

Estimating the Ordering of Variables in a VAR using a Plackett-Luce Prior*

Ping Wu[†] Gary Koop[‡]

June 2023

Abstract

Estimating Bayesian Vector Autoregressions (VARs) involving the Cholesky decomposition is sensitive to the ordering of variables. We treat the ordering as unknown, develop a prior over variable orderings and Markov Chain Monte Carlo (MCMC) methods for posterior sampling over orderings.

JEL classification: C11, C32

Keywords: Variables ordering, Plackett-Luce model

*We would like to thank Rafaella Giacomini and participants at the NBER-NSF SBIES conference 2022 for useful comments.

[†]University of Strathclyde (Address: Department of Economics, University of Strathclyde, 199 Cathedral Street, Glasgow, G4 0QU. E-mail: ping.wu@strath.ac.uk)

[‡]University of Strathclyde

1 Introduction

Bayesian VARs have traditionally been estimated in reduced form, where the left hand side of the VAR equation involves an $n \times 1$ vector of dependent variables, \mathbf{y}_t , and the error covariance matrix for the VAR, Σ_t , is unrestricted (apart from being positive definite). However, as VARs have become larger and larger, researchers have increasingly been estimating VARs in a structural form where the left hand side of the VAR is $\mathbf{B}_0 \mathbf{y}_t$ and the error covariance matrix is diagonal. \mathbf{B}_0 is a lower triangular matrix based on the Cholesky decomposition of Σ_t . The fact that the error covariance matrix of the structural VAR is diagonal means that Bayesian estimation can proceed one equation at a time. As shown, e.g., in Carriero et al. (2019) the MCMC algorithm based on the reduced form VAR requires $O(n^6)$ elementary operations to take one draw of the VAR coefficients. This is reduced to $O(n^4)$ with the structural VAR. Thus, in larger VARs, it can be computationally impractical to work in the reduced form, but feasible when working in the structural form.¹

However, the use of the Cholesky decomposition of the reduced form error covariance matrix leads to order dependence. That is, posterior and predictive results will vary depending on the way that the variables are ordered in the VAR. To clarify the precise nature of this order dependence, we highlight the discussion of sub-section 3.1 of Carriero et al. (2019) who demonstrate that the posterior of the structural form VAR coefficients, conditional on Σ_t is invariant to ordering. The lack of order invariance arises due to the fact that the implied prior on Σ_t is not order invariant. Bognanni (2018) and Arias et al. (2021) demonstrate the importance of the ordering issue in VARs with stochastic volatility (SV) both theoretically and empirically. Arias et al. (2021) shows that, although point forecasts are not sensitive to the way variables are ordered, predictive standard deviations can be substantially affected by the way that the variables are ordered. The VARs considered in Arias et al. (2021) are all low dimensional. Chan et al. (2021) consider ordering issues in high dimensional VARs and demonstrate that the theoretical and empirical findings of Arias et al. (2021) hold with additional force in higher dimensions. Thus, there is growing theoretical and empirical evidence that ordering issues are important, particularly in the large VARs that cannot easily be estimated in reduced form due to the computational burden.

These considerations have stimulated interest in order invariant approaches. Reduced form estimation of VARs is order invariant with commonly-used priors, but is not scalable to large VARs. Chan et al. (2021) critiques various order invariant approaches and proposes a new order invariant approach which avoids the use of the Cholesky decomposition, relies on stochastic volatility to identify the model and is scalable. Wu and Koop

¹Numerical instability issues can also plague the Bayesian estimation of reduced form VARs due to the need to invert and/or take Cholesky decompositions of enormous posterior covariance matrices for the VAR coefficients.

(2022) develop an order-invariant approach based on Eigendecomposition. The present paper adopts a different strategy to address the ordering issue. We retain the Cholesky decomposition of the reduced form error covariance but develop methods for finding the optimal ordering of the variables. In this, it shares some similarities to Levy and Lopes (2021) which also uses the Cholesky decomposition for the error covariance matrix for a multivariate time series model and develops Bayesian methods for searching over variable orderings. But the model class they use is different from ours and they use approximate discounting methods which contrasts with our use of MCMC methods. In addition, they do not use a prior over variable orderings.

We treat the ordering of variables in the VAR as unknown and estimate it using Bayesian methods. We do so by developing a prior over variable orderings. The Plackett-Luce distribution is commonly used for the likelihood function for rank ordered data. We use it, not for the likelihood function, but rather as a prior. We develop MCMC methods for the posterior that results. Our model is a Bayesian VAR with SV and Placket-Luce prior over variable orderings which we abbreviate to BVARSV-PL. In a macroeconomic forecasting exercise involving a VAR with 20 variables we compare our BVARSV-PL to a BVARSV which is identical in all respects except that a single ordering is used. We choose this model to be that of Chan (2021) and, for BVARSV, use the same ordering as in this paper. This can be interpreted as a model which is a special case of ours, but has a prior which dogmatically imposes a particular variable ordering. We find that our BVARSV-PL forecasts better than this BVARSV, thus demonstrating the importance of ordering choice in VARs for macroeconomic forecasting.

2 Bayesian Inference on Orderings in VARs

In this section, we develop Bayesian methods for carrying out inference on ways of ordering the variables in VARs. After defining the likelihood function for our VAR, we develop our Plackett-Luce prior on variable ordering. Subsequently, we develop MCMC methods which allow for posterior inference and prediction.

Our VAR with stochastic volatility is:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{B}_0^{-1} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \quad (1)$$

where \mathbf{a} is an $n \times 1$ vector of intercepts, $\mathbf{A}_1, \dots, \mathbf{A}_p$ are $n \times n$ matrices of VAR coefficients, \mathbf{B}_0 is an $n \times n$ matrix, and $\mathbf{D}_t = \text{diag}(e^{h_{1,t}}, \dots, e^{h_{n,t}})$ is diagonal. Each of the log-volatilities follows a stationary AR(1) process:

$$h_{i,t} = \phi_i h_{i,t-1} + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}(0, \omega_i^2)$$

where $\mathcal{N}(\cdot, \cdot)$ denotes the Gaussian distribution.²

²It is simple to extend our methods to the homoskedastic VAR or to allow for \mathbf{A}_i for $i = 1, \dots, p$ to

\mathbf{B}_0 is

$$\mathbf{B}_0 = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & 1 & & b_{2,n} \\ \vdots & & \ddots & \vdots \\ b_{n,1} & \dots & b_{n,n-1} & 1 \end{pmatrix}. \quad (2)$$

A common assumption, based on taking the Cholesky decomposition of the reduced form error covariance matrix, is that \mathbf{B}_0 is a lower triangular matrix with ones on the diagonal. It is this which leads to order dependence as shown, e.g., in Chan et al. (2021). But the lower triangularity assumption has the advantage that it leads to simple and fast Bayesian computation. Hence, in this paper, we retain the lower triangularity assumption but express uncertainty over which lower triangular form is appropriate. To be precise, we use priors which rule out all forms for \mathbf{B}_0 except those that can be, via permutations of the columns of the matrix, put into a lower triangular form with ones on the diagonal.

Before discussing our prior for \mathbf{B}_0 , we emphasize that we make standard assumptions for the other parameters in this model (i.e. the VAR coefficients and those in the SV processess). There are a range of possible priors for the VAR with SV which could be used. In our empirical work, we use the prior of Chan (2021).³

To describe the prior for the orderings, we introduce a discrete random $1 \times n$ vector $\boldsymbol{\rho}$ which can take on realizations $\boldsymbol{\rho}^l$ for $l = 1, \dots, L$. Each $\boldsymbol{\rho}^l$ defines an ordering of the variables with ρ_1^l being the variable which takes the first place under ordering l , ρ_2^l the second, etc. For instance, in a VAR with 3 variables there are 6 possible orderings (e.g. 1,2,3; 1,3,2; 2,1,3; etc.) and, thus, $L = 6$. If $\boldsymbol{\rho}^2 = (1, 3, 2)$, then the prior for \mathbf{B}_0 conditional on $\boldsymbol{\rho}^2$ will imply a lower triangular form for \mathbf{B}_0 consistent with the variables being ordered as (1,3,2).

We stress that the variables in \mathbf{y}_t are ordered in a particular way (from variable 1 through variable n) and that they always appear in this order in the model. When we refer to variable orderings and lower triangular forms this relates to \mathbf{B}_0 with the idea being that, through appropriate permutations of its columns, it becomes lower triangular. That is, formally the variables in \mathbf{y}_t never change order, nor is \mathbf{B}_0 itself necessarily lower triangular. However, we consider a prior which only allows for choices of \mathbf{B}_0 which can be transformed into lower triangular form after appropriate switching of its columns. This

be time varying. Allowing for \mathbf{B}_0 to be time varying would be a greater challenge, but could be done by replacing our Plackett-Luce prior with a dynamic Plackett-Luce prior.

³A word is in order about identification. The homoskedastic version of our model is unidentified, although adding SV identifies the model as shown in Bertsche and Braun (2022). It is well-known that identification is not necessary to carry out Bayesian prediction in models which have proper priors. Even for the homoskedastic version of our model, updating will occur in the sense that the posterior will differ from the prior. This is an example of what Giacomini et al. (2022) call uncertain identification.

is equivalent to considering all possible structural VARs with lower triangular triangular impact matrices for different variable orderings. When we use phrases below which refer to different orderings of the variables, they should be interpreted in this context.

For each choice of l we restrict \mathbf{B}_0 to lower triangular form (after permutations) through a Dirac spike-and-slab prior. Lower triangularity is obtained if, for $i < j$, $i, j = 1, \dots, n$, variable ρ_i^l is ordered before variable ρ_j^l , $i \neq j$, then it is zero, otherwise it is non-zero. Thus,

$$p\left(b_{\rho_i^l, \rho_j^l} \mid \boldsymbol{\rho} = \boldsymbol{\rho}^l\right) = \mathbb{1}_{(i < j)} \Delta_0\left(b_{\rho_i^l, \rho_j^l}\right) + \mathbb{1}_{(i > j)} \mathcal{N}\left(a_b, V_b\right), \quad i \neq j. \quad (3)$$

We set $c = 0.0001$ and $V_b = 1$.

This defines the prior for \mathbf{B}_0 conditional on a specific ordering.

We now need a prior over $\boldsymbol{\rho}$. When dealing with ordered or ranked data, the Plackett-Luce model, (Luce, 1959; Plackett, 1975), is a popular one.⁴ The Plackett-Luce distribution is parameterized by the parameter $\lambda_j > 0$ for $j = 1 \dots n$ which represents the skill rating or ability of each variable (i.e. λ_j is the probability that variable $y_{j,t}$ is ordered first in the VAR). Our prior is

$$p\left(\boldsymbol{\rho} = \boldsymbol{\rho}^l \mid \lambda\right) = \prod_{i=1}^n \frac{\lambda_{\rho_i^l}}{\sum_{k=i}^n \lambda_{\rho_k^l}} \quad (4)$$

where $\lambda = \{\lambda_i\}_{i=1}^n$. It is a hierarchical prior in that λ is treated as unknown and estimated.

There is an alternative way of writing the Plackett-Luce model which involves introducing a latent variable z_i for each variable

$$z_i \mid \lambda_i \sim \mathcal{E}\left(\lambda_i\right), \quad i = 1, 2, \dots, n \quad (5)$$

where \mathcal{E} denote the exponential distribution. If we define

$$p\left(\boldsymbol{\rho} = \boldsymbol{\rho}^l \mid \mathbf{z}\right) = \Pr\left(\rho_1^l = \min\{z_{\rho_1^l}, \dots, z_{\rho_n^l}\}\right) \times \dots \times \Pr\left(\rho_{n-1}^l = \min\{z_{\rho_{n-1}^l}, z_{\rho_n^l}\} \mid z_{\rho_1^l}, \dots, z_{\rho_{n-2}^l}\right). \quad (6)$$

and calculate

$$p\left(\boldsymbol{\rho} = \boldsymbol{\rho}^l \mid \lambda\right) = \int p\left(\boldsymbol{\rho} = \boldsymbol{\rho}^l \mid \mathbf{z}\right) p\left(\mathbf{z} \mid \lambda\right) d\mathbf{z}, \quad (7)$$

⁴The Plackett-Luce model assumes the ranking is complete, an assumption we maintain in this paper. However, it is worth noting that Bayesian methods for the extended Plackett-Luce model are available (Johnson et al., 2021) and could easily be used in our approach. This model allows for the ranking to be incomplete which, in some cases, could be useful for VAR ordering problems. For instance, a researcher may wish to know whether a block of macroeconomic variables is ordered before or after a block of financial variables but does not care about the ordering of variables within each block.

it can be shown to be the same prior as in (4).

Finally, we require a prior for λ , which is the vector of abilities for the variables. For these, we adopt independent Gamma distributions:

$$p(\lambda) = \prod_{i=1}^n \mathcal{G}(\lambda_i; a_{\lambda,i}, b_{\lambda,i}).$$

As explained in Caron and Doucet (2012), the hyperparameters $b_{\lambda,i}$ are just scaling parameters on λ_i . As the likelihood is invariant to a rescaling of the λ_i , these hyperparameters can be fixed without influencing inference. Following Henderson and Kirrane (2018) we set $b_{\lambda,i} = 1$. Following Caron and Doucet (2012) we estimate the other set of prior hyperparameters, a_λ . We begin by setting $a_{\lambda,i} = a_\lambda$ for $i = 1, \dots, n$ which leads to an exchangeable prior specification in which we believe that some variables are better than others but we have no beliefs about which the stronger and the weaker variables are (Henderson and Kirrane, 2018). Thus, we assume $\lambda_i \sim \mathcal{G}(a_\lambda, 1)$ and treat a_λ as an unknown parameter and give it a noninformative prior (i.e. Uniform over the positive real line).

Posterior inference is carried out using an MCMC algorithm. The MCMC algorithm for the VAR coefficients, \mathbf{B}_0 and the SV processes, conditional on a given ordering, is standard, see Chan et al. (2021) and will not be given here. The algorithm for drawing \mathbf{z} , which can then be used to provide draws of $\boldsymbol{\rho}$ using (6), is provided in the Appendix. Intuitively, it will first search over all n variables to draw the variable to be ordered first, then search over the remaining $n - 1$ variables to find the variable ordered second, etc. This we call the forward step and it leads to a draw we label z_j^1 for $j = 1, \dots, n$. We then repeat the process in reverse ordering (i.e. first drawing the variable ordered last, then the variable ordered second last, etc.), leading to z_j^2 for $j = 1, \dots, n$. This is the backward step. Theoretically, we could use only the forward or the backward step in a valid MCMC algorithm, but we find MCMC convergence to be much better by including both steps. These appear in the conditional posterior for λ . In particular, for $i = 1, \dots, n$, we draw the ability parameters from:

$$\lambda_i \mid \mathbf{z}, a_\lambda \sim \mathcal{G} \left(a_\lambda + w_i, b_\lambda + \sum_{k=1}^2 \sum_{j=1}^n z_j^k \right). \quad (8)$$

where w_i denotes the total number of wins of variable i (i.e. cases where variable i is ordered before another variable). If we let w_{ij} denote the number of cases where i beats j , then $w_i = \sum_{j=1, j \neq i}^n w_{ij}$.

For a_λ , we use the Metropolis-Hastings step proposed in Caron and Doucet (2012). It proceeds by taking candidate draws, denoted $\log(a_\lambda^c)$, from $\mathcal{N}(\ln(a_\lambda), \omega^2)$, where ω^2 is a tuning parameter and we set $\omega^2 = 0.1^2$ in our application. The acceptance probability

for a_λ^c is given by

$$\min \left\{ 1, b_\lambda^{n(a_\lambda^c - a_\lambda)} \left(\frac{\Gamma(a_\lambda)}{\Gamma(a_\lambda^c)} \right)^n (\lambda)^{a_\lambda^c - a_\lambda} \right\},$$

where $\Gamma(\cdot)$ denotes the Gamma function.

3 Empirical Application

We use the first twenty variables, transformed and ordered in the same way, from the data set used in Chan (2021) but we update the data set so that it runs from 1960Q1 through 2021Q3. For brevity, we do not provide details here but note that four important variables are GDP (GDPC1), industrial production (INDPRO), the unemployment rate (UNRATE) and the consumer price index (CPIAUCSL). The benchmark model is the BVARSV with variables in the same order as in Chan (2021). We refer the reader to the the label on the Y-axis of Figure 1 which lists the variables in the order which they enter the BVARSV. All other modelling choices (i.e. prior and lag length) are the same as in Chan (2021) for both BVARSV and BVARSV-PL. For BVARSV-PL we additionally require the prior over the orderings which is described in the preceding section. The forecasting exercise uses an evaluation period beginning in 1988Q1. Forecasts are iterative for horizons $h = 1$ and $h = 4$. To assess forecasting accuracy, we use root mean square forecast errors (RMSFEs) for point forecasts and average log predictive likelihoods (ALPLs) for density forecasts. These are benchmarked so that they represent percentage improvements in forecast performance of the BVARSV-PL relative to the BVARSV.

MCMC convergence was checked by analyzing the difference between multiple Markov chains obtained from different starting values. Twenty Monte Carlo Markov chains, each initiated at a random draw of $\boldsymbol{\rho}^l, l = 1, \dots, 20$ are taken. Initial values for \mathbf{B}_0 are either set to 0 or 0.5 depending on the initial value for $\boldsymbol{\rho}^l$. Each Markov chain was run for 20,000 iterations. Convergence of each parameter in $\boldsymbol{\rho}$ was checked with the Gelman and Rubin diagnostics in Gelman and Rubin (1992). Table 1 reports the 10 largest diagnostics. As Brooks and Gelman (1998) have suggested, if diagnostics are less than 1.2, one can be confident that convergence has been reached.

Table 1: Gelman and Rubin diagnostics (the 10 largest)

ρ_3	ρ_{18}	ρ_{19}	ρ_5	ρ_7	ρ_{14}	ρ_4	ρ_1	ρ_{13}	ρ_{20}
1.12	1.03	1.02	1.02	1.02	1.02	1.01	1.01	1.01	1.01

Evidence on posterior ordering is presented in Figure 1. This presents the probability (obtained by averaging over MCMC draws) that each variable appears in each possible

order. For instance, the fact that box in the row labeled FEDFUNDS and column labelled 1 is dark blue means there is a probability near one that the FEDFUNDS should be ordered first. It turns out that there are many probabilities near one or near zero, with few values in intermediate regions and is choosing an ordering of variables that is very different from that of Chan (2021). Note that the ordering used in Chan (2021), which puts macroeconomic variables first and financial variables last, is commonly used in many large VAR forecasting papers. But we are finding strong evidence that the optimal ordering involves financial variables to be ordered first.

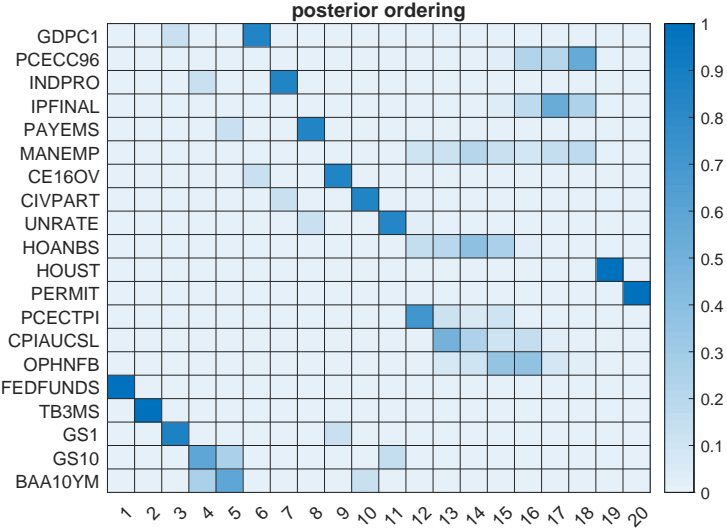


Figure 1: Posterior Ordering: BVARSV-PL

Table 2 shows that estimating the ordering leads to substantial forecast improvement relative to a strategy of following a single conventional ordering of variables. We find that there is not much difference between BVARSV-PL and BVARSV in terms of their point forecasts (i.e. RMSFEs for one model do not consistently beat the other), but that the density forecasts produced by BVARSV-PL are substantially better (i.e. ALPLs are much higher for all variables and for both forecast horizons).

Table 2: Forecasting Results

	% gains in RMSFE		% gains in ALPL	
	h=1	h=4	h=1	h=4
GDPC1	-5.51	-0.49	13.51	51.20
INDPRO	0.34	-0.83	14.70	2.79
UNRATE	-9.50	-18.65	97.68	265.95
CPIAUCSL	0.23	1.32	3.15	16.98

4 Conclusions

Bayesians commonly work with structural VARs which involve the Cholesky decomposition of the error covariance matrix. But posterior and predictive results can depend on the way the variables are ordered in the VAR. To overcome this problem, while retaining the Cholesky-transformed VAR, this paper has developed an algorithm for finding the optimal ordering of variables. We find the algorithm to lead to forecast improvements relative to a strategy of choosing a particular ordering a priori.

References

- J. E. Arias, J. F. Rubio-Ramirez, and M. Shin. Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models. *Federal Reserve Bank of Philadelphia Working Papers*, 2021.
- D. Bertsche and R. Braun. Identification of structural Vector Autoregressions by stochastic volatility. *Journal of Business & Economic Statistics*, 40:328–341, 2022.
- M. Bognanni. A class of time-varying parameter structural vars for inference under exact or set identification. *Federal Reserve Bank of Cleveland, Working Paper No. 18-11*. <https://doi.org/10.26509/frbc-wp-201811>, 2018.
- S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.
- F. Caron and A. Doucet. Efficient Bayesian inference for generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, 2012.
- A. Carriero, T. E. Clark, and M. Marcellino. Large Bayesian Vector Autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212:137–154, 2019.
- J. Chan. Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting*, 37:1216–1226, 2021.
- J. C. Chan, G. Koop, and X. Yu. Large order-invariant Bayesian VARs with stochastic volatility. *arXiv preprint arXiv:2111.07225*, 2021.
- A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.
- R. Giacomini, T. Kitagawa, and A. Volpicella. Uncertain identification. *Quantitative Economics*, 13:95–123, 2022.

- D. A. Henderson and L. J. Kirrane. A comparison of truncated and time-weighted plackett–luce models for probabilistic forecasting of formula one results. *Bayesian Analysis*, 13(2):335–358, 2018.
- S. Johnson, D. Henderson, and R. Boys. On Bayesian inference for the extended Plackett-Luce model. *Bayesian Analysis*, 17:1–26, 2021.
- B. Levy and H. Lopes. Dynamic ordering learning in multivariate forecasting. <https://arxiv.org/abs/2101.04164>, 2021.
- R. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.
- R. Plackett. The analysis of permutations. *Applied Statistics*, 24:193–202, 1975.
- P. Wu and G. Koop. Fast, order-invariant Bayesian inference in VARs using the eigen-decomposition of the error covariance matrix. 2022.

Appendix: MCMC algorithm for ρ

This appendix describes how draws from the orderings are taken from their conditional posterior density: $p(\rho \mid \mathbf{y}, \mathbf{A}, \mathbf{D}, \mathbf{B}_0, \mathbf{z}, \lambda, a_\lambda)$.

Our algorithm consists of two steps: the forward step and the backward step.

The forward step:

1) we first decide which variable will take the first place. There are n variables in total, so we compare the n variables. The winner needs to beat all other $n - 1$ variables. The success of variable i means all other variables will not appear in this equation, that is, the i -th row in \mathbf{B}_0 will be zero and the prior will come from the Dirac spike component. The success of variable i also means this variable will appear in all other equations, that is, the i -th column in \mathbf{B}_0 will be non-zero and the prior will come from the slab component (except itself, because it always appears in its own equation. This is not influenced by the ordering).

Let b_i denote the i -th row in \mathbf{B}_0 and delete the i -th element (which is the variable itself), $b_{\cdot i}$ denote the i -th column in \mathbf{B}_0 and delete the i -th element. Let p_i denote the success probability of variable i . Then:

$$p_i = p(\rho_1 = i \mid \mathbf{B}_0) \propto \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} \phi(b_{i\cdot}; 0, c) \phi(b_{\cdot i}; a_b, V_b),$$

where ϕ denotes the pdf of the Normal distribution. Hence, after normalization, we obtain

$$p_i = \frac{p_i}{\sum_{k=1}^n p_k}.$$

We can sample from the multinomial distribution as follows:

- a) Create an array p containing the cumulative probabilities of p_i , $i = 1, 2, \dots, n$;
- b) Generate U , a uniform(0,1) random value;
- c) Select the first index such that $p_i > U$.

2) conditioning on the first place, we can decide which variable will take the second place. There are $n - 1$ variables in total, so we compare the $n - 1$ variables. The winner needs to beat all other $n - 2$ variables. The success of variable j means all other variables will not appear in this equation, that is, the j -th row in \mathbf{B}_0 will be zero and the prior will come from the Dirac spike component (except the variable which takes the first place). The success of variable j also means this variable will appear in all other equations, that is, the j -th column in \mathbf{B}_0 will be non-zero and the prior will come from the slab component (also, except the variable which takes the first place).

Let $b_{j\cdot}$ denote the j -th row in \mathbf{B}_0 , delete the j -th element (which is the variable itself) and delete the ρ_1^* -th element (which takes the first place), $b_{\cdot j}$ denote the j -th column in \mathbf{B}_0 , delete the j -th element and delete the ρ_1^* -th element. Let p_j denote the success probability of variable j . Then:

$$p_j = p(\rho_2 = j \mid \mathbf{B}_0) \propto \frac{\lambda_j}{\sum_{k=1, k \neq \rho_1^*}^n \lambda_k} \phi(b_{j\cdot}; 0, c) \phi(b_{\cdot j}; a_b, V_b).$$

Hence, after normalization, we obtain

$$p_j = \frac{p_j}{\sum_{k=1, k \neq \rho_1^*}^n p_k}.$$

⋮

We can proceed until the $(n - 1)$ -th place. The remaining variable will automatically take the last place.

The backward step:

This is almost the same as the forward step, except that we first draw the variable ordered last, then draw the variable ordered second last, etc..