

In Fig. 6 we report the posterior estimates of the coefficient tensor of the pooled model. We used the same dataset described in the application section of the main paper.

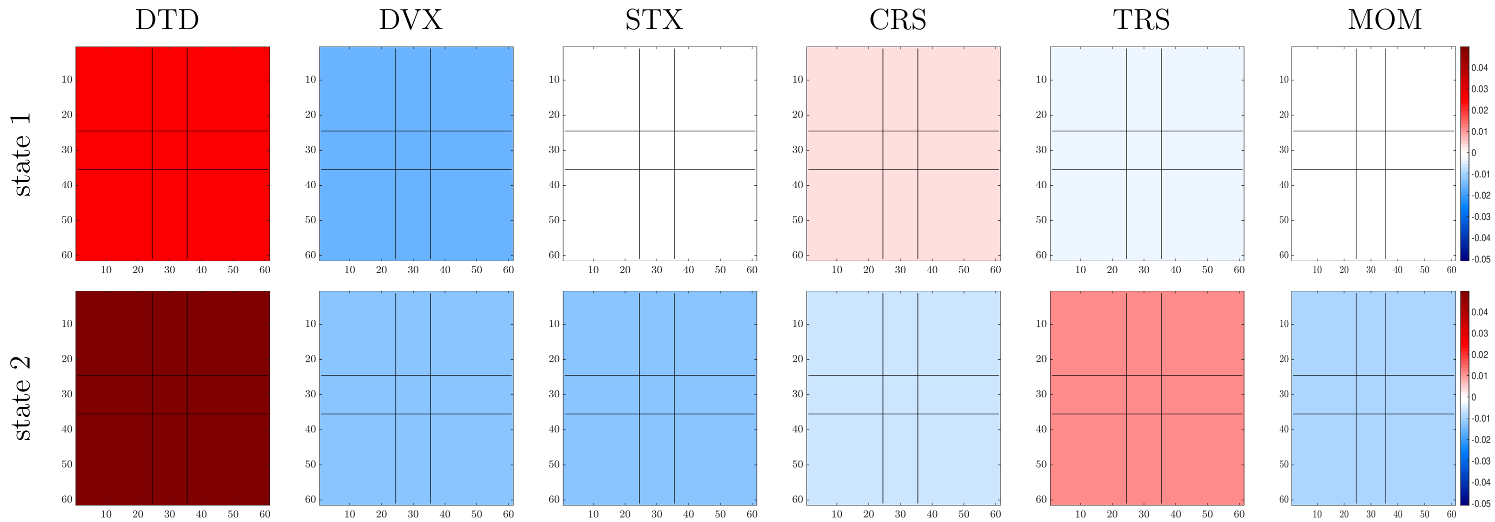


Figure 6: Pooled model. Posterior mean of the coefficient tensor, in matricised form, in the sparse (*top*) and dense (*bottom*) state of the hidden Markov chain. In each plot, entry (i, j) represents the effect of the covariate reported in column on the probability of observing the edge between institution i and institution j. Black lines separate groups of institutions: banks (i and j in $\{1, \ldots, 25\}$), insurance ($\{26, \ldots, 36\}$) and investment companies ($\{37, \ldots, 61\}$). Same color scale, with red, blue and white colors indicating positive, negative and zero valued coefficients, respectively.

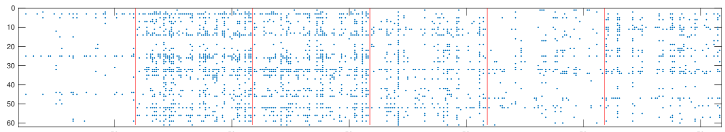
S.5 Data Description

We apply the proposed methodology to a financial network dataset estimated as in Billio et al. (2012), Bianchi et al. (2019). The dataset consists of binary, directed networks estimated via the Granger causality approach, where the nodes are European financial institutions. Other methods for extracting the network structure from data can be used, as this is not relevant for our inference framework, which applies to any sequence of binary tensors.

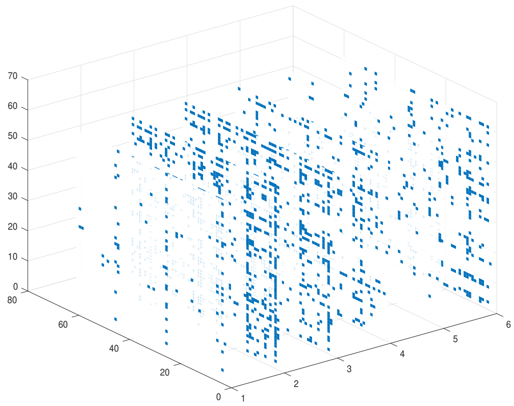
The original dataset is composed by the daily closing price series at a daily frequency from 29th December 1995 to 16th January 2013 of all the European financial institutions active and dead in order to cope with survivorship bias. It covers a total of 770 European financial firms which are traded in 10 European financial markets (core and peripheral). The pairwise Granger causalities are estimated on daily returns using a rolling window approach with a length for each window of 252 observations (approximately 1 year), controlling for heteroskedasticity and common macroeconomic factors. We obtain a total of 4197 adjacency matrices during the period from 8th January 1997 to 16th January 2013.

Then, we define a binary adjacency matrix for each month by setting an entry to 1 only if the corresponding Granger-causality link existed for the whole month (i.e. for each trading day of the corresponding month), and setting the entry to 0 otherwise. Since the panel is unbalanced due to entry and exit of financial institutions from the sample over time, we consider a subsample of length $T = 110$ months (from December 2003 to January 2013) made of 61 financial institutions.

We can visualize a sequence of adjacency matrices representing a time series of networks in several ways. subfig. 7(a) shows a stacked representation of a subsample composed by six adjacency matrices, while Fig. 7(b) plots a 3-dimensional array representation of the same data. In the first case, all matrices are stacked horizontally. Instead, the 3-dimensional representation plots each matrix in front of the other, as frontal slices of an array. It is possible to interpret the two plots as equivalent representations of a 3-order tensor: in this case, Fig. 7(a) shows the matricised form (along mode 1) of the tensor, while Fig. 7(b) plots its frontal slices. Finally, Fig. 8 plots the graph associated to two of these adjacency matrices. Though this representation allows for visualising the topology of a network, it is impractical for giving a compact representation of the whole time series of networks. Thus, we provide in Fig. 9 the stacked representation of the whole network sequence. Each row plots twelve time-consecutive adjacency matrices, starting from the top-left corner. The most striking features emerging from Fig. 9 are the time-varying degree distribution and the temporal clustering of sparse and dense networks.



(a) Stacked representation.



(b) 3D representation.

Figure 7: Stacked (a) and 3-dimensional (b) representations of a subsample of adjacency matrices (months $t = 65, 69, 73, 77, 81, 85$). Blue dots are existing edges, white dots are absent edges. A red line is used to separate each matrix (or tensor slice).

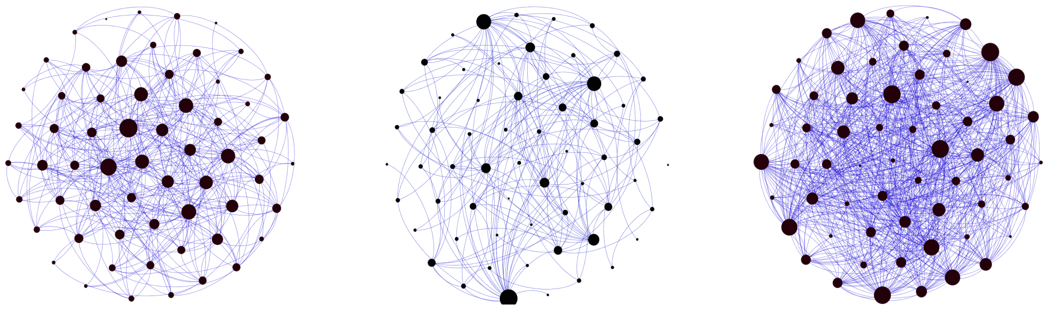


Figure 8: Graphical representation of networks at time $t = 25$ (Dec 2005), $t = 59$ (Oct 2008) and $t = 69$ (Aug 2009), respectively. Node size is proportional to its total degree. Edges are clockwise directed.

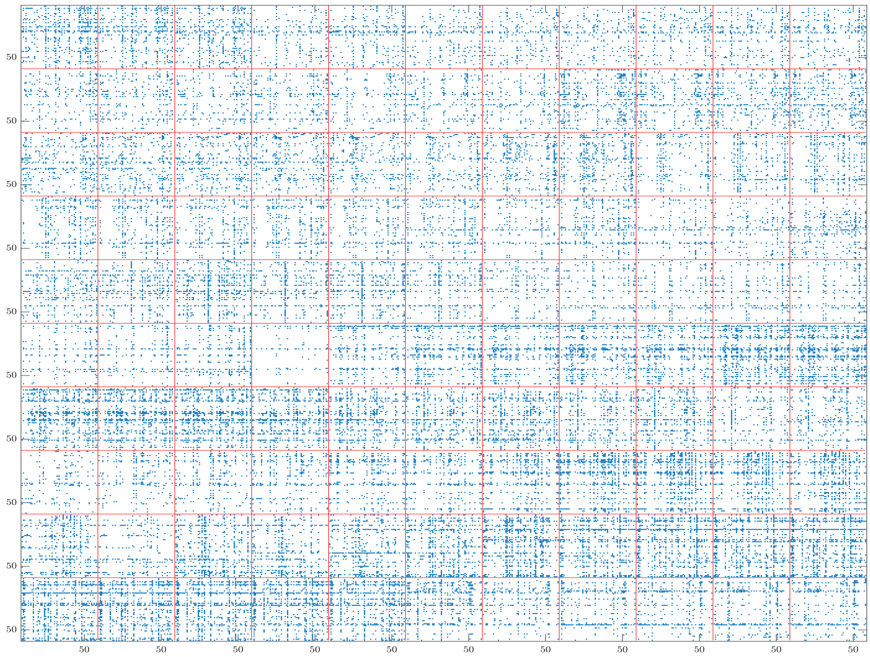


Figure 9: Full network dataset. In each red box there is an adjacency matrix, starting from top-left at time $t = 1$, the first row contains matrices from time $t = 1$ to $t = 11$, the second row from $t = 12$ to $t = 22$ and so on. Blue dots are existing edges, white dots are absent edges. Red lines are used to delimit the matrices.

The set of covariates $\mathbf{z}_t$ used to explain each edge's probability includes a constant term and:

- DTD: the lagged network total degree, defined as the total number of edges in the network at time $t-1$;
- DVX: the monthly change of the VSTOXX index, which is the volatility index for the STOXX50 (and may considered the counterpart of the VIX for Europe);
- STX: the monthly log-returns on the STOXX50 index, taken as a European equivalent to the US S&P500 index;
- CRS: the credit spread, defined as the difference between BAA and AAA indices provided by Moody's;
- TRS: the term spread, defined as the difference between the 10-year returns of reference Government bonds and the 6-months EURIBOR;
- MOM: the momentum factor.

The variables CRS, TRS are obtained using Datastream. VTOXX, STOXX50 are obtained from the ECB Data warehouse. The momentum factor is obtained from Ken French's website. All covariates have been standardised and included with one lag of delay, except DVX which is contemporaneous to the response, following the standard practice in volatility modelling (e.g., see, Corsi et al. (2013), Delpini and Bormetti (2015) Majewski et al. (2015)).

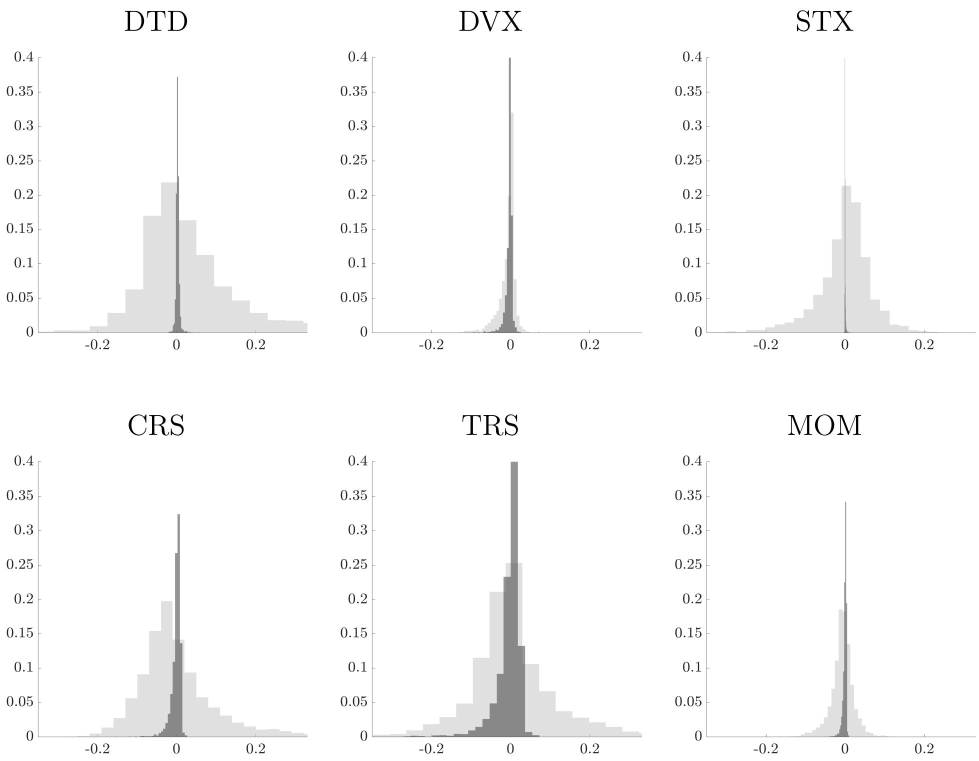


Figure 10: In all plots light (dark) grey identifies the dense (sparse) regime. Distribution of the entries of each slice (corresponding to each covariate) of the estimated coefficient tensor.

Fig. 10 provides evidence of different heterogeneity in the degree cross-sectional posterior distribution of the estimated coefficients. We find that in the sparse regime all distributions are narrowly concentrated around zero. The left skewness of credit and term spread coefficients accounts for the weakly negative effect of these covariates on the institutions connectivity in the sparse regime. Instead, in the dense regime there is evidence of larger impact of the network degree, credit spread and term spread on edge probability, as reflected by the fat tailed distribution of the corresponding coefficients.

Fig. 11 reports the effects of the credit spread on the network in the dense regime. We find that the credit spread has a strong, positive impact on the connections from banks and insurances to investment companies. There is evidence of higher between-groups connectivity of central institutions than within groups. The connections to the most central bank are positively affected by the credit spread, while linkages are negatively impacted. The linkages of the central insurance are the most affected by the credit spread, which has an average positive effect on the connections to the investment companies. There is no evidence of strong impact of the credit spread on the linkages coming from the most central investment company.

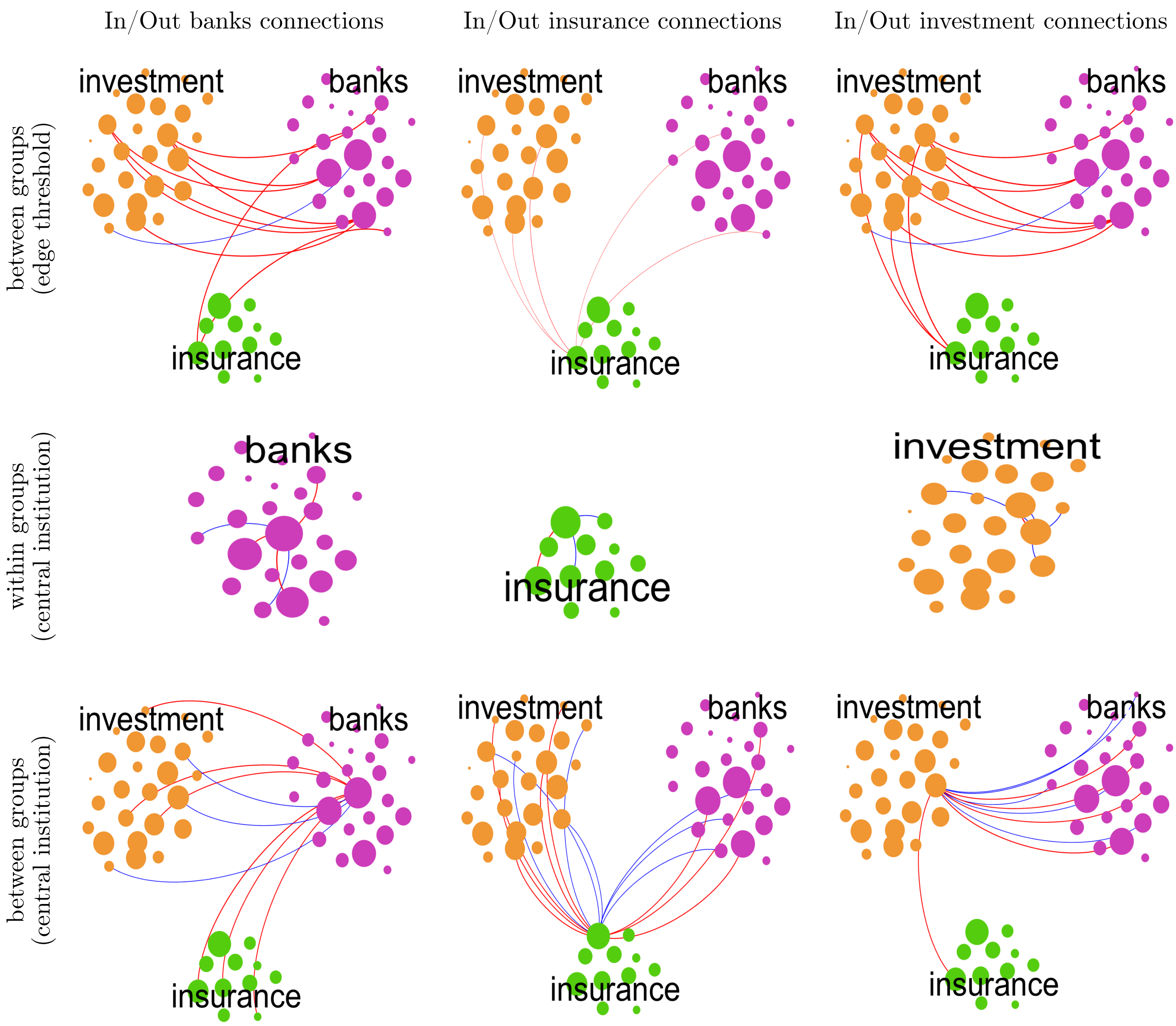


Figure 11: CRS coefficients in the dense regime. In the columns the effect of TRS on the edges from and to a specific group of nodes: bank (purple), insurance (green), investment companies (orange). In the rows the effects of CRS on between and within groups connectivity, filtering relevant effects (first row) and central institutions (second and third row). Node size: proportional to the total degree averaged over time within each regime. Edge color: blue for negative, red for positive. We show only edges with significant CRS coefficient.

References

Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2019). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics 2010*(1), 58–74.

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

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*(2), 465–480.

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

Cichocki, A., R. Zdunek, A. H. Phan, and S.-i. Amari (2009). *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation.* John Wiley & Sons.

Corsi, F., N. Fusari, and D. La Vecchia (2013). Realizing smiles: options pricing with realized volatility. *Journal of Financial Economics 107*(2), 284–304.

Delpini, D. and G. Bormetti (2015). Stochastic volatility with heterogeneous time scales. *Quantitative Finance 15*(10), 1597–1608.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models.* Springer.

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

Harris, M. N. and X. Zhao (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics 141*(2), 1073–1099.

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*.

Kiers, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chameometrics 14*(3), 105–122.

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

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.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), 1–14.

Majewski, A. A., G. Bormetti, and F. Corsi (2015). Smile from the past: a general option pricing framework with multiple volatility and leverage components. *Journal of Econometrics 187*(2), 521–531.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, J. L. Galin, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Chapter 5. Chapman & Hall /CRC.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association 108*(504), 1339–1349.