

VU Research Portal

Bayesian nonparametric graphical models for time-varying parameters VAR

Iacopini, Matteo; Rossini, Luca

2019

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Iacopini, M., & Rossini, L. (2019). *Bayesian nonparametric graphical models for time-varying parameters VAR*. (arXiv.org).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Bayesian nonparametric graphical models for time-varying parameters VAR*

Matteo Iacopini^{a†} Luca Rossini^{b‡}

^aScuola Normale Superiore of Pisa, Italy ^bVrije Universiteit Amsterdam, The Netherlands

6th June 2019

Abstract

Over the last decade, big data have poured into econometrics, demanding new statistical methods for analysing high-dimensional data and complex non-linear relationships. A common approach for addressing dimensionality issues relies on the use of static graphical structures for extracting the most significant dependence interrelationships between the variables of interest. Recently, Bayesian nonparametric techniques have become popular for modelling complex phenomena in a flexible and efficient manner, but only few attempts have been made in econometrics.

In this paper, we provide an innovative Bayesian nonparametric (BNP) time-varying graphical framework for making inference in high-dimensional time series. We include a Bayesian nonparametric dependent prior specification on the matrix of coefficients and the covariance matrix by mean of a Time-Series DPP as in Nieto-Barajas et al. (2012). Following Billio et al. (2019), our hierarchical prior overcomes over-parametrization and over-fitting issues by clustering the vector autoregressive (VAR) coefficients into groups and by shrinking the coefficients of each group toward a common location. Our BNP time-varying VAR model is based on a spike-and-slab construction coupled with dependent Dirichlet Process prior (DPP) and allows to: (i) infer time-varying Granger causality networks from time series; (ii) flexibly model and cluster non-zero time-varying coefficients; (iii) accommodate for potential non-linearities.

In order to assess the performance of the model, we study the merits of our approach by considering a well-known macroeconomic dataset. Moreover, we check the robustness of the method by comparing two alternative specifications, with Dirac and diffuse spike prior distributions.

Keywords: Bayesian Nonparametrics; Dependent Dirichlet process; Large vector autoregression; Sparsity; Time-Varying networks.

AMS 2000 subject classifications: Primary 62; secondary 91B84.

JEL Classification: C11, C32, C51, C53

*We are grateful to Deborah Gefang, Roberto León-Gonzalez and Sylvia Kaufmann for their comments and suggestions. Moreover, we thank the participants at: “4th Vienna Workshop on High-dimensional Time Series in Macroeconomics and Finance” in Wien, 2019. This research used the SCSCF multiprocessor cluster system at Ca’ Foscari University of Venice.

Luca Rossini acknowledges financial support from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 796902.

[†]e-mail: matteo.iacopini@sns.it

[‡]e-mail: l.rossini@vu.nl

1 Introduction

Over the last decade, the availability of large datasets in economics and finance has allowed the introduction of high dimensional models. In particular, large datasets in macroeconomics help to improve the forecasts, while in finance some authors have investigated the use of large datasets to analyse financial crises, contagion effects and their impact on the real economy. In order to deal with high dimensional models, the introduction of Bayesian nonparametric techniques have become popular in different fields (such as statistics and machine learning), but only few attempts have been made in econometrics. In particular, Bayesian nonparametric approach allows to improve the estimation efficiency and the prediction accuracy in time series analysis.

Recently, time-varying parameter (TVP) models provide an interesting alternative to process multiple change points; for example, time-varying structural vector autoregressive (VAR) models have been used in Primiceri (2005) for study monetary policy application; Dangel and Halling (2012) forecast equity returns by mean of TVP models; and in Belmonte et al. (2014) the European inflation has been studied via a time-varying parameters model. As shown in Primiceri (2005), Del Negro and Primiceri (2015) and Bitto and Frühwirth-Schnatter (2019), the advantage in capturing gradual changes is due to the flexibility of TVP models. We combine the ideas behind time-varying parameters models and Bayesian nonparametric techniques, thus allowing to model complex phenomena in a flexible and efficient manner. Moreover, we provide an innovative Bayesian nonparametric time-varying graphical framework for making inference in high-dimensional time series.

In this paper, we allow coefficients to be sparse, meaning that only a fraction of the time varying parameters have significant effects, but we retain flexibility in modelling non-zero coefficients, by including temporal dependence in the prior structure. In order to achieve these goals, we define a shrinkage prior on the VAR coefficients by means of a Bayesian nonparametric prior (BNP). This distribution is a spike-and-slab prior, where on the spike (parametric) distribution, we impose two different specifications: a Dirac spike and a “diffuse” spike. On the other hand, on the (non-parametric) slab distribution, we use a well know Bayesian nonparametric Lasso prior as in Billio et al. (2019).

The prior previously described groups the time-varying parameter vector autoregressive (TVP-VAR) coefficients into clusters and shrinks the coefficients within a cluster toward common notation. Differently from Markov-switching approach (Krolzig, 1997) and random walk processes (Primiceri, 2005; Del Negro and Primiceri, 2015), we impose time variation on the distribution of the VAR coefficients. In the literature of time-varying coefficients, the VAR coefficients are represented as a direct dependence, in practice they can be represented as state-space models, where they are functions of the previous time. On the other hand, we introduce a different structure, the indirect dependence on the VAR coefficients. In this case, we have a dependence construct through the atoms of the Dirichlet process and not on a direct way. Thus, we include a Bayesian nonparametric dependent prior specification on the VAR coefficients and the covariance matrix by means of a time-series dependent Dirichlet process (tsDDP) as in Nieto-Barajas et al. (2012).

Following Billio et al. (2019), our hierarchical prior overcomes overparametrization and overfitting issues by clustering the VAR coefficients into groups and by shrinking the coefficients of each group toward a common location. This hierarchical prior allows to contemporaneously estimate the (potentially) sparse time-varying causal network structure and to cluster the corresponding coefficients. In our BNP-TVP-VAR model, time-varying coefficients allow to (i) estimate the temporal networks of contemporaneous and causal structures, (ii) identify

different sources of time variation, from the size of shocks and/or the propagation mechanism, and (iii) accommodate for potential non-linearities.

We also contribute to the literature on financial and macroeconomic contagion (see Billio et al. (2012); Bianchi et al. (2019) and Barigozzi and Brownlees (2019)) through the lens of Granger causality and graphs/networks representation. Our BNP prior is particularly suited for studying Granger causality from time series and in particular it allows to estimate the most significant time-varying dependence interrelationships between the variables of interest. As explained above, we can extract time-varying graphs by using the posterior random partition induced by the non-parametric (slab) distribution, which allows to cluster the edges into groups.

1.1 Literature

Since their introduction in macroeconomics (see Sims (1980)), vector autoregressive (VAR) models have been extensively used in econometrics and time series statistics. Large VAR models have been used to analyse and forecast high-dimensional macroeconomic data (e.g., McCracken and Ng (2016)) and financial panels (e.g., Barigozzi and Brownlees (2019)). Moreover, in recent years VAR models have been used for studying financial and macroeconomic contagion (e.g., Cogley and Sargent (2005), Stock and Watson (2007), Diebold and Yilmaz (2012) and Bianchi et al. (2019)). Although, VAR models have been extensively used for assessing the impact and spread of external shocks (i.e., to perform impulse-response analysis), forecasting, estimating networks from Granger-causal relationships and to study systemic risk and financial contagion (e.g., Diebold and Yilmaz (2009), Billio et al. (2012) and Barigozzi and Brownlees (2019)).

Despite being a potentially very flexible statistical tools, the high number of parameters and the typical limited length of standard macroeconomic datasets make unrestricted inference daunting as the cross-sectional size increases. This has favoured the use of penalised regression and Bayesian methods for dealing with the problem of over-parametrisation. The general idea is to use informative priors to shrink the unrestricted model towards a more parsimonious setting, thereby reducing parameter uncertainty and improving forecast accuracy (see Karlsson (2013), Koop and Korobilis (2010) for a survey).

In the Bayesian VAR (a.k.a. BVAR) literature, a plethora of different prior distributions have been proposed to perform sparse estimation (e.g., see Giannone et al. (2014)). Starting from the well-known Minnesota prior (see Doan et al. (1984), Litterman (1986)), which specifies an objective prior on the coefficient and covariance matrices of a VAR, several parametric approaches have been developed exploiting hierarchical structures and finite mixtures (e.g., Kalli and Griffin (2014), Gefang (2014), Huber and Feldkircher (2019), Kastner and Huber (2018)).

Among the recent contributions for dealing with large dimensional models, we distinguish two approaches: the first attempts to reduce the size of the data to handle or to process during each step of the inferential algorithm, while the second is concerned with the reduction of the size of the parameter space. Within the first class, we mention the Bayesian compressed VAR of Koop et al. (2018), who tackled the dimensionality issue by using random projections to compress the data, and the Bayesian composite likelihood approach of Chan et al. (2018). On the other hand, Gefang et al. (2019) and Koop and Korobilis (2018) adopted a variational Bayes approach for performing efficient approximate posterior inference in large parameter spaces. Also, Kastner and Huber (2018) exploited factor models and hierarchical shrinkage priors for providing a parsimonious parametrisation of the covariance matrix which allows for equation-by-equation estimation. Additional contributions for estimating large VAR and

VARMA models include Koop and Korobilis (2013), Korobilis (2016) and Chan et al. (2016).

In addition to the large cross-sectional dimensionality, also the temporal length of many economic and financial datasets is steadily increasing. Thus, the possible relations between different variables of interest can be described by static matrix of coefficients. This assumption can be elapsd by introducing a time-variation of the matrix of coefficients of the time series. In particular, the most common approach consists in specifying a process governing the evolution of the parameters of interest. According to the force driving this dynamics, we distinguish observation-driven and parameter-driven time-varying parameter (TVP) models. The first class is mainly represented by generalised autoregressive score models (GAS, see Creal et al. (2013)), while the second one includes Markov switching (e.g., Hamilton (1989), Krolzig (1997)), change point (e.g. Pesaran et al. (2006)) and random walk models (e.g. Del Negro and Primiceri (2015), Primiceri (2005)). These processes are able to describe parameters whose evolution is subject to switching regimes, structural breaks or smooth changes, respectively.

In the Bayesian and frequentist literature, the use of parametric models has been widely studied by applying different shrinkage methods (such as the Least Absolute Shrinkage and Selection Operator, known as LASSO). In particular, important papers focus on sparse and efficient estimation in high-dimensional datasets. However, more recently increasing attention is being devoted to the issue of over-shrinkage and to the modelling of non-zero coefficients (e.g., Giannone et al. (2018)). Consequently, there is an increasing need for adequate statistical tools capable of flexibly model the dynamics described by a VAR process, allowing for sparsity without incurring into over-shrinking.

In this paper, we aim to contribute to the growing literature on the use of Bayesian nonparametrics in time series analysis. In particular, Bayesian nonparametric techniques are widely in statistics, machine learning and data analysis as powerful tools for flexible modelling of complex data structure. Only recently, Bayesian nonparametrics has increased popularity in econometrics and in economic time series modelling to capture observation clustering effects (see e.g. Bassetti et al. (2014), Kalli and Griffin (2018) and Billio et al. (2019)).

Up to our knowledge, our paper is the first to provide sparse Bayesian nonparametric VAR model when the coefficients are time-varying and the proposed two-stage prior specification can be easily extended to other classes, such as the seemingly unrelated regression (SUR) models. We propose a novel Bayesian nonparametric prior structure, which provides a sparse estimation of the coefficient matrix of a VAR model. This representation allows to manage the flexibility of non-zero entries and most importantly, to manage the time-variation in the matrix of coefficients through the atoms of the Dirichlet process and not through a state-space representation.

Our approach substantially differs from the existing literature in two aspects as described below. First, we consider a spike-and-slab prior distribution for each entry of the coefficient matrix, where on the spike we have a parametric prior specification by mean of Dirac or diffuse prior. On the hand, the slab component has random nonparametric prior. Second, we impose prior dependence on the coefficients by specifying a Markov process for their random distribution. As a by-product of the estimation procedure, we are able to extract a time series of dependent Granger-causality graphs. This shows how the BNP-TVP-VAR contributes to the literature on the estimation of time-varying networks from economic and/or financial series.

The paper is organized as follows. Section 2 introduces the modelling framework and presents the BNP-TVP prior structure, then Section 3 presents posterior approximation and describes how to extract Granger-causal time varying graphs from time series. Finally, Section 4 draws the conclusions.

2 A Bayesian Time-Varying VAR Model

2.1 TVP-VAR models

Let n be the number of units in a dataset and $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})$ a vector of n variables available at time t . A time-varying parameters vector autoregressive model of order p (TVP-VAR(p)) is defined as

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

where B_t is the $(n \times n)$ matrix of time-varying coefficients an $t = p, \dots, T$ is the time period. We assume that the error terms $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \dots, \epsilon_{n,t})'$ are i.i.d. for t with Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Eq. (1) can be written in the more compact form as

$$\mathbf{y}_t = \sum_{i=1}^p \mathbf{X}_{t-i+1} \boldsymbol{\beta}_{t-i+1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \quad (2)$$

where we define $\boldsymbol{\beta}_t = \text{vec}(\mathbf{B}_t)$; $\mathbf{X}_t = (\mathbf{y}'_{t-1} \otimes \mathbf{I}_n)$; \otimes is the Kronecker product and $\text{vec}(\cdot)$ the column-wise vectorization operator that stacks the columns of a matrix into a column vector.

2.2 Prior specification

Let us consider the problem of defining a flexible prior for a time-varying parameter model and we define the following TVP-VAR(p) with p equal to 1 as

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \quad (3)$$

In Eq. (3), the number of parameters of the n -dimensional TVP-VAR(1) model is $(T-1)n^2 + n(n+1)/2 = O(n^2)$, thus scales quadratically in n . In macroeconomic and financial applications, the number of variables of interest ranges from $n = 3$ (small size model) to $n = 20$ (large model) and even $n = 100$ (huge model). This highlights the twofold need for shrinkage estimation methods and in particular, for the introduction of sparse estimation of the coefficient matrix. In fact, it is very hard both to provide a meaningful interpretation for a large VAR with full time-varying matrix \mathbf{B}_t and to have an efficient and computationally feasible algorithm for an unrestricted estimation.

Motivated by this fact, we provide a prior distribution, which allows for sparse estimation in a time-varying parameter setting. For each coefficient of the matrix of parameters B_t , we introduce a mixture prior with independent location and scale parameters:

$$P(\boldsymbol{\beta}_t) = \prod_{j=1}^{n^2} P(\beta_{j,t}), \quad t = 2, \dots, T.$$

where $P(\beta_{j,t})$ is the probability distribution of the vector matrix of coefficients (for example, we can choose it as a Double Exponential or Laplace distribution). One of the most successful and widespread approach in the Bayesian literature consists in the use of (independent) spike-and-slab prior distributions (e.g., Mitchell and Beauchamp (1988), George and McCulloch (1993), Smith and Kohn (1996), George and McCulloch (1997)) for each coefficient $\beta_{j,t}$. Based on it, we specify a spike-and-slab prior distribution for each $\beta_{j,t}$, with $j = 1, \dots, n^2$ and $t = 2, \dots, T$,

of the form

$$\beta_{j,t} \sim \pi_t R(\beta_{j,t}) + (1 - \pi_t) Q(\beta_{j,t}), \quad (4)$$

where R, Q correspond to the spike and slab distributions, respectively, and π_t is the time-varying mixing probability (i.e., the prior probability of the spike component). In the literature, we have two commonly choices for D : a Dirac mass at 0 such that $R(\beta_{j,t}) = \delta_{\{0\}}(\beta_{j,t})$; and a centered (in zero) Normal distribution $R(\beta_{j,t}) = \mathcal{N}(\beta_{j,t}|0, \tau_0)$. The Dirac spike is a degenerate distribution that allows for variable selection as a by-product of the estimation. Instead, the choice of a continuous, diffuse prior (like a Gaussian) allows for shrinkage of the coefficients and is computationally faster, but requires the post-processing specification of threshold for the sake of variable selection.

The standard choice for the slab component Q is a heavy-tailed distribution belonging to the family of Generalised Hyperbolic distribution (e.g., Double Exponential, Cauchy, t -Student), since the aim of this component is to capture potentially large non-zero coefficients. In the case of Dirac spike, the prior for each coefficient, for $j = 1, \dots, n^2$ and $t = 2, \dots, T$, is given by

$$\beta_{j,t} | \mu_j, \lambda_j, \pi_t \sim \pi_t \delta_{(0)}(\beta_{j,t}) + (1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_j, \lambda_j), \quad (5)$$

while for a diffuse (Gaussian) spike we have

$$\beta_{j,t} | \mu_j, \lambda_j, \pi_t, \tau_0 \sim \pi_t \mathcal{N}(\beta_{j,t} | 0, \tau_0) + (1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_j, \lambda_j). \quad (6)$$

Example 1. In *Figure 1* we report an example of spike-and-slab prior, with centred Gaussian spike distribution (in blue) and centred double exponential slab distribution (in red). From the left panel, which shows the two distributions, we can see that the Gaussian accounts for most of the prior mass on 0 while the double exponential governs the tails. This is reflected in the plot on the right, which shows that the mixture distribution (with equal weights) has fatter tails than the Gaussian and more mass in a neighborhood of 0 than the double exponential.

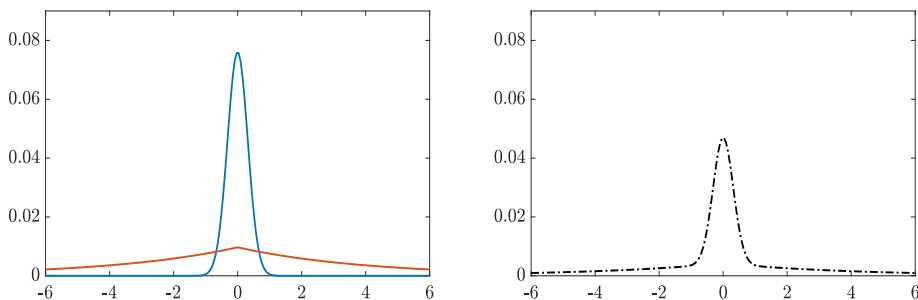


Figure 1: Example of spike-and-slab distribution as in eq. (6). *Left*: $\mathcal{N}(0, \sqrt{0.1})$ spike (blue) and $\mathcal{DE}(0, 4)$ slab (red) components. *Right*: mixture distribution, with mixing probability $\pi_t = 0.5$.

As previously described, we study the performance of the two different specification of the spike component by comparing the performances of the two constructions in extracting time-varying Granger causality networks from time series data.

The literature on time-varying parameter (TVP) models is vast. Some common parametric specifications include the threshold AR (TAR, e.g., Tong and Lim (1980)), smooth transition AR (STAR, e.g., Teräsvirta (1994)), along with their multivariate generalisations, Markov switching process (e.g., Hamilton (1989), Krolzig (1997)), change point process (e.g. Pesaran et al. (2006)) and random walk process (e.g. Del Negro and Primiceri (2015), Primiceri

(2005)). The choice of the particular specifications has been motivated by the intent to capture a particular feature of the dynamic evolution of the coefficients, such as changing regimes, structural breaks or smooth variations.

Differently from the existing literature on TVP-VAR models, we model the temporal dependence of the autoregressive parameters $\beta_{j,t}$ via assuming that the underlying prior (random) distributions evolve according to a discrete-time Markov process. A standard approach in Bayesian nonparametrics involves the specification of a Dirichlet Process (a.k.a. DP, see Ferguson (1973)) or a Dirichlet Process mixture (a.k.a. DPM, see e.g., Lo (1984)) prior for the distribution of the parameters of interest. The use of DP and related priors for a random probability measure P allows for clustering of the variables $x_i \stackrel{iid}{\sim} P$.

We proceed by introducing the prior temporal dependence between the random measures P_2, \dots, P_T via the time series Dirichlet Process (tsDDP) of Nieto-Barajas et al. (2012). As stated in the paper, for time series models, it is convenient to use dependence on the weights and common location. In practice, we apply a common discretization over the sequence of random measures, while the assumption of common weights and dependent location will lead to a discretization over the probability scale. In opposite to Taddy (2010), who was working with equally spaced time points, we accommodate for unequal time points. In our analysis a latent binomial process to induce the desired correlation has been used, differently from the stick-breaking random probability measures as in Taddy (2010), which use a beta autoregression on the fractions of the stick-breaking constructions by mean of two sets of latent variables.

By exploiting the stick-breaking construction of Sethuraman (1994), the time series Dirichlet Process imposes a dependence for the random probability measures

$$P_t(\cdot) = \sum_{i=1}^{\infty} w_{i,t} \delta_{(\theta_{i,t})}(\cdot), \quad (7)$$

where the locations are fixed $\theta_{i,t} = \theta_i$ and the weights $w_{i,t}$ vary over time. The dependence is described by a Markov process for each un-normalised stick-breaking weight $v_{i,t}$, with $i = 1, \dots, \infty$, via auxiliary variables $z_{i,t}$ (in the spirit of Pitt et al. (2002), Pitt and Walker (2005)), as follows

$$\begin{aligned} v_{i,1} &\sim \mathcal{Be}(1, \alpha) \\ z_{i,t} | v_{i,t} &\sim \mathcal{Bin}(m_{i,t}, v_{i,t}) \\ v_{i,t+1} | z_{i,t} &\sim \mathcal{Be}(1 + z_{i,t}, \alpha + m_{i,t} - z_{i,t}). \end{aligned} \quad (8)$$

The hyper-parameter $m_{i,t}$ tunes the strength of the dependence between P_t and P_{t+1} , such that $m_{i,t} = 0$ implies $P_t \perp P_{t+1}$, while $m_{i,t} \rightarrow \infty$ implies $P_t = P_{t+1}$ with probability 1 (see Nieto-Barajas et al. (2012)). Note that this construction implies that at each time $t = 2, \dots, T$, the marginal distribution of each random measure is a Dirichlet Process, that is

$$P_t \sim DP(\alpha, P_0).$$

with total mass parameter α and base measure P_0 , such that the base measure defines the expectation and the mass parameter is interpreted as the precision parameter.

Eq. (8) explains the joint distribution of $z_{i,t}$ and $v_{i,t}$ and it allows us to define the joint model for (P_1, \dots, P_T) as a *tsDDP*(α, P_0, \mathbf{m}), where \mathbf{m} is the sequence of the strength of dependence, $m_{i,t}$ for $i = 1, 2, \dots$ and $t = 1, \dots, T$. In Figure 2, we show a single draw of $(P_1, \dots, P_T) \sim \text{tsDDP}(\alpha, P_0, \mathbf{m})$, with total mass parameter α equal to 10; base measure P_0

as a normal distribution with zero mean and variance $\sqrt{2}$, i.e. $P_0 \sim \mathcal{N}(0, \sqrt{2})$ and total timing $T = 6$.

In order to assess the dependence structure induced by the time series Dirichlet Process, we consider the correlation between two random probability measures P_t and P_{t+1} . The following proposition is explaining this correlation:

Proposition 2.1 (Nieto-Barajas et al. (2012)).

Let $A \subset \mathbb{R}$ be measurable. For $(P_1, \dots, P_T) \sim tsDDP(\alpha, P_0, \mathbf{m})$ and any $t = 1, \dots, T - 1$ let $\rho_t(A) := \text{Corr}(P_t(A), P_{t+1}(A))$. Then

$$\rho_t(A) = (1 + \alpha) \sum_{h=1}^{\infty} a_{th} \prod_{i=1}^{h-1} b_{ti} + \frac{P_0(A)}{1 - P_0(A)} \left[\sum_{h=1}^{\infty} [2 - (1 + \alpha)a_{th}] \prod_{i=1}^{h-1} b_{ti} - (1 + \alpha) \right],$$

where

$$a_{ti} = \frac{2(1 + m_{ti}) + \alpha}{(1 + \alpha + m_{ti})(1 + \alpha)(2 + \alpha)}, \quad b_{ti} = \frac{\alpha - 1}{1 + \alpha} + a_{ti}$$

Remark 1. The correlation between (P_t, P_{t+1}) is larger in regions where the prior mean P_0 assigns more probability, meaning that the *tsDDP* prior places strongest dependence in P_0 -most probable regions. Note that strong dependence between (P_t, P_{t+1}) does not imply strong dependence between their outcomes $(\beta_{i,t}, \beta_{j,t+1})$.

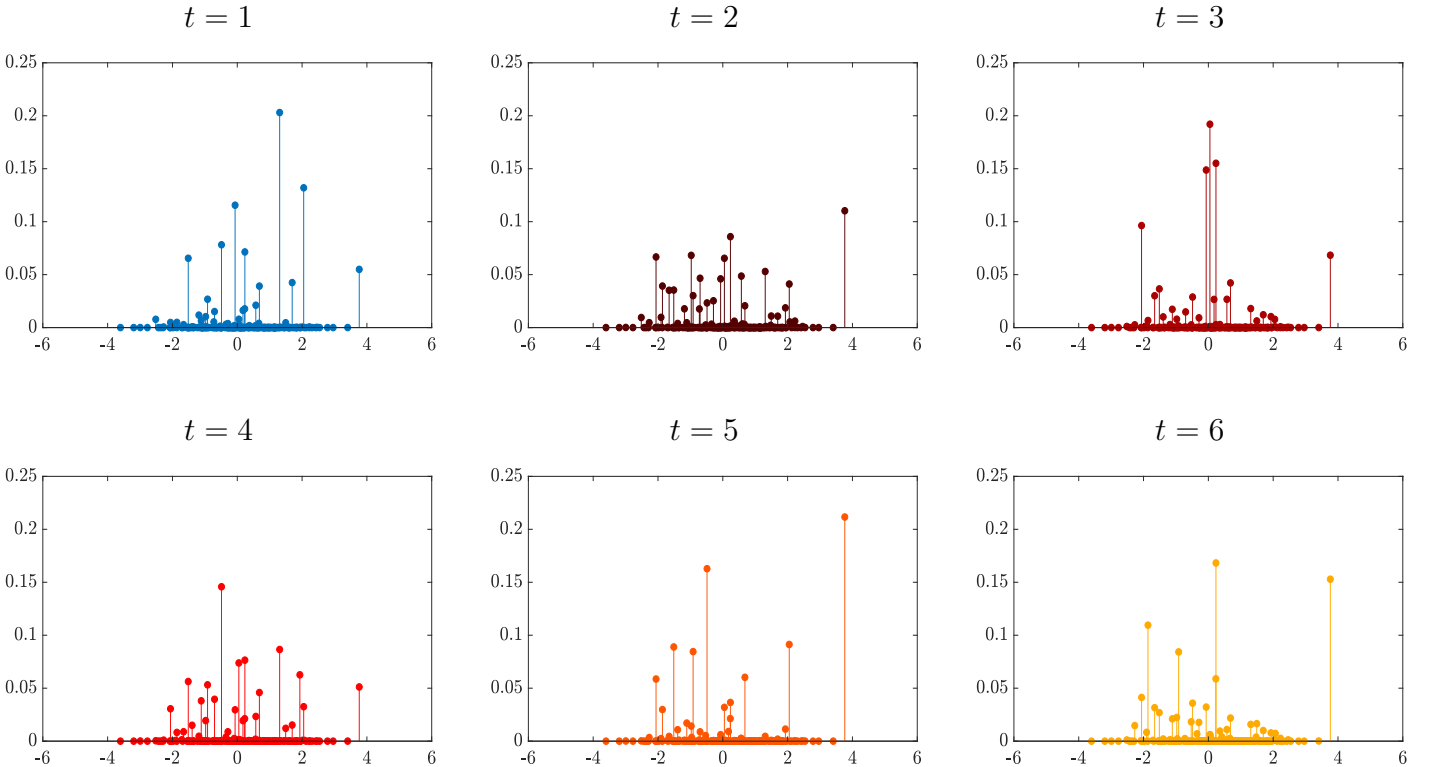


Figure 2: Sample from $tsDDP(\alpha, P_0, \mathbf{m})$, with base measure $P_0 \sim \mathcal{N}(0, \sqrt{2})$, concentration parameter $\alpha = 10$ and strength of dependence $\mathbf{m} = m_{i,t} = 4$. The size of the time series is assumed to be $T = 6$.

We can summarize what we have described above in the following prior structure¹, for

¹We use the shape-scale parametrisation of the Gamma distribution (thus $\mathbb{E}[x] = ab$ and $\mathbb{V}[x] = ab^2$) and

$j = 1, \dots, n^2$ and $t = 2, \dots, T$ as follows

$$\beta_{j,t} | \mu_j, \lambda_j, \pi_t \sim \pi_t R(\beta_{j,t}) + (1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_j, \lambda_j) \quad (9)$$

$$\lambda_j | \tau_j \sim \mathcal{Exp}(\lambda_j | 2/\tau_j) \quad (10)$$

$$(\mu_j, \tau_j) | P_t \sim P_t, \quad (11)$$

$$(P_2, \dots, P_T) \sim tsDDP(\alpha, P_0, \mathbf{m}) \quad (12)$$

$$P_0(\mu_j, \tau_j) \sim \mathcal{N}(\mu_j | c, d) \mathcal{Ga}(\tau_j | a_1, b_1) \quad (13)$$

$$\pi_t \sim \mathcal{Be}(1, \eta) \quad (14)$$

where $R(\beta_{j,t})$ is either a Dirac mass at 0 or a Normal distribution centered in zero and with variance τ_0 , which marginally has Inverse Gamma prior distribution $\tau_0 \sim \mathcal{IG}(\tau_0 | a_0, b_0)$. If we marginalize over λ_j , we have a Double Exponential slab distribution for each entry of the coefficient matrix. Following the notation of Eq. (4) we have $Q(\beta_{j,t}) = \mathcal{DE}(\beta_{j,t} | \mu_j, \tau_j)$, resulting in

$$\beta_{j,t} | \mu_j, \tau_j, \pi_t \sim \pi_t R(\beta_{j,t}) + (1 - \pi_t) \mathcal{DE}(\beta_{j,t} | \mu_j, \tau_j).$$

For the covariance matrix Σ , we assume an Inverse Wishart prior distribution that is:

$$\Sigma \sim \mathcal{IW}(\nu, \Psi) \quad (15)$$

where ν and Ψ are the degrees of freedom and scale hyperparameters, respectively.

In summary, the observational model in Eq. (2) together with the prior structure in Eqs. (9) to (14) lead to the BNP-TVP-VAR(1) model. Eq. (9)–(14) represent our hierarchical prior and in Figure 3 we represent them through a Directed Acyclic Graph (DAG) for the Normal spike specification. The observable and non-observable random variables are indicated through shadow and empty circles, respectively. On the left side we have the prior for Σ , while on the right side, we have the hierarchical prior for β_t , with a description of the first and second stage of the hierarchy by means of the shrinking parameters μ , τ and λ .

Sufficient conditions for stationarity of TVP autoregressive models are given for the univariate case (despite the proof is valid also in the multivariate setting) in Brandt (1986), while Bourgerol and Picard (1992) provides conditions for multivariate regressions where the coefficients are independent and identically distributed. The sufficient conditions given by Brandt (1986) is reported below.

Theorem 2.1 (Brandt (1986)).

Let $\{(\mathbf{B}_t, \boldsymbol{\epsilon}_t), t \in \mathbb{Z}\}$ be a strictly stationary ergodic process such that both $\mathbb{E}[\log^+(\|\mathbf{B}_0\|)]$ and $\mathbb{E}[\log^+(\|\boldsymbol{\epsilon}_0\|)]$ are finite. Suppose that the top Lyapunov exponent γ defined by

$$\gamma := \inf_{t \in \mathbb{N}} \mathbb{E} \left[\frac{1}{t+1} \log(\|\mathbf{B}_0 \mathbf{B}_{-1} \cdots \mathbf{B}_{-t}\|) \right]$$

Inverse Gamma distribution, whose probability density functions are, respectively

$$\begin{aligned} x \sim \mathcal{Ga}(x|a, b) &\iff p(x|a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}, & x \in (0, \infty), \\ x \sim \mathcal{IG}(x|a, b) &\iff p(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x}, & x \in (0, \infty). \end{aligned}$$

The exponential distribution is obtained as a particular case when $a = 1$, that is $\mathcal{Exp}(x|b) = \mathcal{Ga}(x|1, b)$.

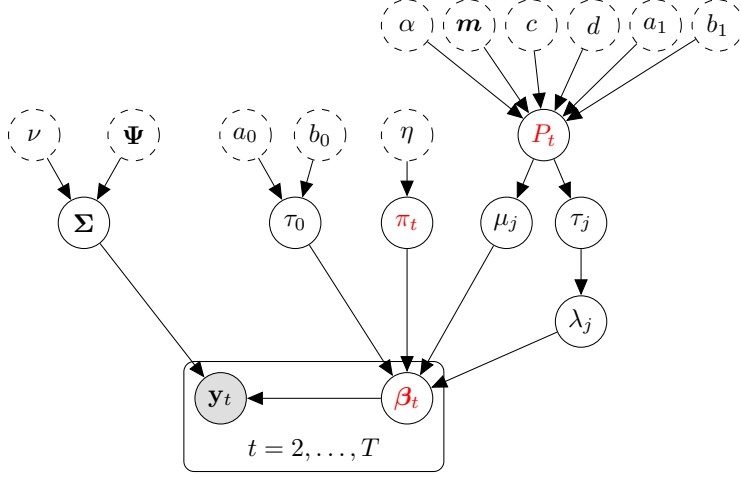


Figure 3: DAG of the BNP-TVP-VAR model, with Normal spike. It exhibits the hierarchical structure of priors and related hyperparameters. The directed arrows show the causal dependence relations of the model.

is strictly negative. Then, for all $t \in \mathbb{Z}$, the series

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{B}_0 \mathbf{B}_{t-1} \cdots \mathbf{B}_{t-i+1} \boldsymbol{\epsilon}_{t-i}$$

converges a.s., and the process $\{\mathbf{y}_t, t \in \mathbb{Z}\}$ is the unique strictly stationary solution of

$$\mathbf{y}_{t+1} = \mathbf{B}_{t+1} \mathbf{y}_t + \boldsymbol{\epsilon}_{t+1}, \quad t \in \mathbb{Z}.$$

2.3 Hyper-parameter elicitation

Following Nieto-Barajas et al. (2012), we assume $m_{i,t} = m$, for each $i = 1, 2, \dots$ and $t = 1, \dots, T$. Higher values of m strengthen the dependence between the un-normalised weights $v_{i,t}$, however when big m may induce the prior to overcome the likelihood, especially when the sample size is small. For this reason they specify a Poisson prior distribution for m , truncated on $\{1, \dots, 5\}$. Given the complexity of our prior specification, we prefer to fix the value of $m = 5$, which is sufficiently small to avoid overweighting of the prior² and then check the robustness of the results to alternative values of m . We choose the following values for the hyper-parameters:

$$\begin{aligned} c = 0 & \quad d = 4.0 & a_1 = 20.0 & b_1 = 0.1 & a_0 = 0.64 & b_0 = 1.25, \\ \alpha = 1.0 & \eta = 1.0 & \nu = n + 12 & \boldsymbol{\Psi} = \mathbf{I}_n/n. \end{aligned}$$

This choice amounts to assuming a uniform prior on each π_t and a rather uninformative prior on the covariance matrix, Σ . The value of the concentration parameter α is set according to standard practice in Dirichlet Process literature. The hyper-parameters (c, d, a_1, b_1) imply that for each new component of the Dirichlet Process the prior distribution of μ_j is centered at zero mean with medium-high variance, whereas the prior for τ_j has mean 2. Instead, the values of (a_0, b_0) imply that the prior variance of the (diffuse) spike distribution is 0.8, reflecting that this component should account for coefficients $\beta_{j,t}$ not significantly different from zero.

²In our empirical application, the sample size is $T = 248$, while Nieto-Barajas et al. (2012) have $T = 8$.

3 Posterior computation

3.1 Sampling method

Since the joint posterior distribution is not tractable and it is complex to be sample from, Bayesian estimator cannot be obtained analytically. In this paper, we rely on simulation based inference methods, and develop a Gibbs sampler algorithm for approximating the posterior distribution.

In order to deal with the finite mixture provided by the spike-and-slab prior and the infinite mixture given by the DPM, we exploited a data augmentation approach. For each $j = 1, \dots, n^2$ and $t = 2, \dots, T$, we introduce two sets of allocation variables $\gamma_{j,t}, d_{j,t}$; a set of stick-breaking variables, $\mathbf{v}_t = \{v_{i,t} : i = 1, 2, \dots\}$; a set of auxiliary variables $z_{i,t}$ (for $i = 1, 2, \dots$) and a set of slice variables, $\mathbf{u}_t = \{u_{j,t} : j = 1, \dots, n^2\}$. The allocation variables, $\gamma_{j,t}$, selects the spike component $R(\cdot)$, when $\gamma_{j,t}$ is equal to zero and the slab component, when it is equal to one. The second allocation variable, $d_{j,t}$, selects the component of the Dirichlet mixture to which each single coefficient $\beta_{j,t}$ is allocated to. The sequence of stick-breaking variables defines the mixture weights, whereas the slice variable, $u_{j,t}$, allows us to deal with the infinite mixture components by identifying a finite number of stick-breaking variables to be sampled and an upper bound for the allocation variables $d_{j,t}$.

Finally, we obtain the following joint posterior distribution

$$\begin{aligned}
 P(\mathbf{B}, \mathbf{U}, \mathbf{V}, \mathbf{Z}, \mathbf{D}, \Gamma, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \Sigma | \mathbf{Y}) &\propto L(\mathbf{Y} | \mathbf{B}, \Sigma) \cdot P(\Sigma) \cdot \prod_{k=1}^{k^*} P(\mu_k) P(\tau_k) \\
 &\cdot \prod_{t=2}^T P(\pi_t) \prod_{j=1}^{n^2} P(\beta_{j,t} | \pi_t, \lambda_j, \boldsymbol{\mu}) P(\lambda_j | \boldsymbol{\tau}) P(d_{j,t} | \mathbf{v}_t, \mathbf{u}_t) P(\gamma_{j,t} | \pi_t) \\
 &\cdot \prod_{i=1} P(u_{i,t} | v_{i,t}) P(v_{i,t} | z_{i,t-1}) P(z_{i,t-1} | v_{i,t-1}),
 \end{aligned} \tag{16}$$

where $\mathbf{U} = \{u_{j,t} : j = 1, \dots, n^2; \text{ and } t = 2, \dots, T\}$ and $\mathbf{V} = \{v_{i,t} : i = 1, 2, \dots \text{ and } t = 2, \dots, T\}$ are the collections of slice variables and stick-breaking components, respectively; $\mathbf{Z} = \{z_{i,t} : i = 1, 2, \dots \text{ and } t = 2, \dots, T\}$ and $\boldsymbol{\lambda} = \{\lambda_j : j = 1, \dots, n^2\}$ are the auxiliary and latent variables, respectively; $\mathbf{D} = \{d_{j,t} : j = 1, \dots, n^2; \text{ and } t = 2, \dots, T\}$ and $\Gamma = \{\gamma_{j,t} : j = 1, \dots, n^2; \text{ and } t = 2, \dots, T\}$ are the allocation variables; $(\boldsymbol{\mu}, \boldsymbol{\tau}) = \{(\mu_k, \tau_k) : k = 1, \dots, k^*\}$ are the atoms, where k ranges from 1 to the number k^* of allocated DP components; $\mathbf{B} = \{\boldsymbol{\beta}_t : t = 2, \dots, T\}$ is the vector of VAR coefficients and $\boldsymbol{\pi} = \{\pi_t : t = 2, \dots, T\}$ are the specific probabilities of shrinking coefficients to zero.

We obtain random samples from the posterior distributions by Gibbs sampling. The Gibbs sampler is based on the algorithm of Hatjispyros et al. (2011) and on the slice sampler approach of Walker (2007) and Kalli et al. (2011) for estimating the weights and locations of each random measure P_t . For improving the mixing of the MCMC, we introduced some Hamiltonian Monte Carlo (see Neal (2011)) steps in spite of drawing from the full conditional posterior distribution. Hereafter, we show the iterative steps by using the conditional independence between variables, for $k = 1, \dots, k^*$, $j = 1, \dots, n^2$, $i = 1, 2, \dots$ and $t = 2, \dots, T$:

- (1) the slice and stick-breaking variables $u_{j,t}$ and $v_{i,t}$ are updated along with the auxiliary variable $z_{i,t}$ given $[d_{j,t}, \gamma_{j,t}]$;
- (2) the latent scale variables λ_j are updated given $[\boldsymbol{\mu}, \boldsymbol{\tau}, (\beta_{j,t}, d_{j,t}, \gamma_{j,t})_t]$;

- (3) the parameters of the stick-breaking locations (μ_k, τ_k) are updated given $[\boldsymbol{\lambda}, \mathbf{B}, \mathbf{D}, \Gamma]$;
- (4) the allocation variables $d_{j,t}, \gamma_{j,t}$ are jointly updated given $[\mu_k, \tau_k, \beta_{j,t}, u_{j,t}, v_{i,t}, \pi_t]$;
- (5) the VAR coefficients $\boldsymbol{\beta}_t$ are jointly updated given $[\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \Sigma, (d_{j,t}, \gamma_{j,t})_j, \mathbf{y}_t]$;
- (6) The covariance matrix Σ is updated given $[\mathbf{B}, \mathbf{Y}]$;
- (7) the mixing probability π_t of having sparse coefficients is updated given $[(\gamma_{j,t})_j]$.

The detailed Gibbs sampler is described in [Appendix A](#) and [Appendix C](#).

3.2 Graph extraction

Based on the Gibbs sampler previously described, we are able to extract time-varying Granger-causal graphs. In the literature, linkages and networks describing the relationships between variables of interest, such as macroeconomics and financial linkages (e.g. Billio et al. (2012) and Barigozzi and Brownlees (2019)) can be used to extract pairwise Granger causality. This approach is generating spurious causality effects and does not consider conditioning on variables of interest. The main problem relies on the high number of variables available relative to the number of data, thus it could lead to overparametrization and inefficiency in gauging the causal relationships. Our proposed prior can be used to extract the networks and pairwise Granger causality while reducing the overfitting and curse of dimensionality problems. Moreover, the introduction of our prior could lead to the extraction of edge-colored graphs, that allows us to identify stylized facts in financial or macroeconomics networks and to show the presence of communities, hubs and linkage heterogeneity.

From the MCMC output of the time-varying coefficient matrix \mathbf{B}_t , we are able to extract time-varying Granger-causal graphs. At each time $t = 2, \dots, T$, we use the posterior random partition induced by the nonparametric (slab) distribution to cluster the edges of the graph (i.e., the entries $\beta_{j,t}$, $j = 1, \dots, n^2$ of the vectorised coefficient matrix $\boldsymbol{\beta}_t$) into groups.

Formally, a graph G is a pair (V, E) , where V is a set of nodes and E is a set of nodes pairs, named links or edges. The nodes are labeled and a link/edge is identified by the pair of nodes it connects, (i, j) . In particular, we have the existence of an edge if and only if the time-varying VAR coefficients of the variable $y_{i,t-1}$ in the equation of $y_{j,t}$ is not null. In our network analysis, we focus on the adjacency matrix constructed a posterior from the allocation variables and it allows to take both values between 0 and 1 if we apply a threshold, while if the values are allowed to vary between 0 and 1, we have a weighted graph. The purpose is to estimate the most significant time-varying dependence interrelationships (in terms of Granger-causality) between the n variables of interest.

Example 2. Consider the TVP-VAR(1) model in Eq. (2) and let $n = 4$. Without loss of generality, focus on the coefficient matrices at three consecutive times $t-1, t$ and $t+1$. Suppose the posterior estimates of the coefficient matrices and allocation variables $d_{j,t}$, respectively, are

as follows

$$\begin{aligned}
\mathbf{B}_{t-1} &= \begin{bmatrix} 0 & 0 & 0 & 0.8 \\ 0 & 0 & 0.8 & 0.2 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & -0.4 & 0 & 0 \end{bmatrix} & \mathbf{D}_{t-1} &= \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 1 \\ 2 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{bmatrix} \\
\mathbf{B}_t &= \begin{bmatrix} 0 & 0 & 0 & 0.8 \\ 0.2 & 0 & 0.8 & -0.4 \\ 0.2 & 0 & 0 & 0.8 \\ 0 & -0.4 & 0 & 0 \end{bmatrix} & \mathbf{D}_t &= \begin{bmatrix} 0 & 0 & 0 & 2 \\ 1 & 0 & 2 & 3 \\ 1 & 0 & 0 & 2 \\ 0 & 3 & 0 & 0 \end{bmatrix} \\
\mathbf{B}_{t+1} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0.8 & -0.4 \\ 0.8 & 0 & 0 & 0.8 \\ 0.2 & 0.2 & 0 & 0 \end{bmatrix} & \mathbf{D}_{t+1} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 3 \\ 2 & 0 & 0 & 2 \\ 1 & 1 & 0 & 0 \end{bmatrix}.
\end{aligned} \tag{17}$$

The corresponding Granger-causal graphs are given in [Figure 4](#), where colours have been used to denote the cluster assignment encoded in the matrices \mathbf{D}_{t-1} , \mathbf{D}_t and \mathbf{D}_{t+1} .

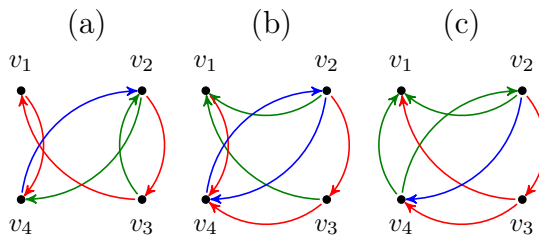


Figure 4: Weighted graphs with $\hat{K} = 3$ intensity levels: $\hat{\mu}_1^* = 0.2$ (green edges), $\hat{\mu}_2^* = 0.8$ (red edges) and $\hat{\mu}_3^* = -0.4$ (blue edges). In each graph the node v_i represents the variable i in the 4-dimensional VAR(1) in eq. (17), a clockwise-oriented edge from node j to node i represents a non-null coefficient for the variable $y_{j,t-1}$ in the i -th equation of the VAR. The vertex set is $V = \{v_1, v_2, v_3, v_4\}$ and the edges are $e_1 = \{v_1, v_4\}$, $e_2 = \{v_2, v_3\}$, $e_3 = \{v_2, v_4\}$, $e_4 = \{v_3, v_1\}$, $e_5 = \{v_3, v_2\}$, $e_6 = \{v_4, v_3\}$, $e_7 = \{v_2, v_1\}$, $e_8 = \{v_3, v_4\}$, $e_9 = \{v_4, v_1\}$. Panel (a): weighted graph $G_{t-1} = (V, E_{t-1})$ with $E_{t-1} = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ induced by edges of intensity level $\hat{\mu}_1^* = 0.2$. Panel (b): the weighted graph $G_t = (V, E_t)$ with $E_t = \{e_1, e_2, e_3, e_4, e_6, e_7, e_8\}$ induced by edges of intensity level $\hat{\mu}_2^* = 0.8$. Panel (c): the weighted graph $G_{t+1} = (V, E_{t+1})$ with $E_{t+1} = \{e_2, e_3, e_4, e_6, e_7, e_8, e_9\}$ induced by edges with intensity $\hat{\mu}_3^* = -0.4$.

4 Conclusions

We proposed the BNP-TVP-VAR model for sparse, nonparametric inference in time-varying VAR models. The use of spike-and-slab priors with time-series dependent Dirichlet Process prior for the slab component allows to contemporaneously shrink the autoregressive coefficients and flexibly modelling time-varying non-zero entries. We applied the proposed methodology using two alternative spike distributions: a Dirac and a Normal distribution. The performance of the resulting models has been compared in terms of: (i) sparse estimation and variable

selection, and (ii) clustering structure. Moreover, we showed how the BNP-TVP-VAR model can be used for extracting Granger-causal time-dependent graphs from multivariate time series.

References

- Barigozzi, M. and Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364.
- Bassetti, F., Casarin, R., and Leisen, F. (2014). Beta-product dependent Pitman–Yor processes for Bayesian inference. *Journal of Econometrics*, 180(1):49–72.
- Belmonte, M. A. G., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94.
- Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics*, 2010(1):58–74.
- Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *Journal of Econometrics*, Forthcoming.
- Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.
- Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97.
- Bourgerol, P. and Picard, N. (1992). Strict stationarity of generalised autoregressive processes. *Annals of Probability*, 20(1):1714–1730.
- Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability*, 18(1):211–220.
- Chan, J., Eisenstat, E., Hou, C., and Koop, G. (2018). Composite likelihood methods for large Bayesian VARs with stochastic volatility. *CAMA Working Paper*.
- Chan, J., Eisenstat, E., and Koop, G. (2016). Large Bayesian VARMA. *Journal of Econometrics*, 192(2):374–390.
- Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.
- Del Negro, M. and Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: A corrigendum. *Review of Economic Studies*, 82(1342–1345).
- Diebold, F. X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

- Diebold, F. X. and Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.
- Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for var shrinkage. *International Journal of Forecasting*, 30(1):1–11.
- Gefang, D., Koop, G., and Poon, A. (2019). Variational bayesian inference in large vector autoregressions with hierarchical shrinkage. *CAMA Working Paper*.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.
- Giannone, D., Lenza, M., and Primiceri, G. E. (2014). Prior selection for vector autoregressions. *The Review of Economics and Statistics*, 97(2):436–451.
- Giannone, D., Lenza, M., and Primiceri, G. E. (2018). Economic predictions with big data: The illusion of sparsity. *FRB of New York Staff Report No. 847*. Available at SSRN: <https://ssrn.com/abstract=3166281>.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.
- Hatjispyros, S. J., Nicolieris, T., and Walker, S. G. (2011). Dependent mixtures of dirichlet processes. *Computational Statistics & Data Analysis*, 55(6).
- Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, 37(1):27–39.
- Kalli, M. and Griffin, J. E. (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793.
- Kalli, M. and Griffin, J. E. (2018). Bayesian nonparametric vector autoregressive models. *Journal of Econometrics*, 203(2):267–282.
- Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and computing*, 21(1):93–105.
- Karlsson, S. (2013). Forecasting with bayesian vector autoregression. In *Handbook of economic forecasting*, volume 2, pages 791–897. Elsevier.
- Kastner, G. and Huber, F. (2018). Sparse Bayesian vector autoregressions in huge dimensions. *arXiv preprint arXiv:1704.03239*.
- Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358.

- Koop, G. and Korobilis, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198.
- Koop, G. and Korobilis, D. (2018). Variational bayes inference in high-dimensional time-varying parameter models. *arXiv preprint arXiv:1809.03031*.
- Koop, G., Korobilis, D., and Pettenuzzo, D. (2018). Bayesian compressed vector autoregressions. *Journal of Econometrics*.
- Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics & Data Analysis*, 101:110–120.
- Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer-Verlag Berlin Heidelberg.
- Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.
- Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics*, 12(1):351–357.
- McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Galin, J. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, chapter 5. Chapman & Hall /CRC.
- Nieto-Barajas, L. E., Muller, P., Ji, Y., Lu, Y., and Mills, G. B. (2012). A Time-Series DDP for Functional Proteomics Profiles. *Biometrics*, 68:859–868.
- Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4):1057–1084.
- Pitt, M. K., Chatfield, C., and Walker, S. G. (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29(4):657–663.
- Pitt, M. K. and Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, 100(470):554–564.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(821–852).
- Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, 4(2):639–650.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

- Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75(2):317–343.
- Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(1):3–33.
- Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association*, 105(492):1403–1417.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218.
- Tong, H. and Lim, K. (1980). Threshold autoregression, limit cycles and cyclical cata (with discussion of the paper). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 42(3):245–292.
- Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, 36(1):45–54.

A Posterior distributions: diffuse DE spike

A.1 Posterior for stick breaking un-normalised weights $v_{i,t}$

Posterior distribution for $v_{i,t}$, for all $t = 1, \dots, T$ and all $k = 1, \dots, k^*$, where k^* is the number of ties. We use the convention $z_{i,0} = m_{i,0} = 0$, for all $i = 1, 2, \dots$

$$\begin{aligned} P(v_{k,t}|z_{k,t}, z_{k,t-1}, \dots) &\propto P(v_{k,t}|z_{k,t-1})P(z_{k,t}|v_{k,t})P(\dots|v_{k,t}) \\ &\propto v_{k,t}^{z_{k,t}-1} (1 - v_{k,t})^{\alpha + m_{k,t-1} - z_{k,t-1} - 1} \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \left[\prod_{j:d_{j,t}=k, \gamma_{j,t}=1} v_{j,t} \right] \left[\prod_{j:d_{j,t}>k, \gamma_{j,t}=1} (1 - v_{j,t}) \right] \\ &\propto \mathcal{Be}(\bar{\xi}_t, \bar{\alpha}_t), \end{aligned}$$

where

$$\begin{aligned} \bar{\xi}_t &= 1 + z_{k,t} + z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} = k, \gamma_{j,t} = 1), \\ \bar{\alpha}_t &= \alpha + m_{k,t} - z_{k,t} + m_{k,t-1} - z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} > k, \gamma_{j,t} = 1). \end{aligned}$$

The normalised weights are then computed by $w_{k,t} = v_{k,t} \prod_{j < k} (1 - v_{j,t})$. The posterior distribution of the latent auxiliary variables $z_{k,t}$, for all $t = 1, \dots, T-1$ and all $k = 1, \dots, k^*$ is given by

$$\begin{aligned} P(z_{k,t}|v_{k,t+1}, v_{k,t}) &\propto P(z_{k,t}|v_{k,t})P(v_{k,t+1}|z_{k,t}) \\ &\propto \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \frac{\Gamma(\alpha + m_{k,t} + 1)}{\Gamma(z_{k,t} + 1)\Gamma(\alpha + m_{k,t} - z_{k,t})} v_{k,t+1}^{z_{k,t}} (1 - v_{k,t+1})^{\alpha + m_{k,t} - z_{k,t} - 1} \\ &\propto \frac{1}{(\Gamma(z_{k,t} + 1))^2 \Gamma(\alpha + m_{k,t} - z_{k,t}) \Gamma(m_{k,t} - z_{k,t} + 1)} \left(\frac{v_{k,t} v_{k,t+1}}{(1 - v_{k,t})(1 - v_{k,t+1})} \right)^{z_{k,t}}. \end{aligned}$$

Finally, the auxiliary variable for the slice sampler has posterior distribution given by

$$P(u_{j,t}|\dots) \propto \begin{cases} \mathbb{I}(u_{j,t} < w_{d_{j,t}})^{\gamma_{j,t}} & \text{if } \gamma_{j,t} = 1, \\ \mathbb{I}(u_{j,t} < 1)^{1-\gamma_{j,t}} & \text{if } \gamma_{j,t} = 0. \end{cases}$$

A.2 Posterior for λ_j

Posterior distribution for λ_j , for $j = 1, \dots, n^2$. Define $\mu_{k_j}^*, \tau_{k_j}^*$ are the location and scale of $\beta_{j,t}$, respectively, when sparse or non-sparse component is chosen.

$$P(\lambda_j|\dots) \propto \prod_{t=2}^T \lambda_j^{-(1-\gamma_{j,t})/2} \exp \left\{ -(1 - \gamma_{j,t}) \frac{1}{2\lambda_j} \beta_{j,t}^2 \right\} \lambda_j^{-\gamma_{j,t}/2} \exp \left\{ -\gamma_{j,t} \frac{1}{2\lambda_j} (\beta_{j,t} - \mu_{d_{j,t}})^2 \right\}$$

$$\begin{aligned}
& \cdot \exp \left\{ (1 - \gamma_{j,t}) \left(-\frac{\lambda_j \tau_0}{2} \right) \right\} \exp \left\{ \gamma_{j,t} \left(-\frac{\lambda_j \tau_{d_{j,t}}}{2} \right) \right\} \\
& \propto \lambda_j^{-\frac{T-1}{2}} \exp \left\{ -\frac{1}{2} \left[\lambda_j \left(\sum_{t=2}^T (1 - \gamma_{j,t}) \tau_0 + \gamma_{j,t} \tau_{d_{j,t}} \right) + \frac{1}{\lambda_j} \left(\sum_{t=2}^T (1 - \gamma_{j,t}) \beta_{j,t}^2 + \gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2 \right) \right] \right\} \\
& \propto GiG(p, a, b)
\end{aligned}$$

where

$$p = \frac{3-T}{2}, \quad a = \sum_{t=2}^T (1 - \gamma_{j,t}) \tau_0 + \gamma_{j,t} \tau_{d_{j,t}}, \quad b = \sum_{t=2}^T (1 - \gamma_{j,t}) \beta_{j,t}^2 + \gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2.$$

The probability density function of a generalised inverse Gaussian, for $p \in \mathbb{R}$, $a > 0$, $b > 0$ and $K_p(\cdot)$ being a modified Bessel function of the second kind, is

$$GiG(x|p, a, b) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-(ax+b/x)/2}, \quad x \in (0, \infty).$$

The vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{n^2})$ represents the diagonal of the diagonal matrix $\boldsymbol{\Lambda}$ of size $(n^2 \times n^2)$.

A.3 Posterior of the stick breaking locations (μ_k, τ_k)

The posterior distribution for the stick breaking locations in the sparse case are given by:

$$\mu_0 | \dots \sim \delta_{(0)}(\mu_0).$$

and

$$\begin{aligned}
P(\tau_0 | \dots) & \propto \mathcal{G}a(\tau_0 | a_0, b_0) \cdot \prod_{(j,t):\gamma_{j,t}=0} \frac{\tau_0}{2} \exp \left\{ -\frac{\tau_0}{2} \lambda_j \right\} \\
& \propto \tau_0^{a_0-1} \exp \left\{ -\frac{\tau_0}{b_0} \right\} \cdot \prod_{(j,t):\gamma_{j,t}=0} \frac{\tau_0}{2} \exp \left\{ -\frac{\tau_0}{2} \lambda_j \right\} \\
& \propto \tau_0^{a_0-1} \exp \left\{ -\frac{\tau_0}{b_0} \right\} \tau_0^{m_0} \exp \left\{ -\frac{\tau_0}{2} \sum_{(j,t):\gamma_{j,t}=0} \lambda_j \right\} \\
& \propto \tau_0^{a_0+m_0-1} \exp \left\{ -\tau_0 \left(\frac{1}{b_0} + \frac{1}{2} \sum_{(j,t):\gamma_{j,t}=0} \lambda_j \right) \right\} \\
& \propto \mathcal{G}a(\bar{a}_0, \bar{b}_0),
\end{aligned}$$

with

$$\bar{a}_0 = a_0 + m_0, \quad \bar{b}_0 = \frac{2b_0}{2 + b_0 \sum_{(j,t):\gamma_{j,t}=0} \lambda_j},$$

where $m_0 = \sum_{(j,t)} \mathbb{I}(\gamma_{j,t} = 0) = \sum_{j=1}^{n^2} \sum_{t=1}^T (1 - \gamma_{j,t})$ and we used the parametrisation of the Gamma distribution with shape $a > 0$ and scale $b > 0$, that is

$$\mathcal{G}a(x|a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}, \quad x \in (0, \infty).$$

Regarding the sample in the non-sparse case, we generate (μ_k, τ_k) for $k = 1, \dots, N^*$ and we have the following full conditional separately for μ_k and τ_k . The posterior distribution for the stick breaking locations in the non-sparse case, is given by:

$$\begin{aligned} P(\mu_k | \dots) &\propto \mathcal{N}(\mu_k | c, d) \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \exp \left\{ -\frac{1}{2\lambda_j} (\beta_{j,t} - \mu_k)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2d} (\mu_k^2 - 2c\mu_k) \right\} \exp \left\{ -\frac{1}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_j^{-1} - 2\mu_k \lambda_j^{-1} \beta_{j,t} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\mu_k^2 d^{-1} - 2c\mu_k d^{-1} + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_j^{-1} - 2\mu_k \lambda_j^{-1} \beta_{j,t} \right] \right\} \\ &\propto \mathcal{N}(\bar{c}, \bar{d}), \end{aligned}$$

where

$$\bar{d} = \left(d^{-1} + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \right)^{-1}, \quad \bar{c} = \bar{d} \left(d^{-1} c + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \beta_{j,t} \right).$$

and for the scale parameter we have the following representation:

$$\begin{aligned} P(\tau_k | \dots) &\propto \mathcal{G}a(\tau_k | a_1, b_1) \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\ &\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\ &\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \tau_k^{m_1} \exp \left\{ -\frac{\tau_k}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right\} \\ &\propto \tau_k^{a_1+m_1-1} \exp \left\{ -\tau_k \left(\frac{1}{b_1} + \frac{1}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right) \right\} \\ &\propto \mathcal{G}a(\bar{a}_1, \bar{b}_1), \end{aligned}$$

with

$$\bar{a}_1 = a_1 + m_1, \quad \bar{b}_1 = \frac{2b_1}{2 + b_1 \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j},$$

where $m_1 = \sum_{(j,t)} \mathbb{I}(\gamma_{j,t} = 1) \mathbb{I}(d_{j,t} = k) = \sum_{j=1}^{n^2} \sum_{t=1}^T \gamma_{j,t} \mathbb{I}(d_{j,t} = k)$.

A.4 Posterior of the allocation variables $(d_{j,t}, \gamma_{j,t})$

Regarding the allocation variables $d_{j,t}, \gamma_{j,t}$, we obtain the following full conditional for the sparse and non-sparse case. Let $u_{j,t} \sim \mathcal{U}([0, 1])$ for all $j = 1, \dots, n^2$ and all $t = 1, \dots, T$, be the auxiliary slice sampling variable. In the non-sparse case, we have:

$$\begin{aligned} P(d_{j,t} = k, \gamma_{j,t} = 1 | \dots) &\propto (1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_k, \lambda_j) \mathcal{E}xp \left(\lambda_j \left| \frac{2}{\tau_k} \right. \right) \mathbb{I}(u_{j,t} < w_k) \\ &\propto \frac{(1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_k, \lambda_j) \mathcal{E}xp \left(\lambda_j \left| \frac{2}{\tau_k} \right. \right)}{\sum_{i \in A_w(u_{j,t})} \mathcal{N}(\beta_{j,t} | \mu_i, \lambda_j) \mathcal{E}xp \left(\lambda_j \left| \frac{2}{\tau_i} \right. \right)}, \quad \forall k \in A_w(u_{j,t}), \end{aligned}$$

where $A_w(u_{j,t}) = \{k \in \{1, 2, \dots, k^*\} : w_k > u_{j,t}\}$, where k^* is the number of ties. On the other hand, in the sparse case we have the non-normalised posterior probability

$$P(d_{j,t} = 0, \gamma_{j,t} = 0 | \dots) \propto \pi_t \mathcal{N}(\beta_{j,t} | 0, \lambda_j) \mathcal{E}xp \left(\lambda_j \left| \frac{2}{\tau_0} \right. \right) \mathbb{I}(u_{j,t} < 1),$$

and full conditional distribution has the following representation:

$$P(d_{j,t} = k, \gamma_{j,t} = 0 | \dots) \propto \begin{cases} \pi_t \mathcal{N}(\beta_{j,t} | 0, \lambda_j) \mathcal{E}xp \left(\lambda_j \left| \frac{2}{\tau_0} \right. \right) & \text{if } k = 0, \\ 0 & \text{if } k > 0. \end{cases}$$

A.5 Posterior for $\beta_{j,t}$

Posterior distribution for β_t , for all $t = 2, \dots, T$. Let $\boldsymbol{\mu}_t^* = (\mu_{d_{1,t}}, \dots, \mu_{d_{n^2,t}})'$ and $\boldsymbol{\gamma}_t = (\gamma_{1,t}, \dots, \gamma_{n^2,t})'$, and denote the Hadamard product by \odot . We have:

$$\begin{aligned} P(\boldsymbol{\beta}_t | \dots) &\propto (2\pi)^{-n^2/2} |\boldsymbol{\Lambda}|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_t - \boldsymbol{\mu}_t^*)' \boldsymbol{\Lambda}^{-1} (\boldsymbol{\beta}_t - \boldsymbol{\mu}_t^*) \right\} \\ &\quad \cdot (2\pi)^{-n^2/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_t' \boldsymbol{\Lambda}^{-1} \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu}_t^*) \right\} \cdot \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{X}_t \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{y}_t) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^{n^2} \frac{1}{\lambda_j} \left(\beta_{j,t}^2 - 2 \beta_{j,t} \mu_{d_{j,t}} \gamma_{j,t} \right) + (\boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{X}_t \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{y}_t) \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[(\boldsymbol{\beta}_t' \boldsymbol{\Lambda}^{-1} \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu}_t^* \odot \boldsymbol{\gamma}_t)) + \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{X}_t \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t \right] \right\} \\ &\propto \mathcal{N}(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t), \end{aligned}$$

where

$$\bar{\boldsymbol{\Sigma}}_t = (\boldsymbol{\Lambda}^{-1} + \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{X}_t)^{-1}, \quad \bar{\boldsymbol{\mu}}_t = \bar{\boldsymbol{\Sigma}}_t (\boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu}_t^* \odot \boldsymbol{\gamma}_t) + \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t).$$

A.6 Posterior for covariance matrix Σ

Posterior distribution for the covariance matrix Σ .

$$\begin{aligned}
P(\Sigma|\dots) &\propto \mathcal{IW}(\nu, \Psi) \cdot \prod_{t=1}^T |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \Sigma^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t) \right\} \\
&\propto |\Sigma|^{-(\nu+n+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} (\Psi \Sigma^{-1}) \right\} |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \text{tr} ((\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \Sigma^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)) \right\} \\
&\propto \mathcal{IW}(\bar{\nu}, \bar{\Psi}),
\end{aligned}$$

where

$$\bar{\nu} = \nu + T, \quad \bar{\Psi} = \Psi + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)'.$$

A.7 Posterior for mixing probability π_t

Posterior distribution for the mixing probability π_t , for all $t = 1, \dots, T$.

$$\begin{aligned}
P(\pi_t|\dots) &\propto \mathcal{Be}(1, \eta) \cdot \prod_{j=1}^{n^2} \pi_t^{\mathbb{I}(\gamma_{j,t}=0)} (1 - \pi_t)^{\mathbb{I}(\gamma_{j,t}=1)} \\
&\propto (1 - \pi_t)^{\eta-1} \pi_t^{n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} (1 - \pi_t)^{\sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} \\
&\propto \mathcal{Be}(\bar{\phi}, \bar{\eta}),
\end{aligned}$$

where

$$\bar{\phi} = 1 + n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1), \quad \bar{\eta} = \eta + \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1).$$

B Posterior distributions: diffuse Normal spike

B.1 Posterior for stick breaking un-normalised weights $v_{i,t}$

Posterior distribution for $v_{i,t}$, for all $t = 1, \dots, T$ and all $k = 1, \dots, k^*$, where k^* is the number of ties. We use the convention $z_{i,0} = m_{i,0} = 0$, for all $i = 1, 2, \dots$

$$\begin{aligned} P(v_{k,t}|z_{k,t}, z_{k,t-1}, \dots) &\propto P(v_{k,t}|z_{k,t-1})P(z_{k,t}|v_{k,t})P(\dots|v_{k,t}) \\ &\propto v_{k,t}^{z_{k,t}-1} (1 - v_{k,t})^{\alpha + m_{k,t-1} - z_{k,t-1} - 1} \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \left[\prod_{j:d_{j,t}=k, \gamma_{j,t}=1} v_{j,t} \right] \left[\prod_{j:d_{j,t}>k, \gamma_{j,t}=1} (1 - v_{j,t}) \right] \\ &\propto \mathcal{Be}(\bar{\xi}_t, \bar{\alpha}_t), \end{aligned}$$

where

$$\begin{aligned} \bar{\xi}_t &= 1 + z_{k,t} + z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} = k, \gamma_{j,t} = 1), \\ \bar{\alpha}_t &= \alpha + m_{k,t} - z_{k,t} + m_{k,t-1} - z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} > k, \gamma_{j,t} = 1). \end{aligned}$$

The normalised weights are then computed by $w_{k,t} = v_{k,t} \prod_{j < k} (1 - v_{j,t})$. The posterior distribution of the latent auxiliary variables $z_{k,t}$, for all $t = 1, \dots, T-1$ and all $k = 1, \dots, k^*$ is given by

$$\begin{aligned} P(z_{k,t}|v_{k,t+1}, v_{k,t}) &\propto P(z_{k,t}|v_{k,t})P(v_{k,t+1}|z_{k,t}) \\ &\propto \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \frac{\Gamma(\alpha + m_{k,t} + 1)}{\Gamma(z_{k,t} + 1)\Gamma(\alpha + m_{k,t} - z_{k,t})} v_{k,t+1}^{z_{k,t}} (1 - v_{k,t+1})^{\alpha + m_{k,t} - z_{k,t} - 1} \\ &\propto \frac{1}{(\Gamma(z_{k,t} + 1))^2 \Gamma(\alpha + m_{k,t} - z_{k,t}) \Gamma(m_{k,t} - z_{k,t} + 1)} \left(\frac{v_{k,t} v_{k,t+1}}{(1 - v_{k,t})(1 - v_{k,t+1})} \right)^{z_{k,t}}. \end{aligned}$$

Finally, the auxiliary variable for the slice sampler has posterior distribution given by

$$P(u_{j,t}|\dots) \propto \begin{cases} \mathbb{I}(u_{j,t} < w_{d_{j,t}})^{\gamma_{j,t}} & \text{if } \gamma_{j,t} = 1, \\ \mathbb{I}(u_{j,t} < 1)^{1-\gamma_{j,t}} & \text{if } \gamma_{j,t} = 0. \end{cases}$$

B.2 Posterior for λ_j

Posterior distribution for λ_j , for $j = 1, \dots, n^2$. Define $\mu_{k_j}^*, \tau_{k_j}^*$ are the location and scale of $\beta_{j,t}$, respectively, when sparse or non-sparse component is chosen.

$$P(\lambda_j|\dots) \propto \prod_{t=2}^T \lambda_j^{-\gamma_{j,t}/2} \exp \left\{ -\gamma_{j,t} \frac{1}{2\lambda_j} (\beta_{j,t} - \mu_{d_{j,t}})^2 \right\} \exp \left\{ \gamma_{j,t} \left(-\frac{\lambda_j \tau_{d_{j,t}}}{2} \right) \right\}$$

$$\begin{aligned} &\propto \lambda_j^{-\frac{1}{2} \sum_{t=2}^T \gamma_{j,t}} \exp \left\{ -\frac{1}{2} \left[\lambda_j \left(\sum_{t=2}^T \gamma_{j,t} \tau_{d_{j,t}} \right) + \frac{1}{\lambda_j} \left(\sum_{t=2}^T \gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2 \right) \right] \right\} \\ &\propto GiG(p, a, b) \end{aligned}$$

where

$$p = 1 - \sum_{t=2}^T \gamma_{j,t}, \quad a = \sum_{t=2}^T \gamma_{j,t} \tau_{d_{j,t}}, \quad b = \sum_{t=2}^T \gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2.$$

The probability density function of a generalised inverse Gaussian, for $p \in \mathbb{R}$, $a > 0$, $b > 0$ and $K_p(\cdot)$ being a modified Bessel function of the second kind, is

$$GiG(x|p, a, b) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-(ax+b/x)/2}, \quad x \in (0, \infty).$$

The vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{n^2})$ represents the diagonal of the diagonal matrix $\mathbf{\Lambda}$ of size $(n^2 \times n^2)$.

B.3 Posterior of the stick breaking locations (μ_k, τ_k)

The posterior distribution for the stick breaking locations in the sparse case are given by:

$$\mu_0 | \dots \sim \delta_{(0)}(\mu_0).$$

and

$$\begin{aligned} P(\tau_0 | \dots) &\propto \mathcal{IG}(\tau_0 | a_0, b_0) \cdot \prod_{(j,t): \gamma_{j,t}=0} \tau_0^{-1/2} \exp \left\{ -\frac{\beta_{j,t}^2}{2\tau_0} \right\} \\ &\propto \tau_0^{-a_0-1} \exp \left\{ -\frac{b_0}{\tau_0} \right\} \tau_0^{-m_0/2} \exp \left\{ -\frac{1}{2\tau_0} \sum_{(j,t): \gamma_{j,t}=0} \beta_{j,t}^2 \right\} \\ &\propto \tau_0^{-a_0-m_0/2-1} \exp \left\{ -\frac{1}{\tau_0} \left(b_0 + \frac{1}{2} \sum_{(j,t): \gamma_{j,t}=0} \beta_{j,t}^2 \right) \right\} \\ &\propto \mathcal{IG}(\bar{a}_0, \bar{b}_0), \end{aligned}$$

with

$$\bar{a}_0 = a_0 + m_0/2, \quad \bar{b}_0 = b_0 + \frac{1}{2} \sum_{(j,t): \gamma_{j,t}=0} \beta_{j,t}^2,$$

where $m_0 = \sum_{(j,t)} \mathbb{I}(\gamma_{j,t} = 0) = \sum_{j=1}^{n^2} \sum_{t=1}^T (1 - \gamma_{j,t})$ and we used the parametrisation of the Inverse Gamma distribution with shape $a > 0$ and scale $b > 0$, that is

$$\mathcal{IG}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x}, \quad x \in (0, \infty).$$

Regarding the sample in the non-sparse case, we generate (μ_k, τ_k) for $k = 1, \dots, N^*$ and we have the following full conditional separately for μ_k and τ_k . The posterior distribution for the

stick breaking locations in the non-sparse case, is given by:

$$\begin{aligned}
P(\mu_k | \dots) &\propto \mathcal{N}(\mu_k | c, d) \cdot \prod_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \exp \left\{ -\frac{1}{2\lambda_j} (\beta_{j,t} - \mu_k)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2\bar{d}} (\mu_k^2 - 2c\mu_k) \right\} \exp \left\{ -\frac{1}{2} \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_j^{-1} - 2\mu_k \lambda_j^{-1} \beta_{j,t} \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\mu_k^2 \bar{d}^{-1} - 2c\mu_k \bar{d}^{-1} + \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_j^{-1} - 2\mu_k \lambda_j^{-1} \beta_{j,t} \right] \right\} \\
&\propto \mathcal{N}(\bar{c}, \bar{d}),
\end{aligned}$$

where

$$\bar{d} = \left(d^{-1} + \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \right)^{-1}, \quad \bar{c} = \bar{d} \left(d^{-1}c + \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \beta_{j,t} \right).$$

and for the scale parameter we have the following representation:

$$\begin{aligned}
P(\tau_k | \dots) &\propto \mathcal{G}a(\tau_k | a_1, b_1) \cdot \prod_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\
&\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \cdot \prod_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\
&\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \tau_k^{m_1} \exp \left\{ -\frac{\tau_k}{2} \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right\} \\
&\propto \tau_k^{a_1+m_1-1} \exp \left\{ -\tau_k \left(\frac{1}{b_1} + \frac{1}{2} \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right) \right\} \\
&\propto \mathcal{G}a(\bar{a}_1, \bar{b}_1),
\end{aligned}$$

with

$$\bar{a}_1 = a_1 + m_1, \quad \bar{b}_1 = \frac{2b_1}{2 + b_1 \sum_{(j,t):\gamma_{j,t}=1, d_{j,t}=k} \lambda_j},$$

where $m_1 = \sum_{(j,t)} \mathbb{I}(\gamma_{j,t} = 1) \mathbb{I}(d_{j,t} = k) = \sum_{j=1}^{n^2} \sum_{t=1}^T \gamma_{j,t} \mathbb{I}(d_{j,t} = k)$.

B.4 Posterior of the allocation variables $(d_{j,t}, \gamma_{j,t})$

Regarding the allocation variables $d_{j,t}, \gamma_{j,t}$, we obtain the following full conditional for the sparse and non-sparse case. Let $u_{j,t} \sim \mathcal{U}([0, 1])$ for all $j = 1, \dots, n^2$ and all $t = 1, \dots, T$, be the auxiliary slice sampling variable. In the non-sparse case, we have:

$$P(d_{j,t} = k, \gamma_{j,t} = 1 | \dots) \propto (1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_k, \lambda_j) \mathcal{E}xp \left(\lambda_j \frac{2}{\tau_k} \right) \mathbb{I}(u_{j,t} < w_k)$$

$$\propto \frac{(1 - \pi_t) \mathcal{N}(\beta_{j,t} | \mu_k, \lambda_j) \mathcal{E}xp\left(\lambda_j | \frac{2}{\tau_k}\right)}{\sum_{i \in A_w(u_{j,t})} \mathcal{N}(\beta_{j,t} | \mu_i, \lambda_j) \mathcal{E}xp\left(\lambda_j | \frac{2}{\tau_i}\right)}, \quad \forall k \in A_w(u_{j,t}),$$

where $A_w(u_{j,t}) = \{k \in \{1, 2, \dots, k^*\} : w_k > u_{j,t}\}$, where k^* is the number of ties. On the other hand, in the sparse case we have the non-normalised posterior probability

$$P(d_{j,t} = 0, \gamma_{j,t} = 0 | \dots) \propto \pi_t \mathcal{N}(\beta_{j,t} | 0, \tau_0) \mathbb{I}(u_{j,t} < 1),$$

and full conditional distribution has the following representation:

$$P(d_{j,t} = k, \gamma_{j,t} = 0 | \dots) \propto \begin{cases} \pi_t \mathcal{N}(\beta_{j,t} | 0, \tau_0) & \text{if } k = 0, \\ 0 & \text{if } k > 0. \end{cases}$$

B.5 Posterior for $\beta_{j,t}$

Posterior distribution for β_t , for all $t = 2, \dots, T$. Let $\boldsymbol{\mu}_t^* = (\mu_{d_{1,t}}, \dots, \mu_{d_{n^2,t}})'$ and $\boldsymbol{\gamma}_t = (\gamma_{1,t}, \dots, \gamma_{n^2,t})'$, and denote the Hadamard product by \odot . We have:

$$\begin{aligned} P(\beta_t | \dots) &\propto (2\pi)^{-n^2/2} |\boldsymbol{\Lambda}|^{-1/2} \exp\left\{-\frac{1}{2}(\beta_t - \boldsymbol{\mu}_t)' \boldsymbol{\Lambda}^{-1}(\beta_t - \boldsymbol{\mu}_t)\right\} \\ &\quad \cdot (2\pi)^{-n^2/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mathbf{X}_t \beta_t)' \boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{y}_t - \mathbf{X}_t \beta_t)\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left[\sum_{j=1}^{n^2} \left(\frac{\beta_{j,t}^2 (1 - \gamma_{j,t})}{\tau_0} + \frac{\beta_{j,t}^2 \gamma_{j,t}}{\lambda_j} - 2 \frac{\beta_{j,t} \mu_{d_{j,t}} \gamma_{j,t}}{\lambda_j} \right) + (\boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{X}_t \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{y}_t) \right]\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left[((\boldsymbol{\beta}_t \odot (1 - \boldsymbol{\gamma}_t))' \tau_0^{-1} (\boldsymbol{\beta}_t \odot (1 - \boldsymbol{\gamma}_t)) + (\boldsymbol{\beta}_t \odot \boldsymbol{\gamma}_t)' \boldsymbol{\Lambda}^{-1} (\boldsymbol{\beta}_t \odot \boldsymbol{\gamma}_t) - 2 \boldsymbol{\beta}_t' \boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu}_t^* \odot \boldsymbol{\gamma}_t)) \right. \right. \\ &\quad \left. \left. + \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{X}_t \boldsymbol{\beta}_t - 2 \boldsymbol{\beta}_t' \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t \right]\right\} \\ &\propto \mathcal{N}(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t), \end{aligned}$$

where

$$\bar{\boldsymbol{\Sigma}}_t = \left(\boldsymbol{\Lambda}_{\tau_0, \gamma}^{-1} + \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{X}_t \right)^{-1}, \quad \bar{\boldsymbol{\mu}}_t = \bar{\boldsymbol{\Sigma}}_t \left(\boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu}_t^* \odot \boldsymbol{\gamma}_t) + \mathbf{X}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t \right)$$

and $\boldsymbol{\Lambda}_{\tau_0, \gamma}$ is a $n^2 \times n^2$ diagonal matrix whose jj -th entry is

$$\boldsymbol{\Lambda}_{\tau_0, \gamma, jj} = \begin{cases} \lambda_j & \text{if } \gamma_{j,t} = 1, \\ \tau_0 & \text{if } \gamma_{j,t} = 0. \end{cases}$$

B.6 Posterior for covariance matrix $\boldsymbol{\Sigma}$

Posterior distribution for the covariance matrix $\boldsymbol{\Sigma}$.

$$P(\boldsymbol{\Sigma} | \dots) \propto \mathcal{IW}(\nu, \boldsymbol{\Psi}) \cdot \prod_{t=1}^T |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mathbf{X}_t \beta_t)' \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{X}_t \beta_t)\right\}$$

$$\begin{aligned} &\propto |\boldsymbol{\Sigma}|^{-(\nu+n+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} (\boldsymbol{\Psi} \boldsymbol{\Sigma}^{-1}) \right\} |\boldsymbol{\Sigma}|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \text{tr} ((\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)) \right\} \\ &\propto \mathcal{IW}(\bar{\nu}, \bar{\boldsymbol{\Psi}}), \end{aligned}$$

where

$$\bar{\nu} = \nu + T, \quad \bar{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)'.$$

B.7 Posterior for mixing probability π_t

Posterior distribution for the mixing probability π_t , for all $t = 1, \dots, T$.

$$\begin{aligned} P(\pi_t | \dots) &\propto \mathcal{Be}(1, \eta) \cdot \prod_{j=1}^{n^2} \pi_t^{\mathbb{I}(\gamma_{j,t}=0)} (1 - \pi_t)^{\mathbb{I}(\gamma_{j,t}=1)} \\ &\propto (1 - \pi_t)^{\eta-1} \pi_t^{n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} (1 - \pi_t)^{\sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} \\ &\propto \mathcal{Be}(\bar{\phi}, \bar{\eta}), \end{aligned}$$

where

$$\bar{\phi} = 1 + n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1), \quad \bar{\eta} = \eta + \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1).$$

C Posterior distributions: Dirac spike

C.1 Posterior for stick breaking un-normalised weights $v_{i,t}$

Posterior distribution for $v_{i,t}$, for all $t = 2, \dots, T$ and all $k = 1, \dots, k^*$, where k^* is the number of ties. We use the convention $z_{i,0} = m_{i,0} = 0$, for all $i = 1, 2, \dots$

$$\begin{aligned} P(v_{k,t}|z_{k,t}, z_{k,t-1}, \dots) &\propto P(v_{k,t}|z_{k,t-1})P(z_{k,t}|v_{k,t})P(\dots|v_{k,t}) \\ &\propto v_{k,t}^{z_{k,t}-1} (1 - v_{k,t})^{\alpha + m_{k,t-1} - z_{k,t-1} - 1} \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \left[\prod_{j:d_{j,t}=k, \gamma_{j,t}=1} v_{j,t} \right] \left[\prod_{j:d_{j,t}>k, \gamma_{j,t}=1} (1 - v_{j,t}) \right] \\ &\propto \mathcal{Be}(\bar{\xi}_t, \bar{\alpha}_t), \end{aligned}$$

where

$$\begin{aligned} \bar{\xi}_t &= 1 + z_{k,t} + z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} = k, \gamma_{j,t} = 1), \\ \bar{\alpha}_t &= \alpha + m_{k,t} - z_{k,t} + m_{k,t-1} - z_{k,t-1} + \sum_j \mathbb{I}(d_{j,t} > k, \gamma_{j,t} = 1). \end{aligned}$$

The normalised weights are then computed by $w_{k,t} = v_{k,t} \prod_{j < k} (1 - v_{j,t})$. The posterior distribution of the latent auxiliary variables $z_{k,t}$, for all $t = 1, \dots, T-1$ and all $k = 1, \dots, k^*$ is given by

$$\begin{aligned} P(z_{k,t}|v_{k,t+1}, v_{k,t}) &\propto P(z_{k,t}|v_{k,t})P(v_{k,t+1}|z_{k,t}) \\ &\propto \frac{m_{k,t}!}{z_{k,t}!(m_{k,t} - z_{k,t})!} v_{k,t}^{z_{k,t}} (1 - v_{k,t})^{m_{k,t} - z_{k,t}} \\ &\quad \cdot \frac{\Gamma(\alpha + m_{k,t} + 1)}{\Gamma(z_{k,t} + 1)\Gamma(\alpha + m_{k,t} - z_{k,t})} v_{k,t+1}^{z_{k,t}} (1 - v_{k,t+1})^{\alpha + m_{k,t} - z_{k,t} - 1} \\ &\propto \frac{1}{(\Gamma(z_{k,t} + 1))^2 \Gamma(\alpha + m_{k,t} - z_{k,t}) \Gamma(m_{k,t} - z_{k,t} + 1)} \left(\frac{v_{k,t} v_{k,t+1}}{(1 - v_{k,t})(1 - v_{k,t+1})} \right)^{z_{k,t}}. \end{aligned}$$

Finally, the auxiliary variable for the slice sampler has posterior distribution given by

$$P(u_{j,t}|\dots) \propto \begin{cases} \mathbb{I}(u_{j,t} < w_{d_{j,t}})^{\gamma_{j,t}} & \text{if } \gamma_{j,t} = 1, \\ \mathbb{I}(u_{j,t} < 1)^{1 - \gamma_{j,t}} & \text{if } \gamma_{j,t} = 0. \end{cases}$$

C.2 Posterior for λ_j

Posterior distribution for λ_j , for $j = 1, \dots, n^2$. Define $\mu_{k_j}^*, \tau_{k_j}^*$ are the location and scale of $\beta_{j,t}$, respectively, when sparse or non-sparse component is chosen.

$$P(\lambda_j|\dots) \propto \lambda_j^{-\gamma_{j,t}/2} \exp \left\{ -\gamma_{j,t} \frac{(\beta_{j,t} - \mu_{d_{j,t}})^2}{2\lambda_j} \right\} \exp \left\{ -\gamma_{j,t} \left(\frac{\lambda_j \tau_{d_{j,t}}}{2} \right) \right\}$$

$$\begin{aligned} &\propto \lambda_j^{-\gamma_{j,t}/2} \exp \left\{ -\frac{1}{2} \left[\lambda_j (\gamma_{j,t} \tau_{d_{j,t}}) + \frac{1}{\lambda_j} (\gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2) \right] \right\} \\ &\propto GiG(p, a, b) \end{aligned}$$

where

$$p = \frac{\gamma_{j,t}}{2}, \quad a = \gamma_{j,t} \tau_{d_{j,t}}, \quad b = \gamma_{j,t} (\beta_{j,t} - \mu_{d_{j,t}})^2.$$

The probability density function of a generalised inverse Gaussian, for $p \in \mathbb{R}, a > 0, b > 0$ and $K_p(\cdot)$ being a modified Bessel function of the second kind, is

$$GiG(x|p, a, b) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-(ax+b/x)/2}, \quad x \in (0, \infty).$$

The vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{n^2})$ represents the diagonal of the diagonal matrix $\mathbf{\Lambda}$ of size $(n^2 \times n^2)$.

C.3 Posterior of the stick breaking locations (μ_k, τ_k)

In the non-sparse case, we generate (μ_k, τ_k) for $k = 1, \dots, k^*$ and we have the following full conditional separately for μ_k and τ_k . The posterior distribution for the stick breaking locations in the non-sparse case, is given by:

$$\begin{aligned} P(\mu_k | \dots) &\propto \mathcal{N}(\mu_k | c, d) \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \exp \left\{ -\frac{1}{2\lambda_j} (\beta_{j,t} - \mu_k)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2d} (\mu_k^2 - 2c\mu_k) \right\} \exp \left\{ -\frac{1}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_j^{-1} - 2\mu_k \lambda_j^{-1} \beta_{j,t} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\mu_k^2 d^{-1} - 2d^{-1}c\mu_k + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \mu_k^2 \lambda_{j,t}^{-1} - 2\mu_k \lambda_{j,t}^{-1} \beta_{j,t} \right] \right\} \\ &\propto \mathcal{N}(\bar{c}, \bar{d}), \end{aligned}$$

where

$$\bar{d} = \left(d^{-1} + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \right)^{-1}, \quad \bar{c} = \bar{d} \left(d^{-1}c + \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j^{-1} \beta_{j,t} \right).$$

and for the scale parameter we have the following representation:

$$\begin{aligned} P(\tau_k | \dots) &\propto \mathcal{G}a(\tau_k | a_1, b_1) \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\ &\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \cdot \prod_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \frac{\tau_k}{2} \exp \left\{ -\frac{\tau_k}{2} \lambda_j \right\} \\ &\propto \tau_k^{a_1-1} \exp \left\{ -\frac{\tau_k}{b_1} \right\} \tau_k^{m_1} \exp \left\{ -\frac{\tau_k}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right\} \end{aligned}$$

$$\begin{aligned} &\propto \tau_k^{a_1+m_1-1} \exp \left\{ -\tau_k \left(\frac{1}{b_1} + \frac{1}{2} \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_j \right) \right\} \\ &\propto \mathcal{G}a(\bar{a}_1, \bar{b}_1), \end{aligned}$$

with

$$\bar{a}_1 = a_1 + m_1, \quad \bar{b}_1 = \frac{2b_1}{2 + b_1 \sum_{(j,t): \gamma_{j,t}=1, d_{j,t}=k} \lambda_{j,t}},$$

where $m_1 = \sum_{(j,t)} \mathbb{I}(\gamma_{j,t} = 1) \mathbb{I}(d_{j,t} = k) = \sum_{j=1}^{n^2} \sum_{t=2}^T \gamma_{j,t} \mathbb{I}(d_{j,t} = k)$.

C.4 Posterior of the allocation variable $(d_{j,t}, \gamma_{j,t})$

Regarding the allocation variables $(d_{j,t}, \gamma_{j,t})$, we obtain the following full conditional for the sparse and non-sparse case. The (conditional) prior distribution for the allocation variable $d_{j,t}$, for each $k = 1, \dots, k^*$, $j = 1, \dots, n^2$, $t = 2, \dots, T$, is given by:

$$\begin{aligned} P(d_{j,t} = k | \gamma_{j,t} = 1) &= w_k, \\ P(d_{j,t} = k | \gamma_{j,t} = 0) &= 0, \\ P(d_{j,t} = 0 | \gamma_{j,t} = 1) &= 0, \\ P(d_{j,t} = 0 | \gamma_{j,t} = 0) &= 1, \end{aligned}$$

while the prior for $\gamma_{j,t}$ is

$$\begin{aligned} P(\gamma_{j,t} = 1) &= (1 - \pi_t) \\ P(\gamma_{j,t} = 0) &= \pi_t \end{aligned}$$

Define $\boldsymbol{\gamma}_{-j,t} = (\gamma_{1,t}, \dots, \gamma_{j-1,t}, \gamma_{j+1,t}, \dots, \gamma_{n^2,t})'$, $\mathbf{d}_{-j,t} = (d_{1,t}, \dots, d_{j-1,t}, d_{j+1,t}, \dots, d_{n^2,t})'$. The joint posterior distribution of the allocation variables $(d_{j,t}, \gamma_{j,t})$, for each $k = 1, \dots, k^*$, $j = 1, \dots, n^2$, $t = 2, \dots, T$, is obtained as:

$$\begin{aligned} P(d_{j,t} = k, \gamma_{j,t} = 1 | \dots) &\propto (1 - \pi_t) P(\mathbf{y}_t | d_{j,t} = k, \gamma_{j,t} = 1, \mathbf{d}_{-j,t}, \boldsymbol{\gamma}_{-j,t}, \dots), \\ P(d_{j,t} = k, \gamma_{j,t} = 0 | \dots) &= 0, \\ P(d_{j,t} = 0, \gamma_{j,t} = 1 | \dots) &= 0, \\ P(d_{j,t} = 0, \gamma_{j,t} = 0 | \dots) &\propto \pi_t P(\mathbf{y}_t | d_{j,t} = 0, \gamma_{j,t} = 0, \mathbf{d}_{-j,t}, \boldsymbol{\gamma}_{-j,t}, \dots) = \pi_t, \end{aligned}$$

where the (conditional) marginal likelihood obtained by integrating out the $\boldsymbol{\beta}_t$, for each $t = 2, \dots, T$, is given by

$$\begin{aligned} P(\mathbf{y}_t | \boldsymbol{\gamma}_t, \mathbf{d}_t, \boldsymbol{\lambda}, \boldsymbol{\Sigma}) &= \int P(\mathbf{y}_t | \boldsymbol{\beta}_t, \boldsymbol{\Sigma}) P(\boldsymbol{\beta}_t | \boldsymbol{\gamma}_t, \mathbf{d}_t, \boldsymbol{\lambda}) d\boldsymbol{\beta}_t \\ &= \int (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t) \right\} \cdot \left[\prod_{j: \gamma_{j,t}=0} \delta_{(0)}(\beta_{j,t}) \right] \\ &\quad \cdot (2\pi)^{-q/2} |\boldsymbol{\Lambda}_{\boldsymbol{\gamma}, \mathbf{d}, t}|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_{\boldsymbol{\gamma}, t} - \boldsymbol{\mu}_{\boldsymbol{\gamma}, \mathbf{d}, t})' \boldsymbol{\Lambda}_{\boldsymbol{\gamma}, t}^{-1} (\boldsymbol{\beta}_{\boldsymbol{\gamma}, t} - \boldsymbol{\mu}_{\boldsymbol{\gamma}, \mathbf{d}, t}) \right\} d\boldsymbol{\beta}_{\mathbf{0}, t} d\boldsymbol{\beta}_{\boldsymbol{\gamma}, t} \\ &= (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} |\boldsymbol{\Lambda}_{\boldsymbol{\gamma}, \mathbf{d}, t}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t + \boldsymbol{\mu}'_{\boldsymbol{\gamma}, \mathbf{d}, t} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}, \mathbf{d}, t}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}, \mathbf{d}, t} \right] \right\} \end{aligned}$$

$$\begin{aligned}
& \cdot \int \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}'_{\gamma,t} (\mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\gamma,d,t} + \boldsymbol{\Lambda}_{\gamma,d,t}^{-1}) \boldsymbol{\beta}_{\gamma,t} - 2(\mathbf{y}_t \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\gamma,d,t} + \boldsymbol{\mu}'_{\gamma,d,t} \boldsymbol{\Lambda}_{\gamma,d,t}^{-1}) \boldsymbol{\beta}_{\gamma,t} \right] \right\} d\boldsymbol{\beta}_{\gamma,t} \\
& = (2\pi)^{-\frac{n}{2}} \left(\frac{|\boldsymbol{\Sigma}_{\beta_{\gamma,t}}|}{|\boldsymbol{\Sigma}|} \right)^{\frac{1}{2}} |\boldsymbol{\Lambda}_{\gamma,d,t}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_t \boldsymbol{\Sigma}^{-1} \mathbf{y}_t + \boldsymbol{\mu}'_{\gamma,d,t} \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} \boldsymbol{\mu}_{\gamma,d,t} - \boldsymbol{\mu}'_{\beta_{\gamma,t}} \boldsymbol{\Sigma}_{\beta_{\gamma,t}}^{-1} \boldsymbol{\mu}_{\beta_{\gamma,t}} \right] \right\},
\end{aligned}$$

where $\mathbf{X}_{\gamma,d,t}$ contains the columns of \mathbf{X}_t corresponding to the j such that $\gamma_{j,t} = 1$ and $d_{j,t} = \mathbf{d}_j$ (similarly for $\boldsymbol{\Lambda}_{\gamma,d,t}$), $\boldsymbol{\mu}_{\gamma,d,t}$ is the sub-vector of $(\mu_{d_{1,t}}, \dots, \mu_{d_{n_2,t}})'$ containing the elements j such that $\gamma_{j,t} = 1$, and

$$\boldsymbol{\Sigma}_{\beta_{\gamma,t}} = (\mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\gamma,d,t} + \boldsymbol{\Lambda}_{\gamma,d,t}^{-1})^{-1}, \quad \boldsymbol{\mu}_{\beta_{\gamma,t}} = \boldsymbol{\Sigma}_{\beta_{\gamma,t}} (\mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{y}_t + \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} \boldsymbol{\mu}_{\gamma,d,t}).$$

C.5 Posterior for $\beta_{j,t}$

Posterior distribution for $\boldsymbol{\beta}_t$, for all $t = 2, \dots, T$. In the sparse case we have

$$P(\beta_{j,t} | \gamma_{j,t} = 0, \dots) = \delta_{(0)}(\beta_{j,t}).$$

In the non-sparse case, denoting $\boldsymbol{\beta}_{\gamma,t}$ the sub-vector of $\boldsymbol{\beta}_t$ corresponding to the coefficients such that $\gamma_{j,t} = 1$, we have

$$\begin{aligned}
P(\boldsymbol{\beta}_{\gamma,t} | \dots) & \propto (2\pi)^{-n^2/2} |\boldsymbol{\Lambda}_{\gamma,d,t}|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_{\gamma,t} - \boldsymbol{\mu}_{\gamma,d,t})' \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} (\boldsymbol{\beta}_{\gamma,t} - \boldsymbol{\mu}_{\gamma,d,t}) \right\} \\
& \cdot (2\pi)^{-n^2/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{X}_{\gamma,d,t} \boldsymbol{\beta}_{\gamma,t})' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{y}_t - \mathbf{X}_{\gamma,d,t} \boldsymbol{\beta}_{\gamma,t}) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}'_{\gamma,t} \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} \boldsymbol{\beta}_{\gamma,t} - 2\boldsymbol{\beta}'_{\gamma,t} \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} \boldsymbol{\mu}_{\gamma,d,t} + \boldsymbol{\beta}'_{\gamma,t} \mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\gamma,d,t} \boldsymbol{\beta}_{\gamma,t} - 2\boldsymbol{\beta}'_{\gamma,t} \mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{y}_t \right] \right\} \\
& \propto \mathcal{N}(\boldsymbol{\mu}_{\beta_{\gamma,t}}, \boldsymbol{\Sigma}_{\beta_{\gamma,t}}),
\end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\beta_{\gamma,t}} = (\mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\gamma,d,t} + \boldsymbol{\Lambda}_{\gamma,d,t}^{-1})^{-1}, \quad \boldsymbol{\mu}_{\beta_{\gamma,t}} = \boldsymbol{\Sigma}_{\beta_{\gamma,t}} (\mathbf{X}'_{\gamma,d,t} \boldsymbol{\Sigma}^{-1} \mathbf{y}_t + \boldsymbol{\Lambda}_{\gamma,d,t}^{-1} \boldsymbol{\mu}_{\gamma,d,t}).$$

C.6 Posterior for covariance matrix $\boldsymbol{\Sigma}$

Posterior distribution for the covariance matrix $\boldsymbol{\Sigma}$.

$$\begin{aligned}
P(\boldsymbol{\Sigma} | \dots) & \propto \mathcal{IW}(\nu, \boldsymbol{\Psi}) \cdot \prod_{t=2}^T |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t) \right\} \\
& \propto |\boldsymbol{\Sigma}|^{-\frac{\nu+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Psi} \boldsymbol{\Sigma}^{-1}) \right\} |\boldsymbol{\Sigma}|^{-\frac{T-1}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=2}^T \text{tr}((\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)) \right\} \\
& \propto \mathcal{IW}(\bar{\nu}, \bar{\boldsymbol{\Psi}}),
\end{aligned}$$

where

$$\bar{\nu} = \nu + T - 1, \quad \bar{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \sum_{t=2}^T (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)'$$

C.7 Posterior for mixing probability π_t

Posterior distribution for the mixing probability π_t , for all $t = 2, \dots, T$.

$$\begin{aligned} P(\pi_t | \dots) &\propto \mathcal{Be}(1, \eta) \cdot \prod_{j=1}^{n^2} \pi_t^{\mathbb{I}(\gamma_{j,t}=0)} (1 - \pi_t)^{\mathbb{I}(\gamma_{j,t}=1)} \\ &\propto (1 - \pi_t)^{\eta-1} \pi_t^{n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} (1 - \pi_t)^{\sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t}=1)} \\ &\propto \mathcal{Be}(\bar{\phi}, \bar{\eta}), \end{aligned}$$

where

$$\bar{\phi} = 1 + n^2 - \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1), \quad \bar{\eta} = \eta + \sum_{j=1}^{n^2} \mathbb{I}(\gamma_{j,t} = 1).$$