Bayesian Estimation of Correlation Matrices of Longitudinal Data

Riddhi Pratim Ghosh[*], Bani Mallick[†], and Mohsen Pourahmadi[‡]

Abstract. Estimation of correlation matrices is a challenging problem due to the notorious positive-definiteness constraint and high-dimensionality. Reparameterizing Cholesky factors of correlation matrices in terms of angles or hyperspherical coordinates where the angles vary freely in the range $[0, \pi)$ has become popular in the last two decades. However, it has not been used in Bayesian estimation of correlation matrices perhaps due to lack of clear statistical relevance and suitable priors for the angles. In this paper, we show for the first time that for longitudinal data these angles are the inverse cosine of the semi-partial correlations (SPCs). This simple connection makes it possible to introduce physically meaningful selection and shrinkage priors on the angles or correlation matrices with emphasis on selection (sparsity) and shrinking towards longitudinal structure. Our method deals effectively with the positive-definiteness constraint in posterior computation. We compare the performance of our Bayesian estimation based on angles with some recent methods based on partial autocorrelations through simulation and apply the method to a data related to clinical trial on smoking.

Keywords: angular parameterization, Cholesky decomposition, longitudinal data, selection, shrinkage, structured correlation matrix.

1 Introduction

Covariance and correlation matrices play a fundamental role in every aspect of multivariate statistics (Anderson, 2003). Flexible modeling and estimation of correlation matrices are daunting tasks due to (1) the positive-definiteness constraint, (2) the number of unknown elements growing quadratically with the dimension and (3) the diagonal entries being the same and equal to one. More specifically, modeling directly the individual elements requires repeatedly checking the positive-definiteness of the estimated matrix in an iterative model fitting procedure. Many strategies involving various matrix decompositions such as spectral, Cholesky decomposition, and factor models have been developed to circumvent the notorious positive-definiteness constraint (Chiu et al., 1996; Pinheiro and Bates, 1996; Pourahmadi, 1999; Fan et al., 2008). Unconstrained parameterization of correlation matrices using angles and partial autocorrelations has been around for a while (Pinheiro and Bates, 1996; Rapisarda et al., 2007; Joe, 2006) where the partial autocorrelations and angles as new parameters vary freely in the ranges

[*]Department of Biostatistics, University of Pennsylvania, Philadelphia, PA 19104, USA, RiddhiPratim.Ghosh@pennmedicine.upenn.edu
[†]Department of Statistics, Texas A&M University, College Station, TX 77840, USA, bmallick@stat.tamu.edu
[‡]Department of Statistics, Texas A&M University, College Station, TX 77840, USA, pourahm@stat.tamu.edu

$[-1, 1]$ (Joe, 2006) and $[0, \pi)$ (Pinheiro and Bates, 1996; Pourahmadi and Wang, 2015; Tsay and Pourahmadi, 2017), respectively.

In Bayesian covariance or precision matrix estimation, inverse-Wishart distribution and its various modifications have been widely used due to its mathematical simplicity. The *separation strategy* in Barnard et al. (2000) writes a covariance matrix as $\Sigma = SRS$ and assumes independent prior for the diagonal matrix of standard deviations S and the correlation matrix R. Specifically they considered two priors: The first one is marginally uniform prior for r_{ij} obtained from marginal distribution on R starting from a standard inverse-Wishart prior on Σ with specific choices of hyperparameters, and the second one is the jointly uniform prior of the form $p(R) \propto 1$ over $\mathcal{R}^k$ where $\mathcal{R}^k$ denotes the space of all valid correlation matrices of dimension k. In the context of graphical model selection, Kundu et al. (2019) used regularized inverse-Wishart prior on Σ and showed its connection to shrinkage in Σ^{-1} through equivalence with a Cholesky-based regularization approach. In this article we restrict our attention only on the estimation of a correlation matrix for ordered data and not the covariance matrix owing to the *separation strategy*. We address some Bayesian modeling and inferential challenges in estimating a correlation matrix by introducing suitable priors on the angles which go beyond the traditional use of the inverse-Wishart distribution, the marginal and joint uniform priors in Barnard et al. (2000).

Structured correlation matrices play an important role in many applications. Liechty et al. (2004) considered various block-structured correlation matrices such as common correlation model, grouped correlation model and grouped variables model, and the priors on R were restricted to the set $\mathcal{R}^k$ of positive-definite correlation matrices. For example, in the common correlation model, the prior on R is

$$p(R|\mu, \sigma^2) = C(\mu, \sigma^2) \prod_{i<j} \exp\{-(r_{ij} - \mu)^2/2\sigma^2\} I\{R \in \mathcal{R}^k\},$$

where $C(\mu, \sigma^2)$ is the normalizing constant with normal and inverse gamma distributions for μ and σ^2, respectively. Due to the lack of conjugacy in the prior model, they used Metropolis-Hastings (MH) algorithm for posterior inference. However, presence of the indicator function $I\{R \in \mathcal{R}^k\}$ makes the computation very expensive. In fact, the proposal density for a particular element r_{ij} in every iteration of the MH step must be restricted to an interval $[l_{ij}, u_{ij}]$ where these limits are quadratic functions of the rest of the elements of R, see (Barnard et al., 2000, p. 1305) for more details and explicit formulae for l_{ij}, u_{ij}.

The challenges of dealing with the positive-definiteness and unit diagonal constraints of a correlation matrix have led to two unconstrained reparameterizations of R for ordered or longitudinal data. The reparameterization in Joe (2006) is based on the notion of partial autocorrelation function (PAC), or the correlation coefficient between two variables given the intermediate variables. The key advantages of using PACs are avoiding the positive-definiteness constraint and providing interpretable parameters. Interestingly, PACs had appeared earlier in Kurowicka and Cooke (2003), Kurowicka and Cooke (2006) in the context of vine graphical models. It has proved useful for elicitation of priors for Bayesian correlation estimation as in Daniels and Pourahmadi (2009) who constructed a flexible prior on R using independent shifted Beta priors on

the PACs, and Wang and Daniels (2014) developed the underlying regression models involving PACs and a triangular prior which shifts the prior weight to a more intuitive choice. Moreover, Gaskins et al. (2014) proposed shrinkage and selection priors on the PACs aiming to estimate the correlation matrix. In all these developments, instead of setting the full partial correlations or the entries in R^{-1} to zero to incorporate sparsity, the goal has been to set some PACs to zero to encourage parsimony in modeling R. Although the angular parameterization has been around much longer (Pinheiro and Bates, 1996; Rapisarda et al., 2007), it has not been used as yet in eliciting priors for correlation matrices.

The aim of this article is to study and deal with some of the computational challenges in Bayesian estimation of correlation matrices by using its Cholesky decomposition, (Pinheiro and Bates, 1996; Rapisarda et al., 2007) and the ensuing angles (hyperspherical coordinates) as the new parameters varying freely in the range $[0, \pi)$. This enables us to deal effectively with the positive-definiteness constraint, resulting in faster computation of the posteriors for our proposed selection and shrinkage priors. For longitudinal data, we show that these angles are directly related to the as yet dormant notion of semi-partial correlations (SPCs). More precisely, we identify the angles as the inverse cosine of $\rho_{ji:1,2,\ldots,j-1}$ between the variables y_i and y_j $(i > j)$ conditioned on the previous variables $y_1, y_2, \ldots, y_{j-1}$, see Huber (1981), Eaves and Chang (1992), and Madar (2015). We propose natural and appealing shrinkage and selection priors on these angles and show that their performance is similar to or in some cases better than the shrinkage and selection priors on the partial autocorrelation (Gaskins et al., 2014).

The rest of this article is organized as follows. In Section 2, we discuss some preliminaries about the angular parameterization and connect the angles to the notion of semi-partial correlations. Section 3 introduces our proposed shrinkage and selection priors on the angles and develops Bayesian estimation of correlation matrices. Sampling scheme under our proposed priors is discussed in Section 4. Section 5 compares the performance of our priors to those based on the PACs through simulations. In Section 6, priors are compared on a data from a smoking cessation clinical trial. Finally Section 7 concludes the article with discussion.

2 Reparameterization of R by angles

This section describes a connection between the well-known hyperspherical coordinates (angles) of the Cholesky factor of a correlation matrix $R = (r_{ij})$ and the less familiar semi-partial correlation coefficients $\rho_{ji:1,2,\ldots,j-1}$ between the variables y_i and y_j $(i > j)$ conditioned on the previous variables $y_1, y_2, \ldots, y_{j-1}$, see Huber (1981), Eaves and Chang (1992), and Madar (2015).

For a $k \times k$ correlation matrix R with 1's on the diagonal, its Cholesky decomposition is given by $R = BB^\top$ where the Cholesky factor B is a lower triangular matrix. Since the rows of B are vectors of unit-length, it turns out that they admit the following representation involving trigonometric functions of some angles (Pinheiro and Bates,

1996; Rapisarda et al., 2007):

$$
B = \begin{bmatrix}
1 & 0 & 0 & 0 & \dots & 0 \\
c_{21} & s_{21} & 0 & 0 & \dots & 0 \\
c_{31} & c_{32}s_{31} & s_{32}s_{31} & 0 & \dots & 0 \\
c_{41} & c_{42}s_{41} & c_{43}s_{42}s_{41} & \prod_{j=1}^{3} s_{4j} & \dots & 0 \\
\vdots & \vdots & \vdots & & \ddots & \vdots \\
c_{k1} & c_{k2}s_{k1} & c_{k3}s_{k2}s_{k1} & c_{k4}\prod_{j=1}^{3} s_{kj} & \dots & \prod_{j=1}^{k-1} s_{kj}
\end{bmatrix}
\tag{2.1}
$$

with $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$, where the angles θ_{ij}'s are measured in radians, $1 \leq j < i \leq k$. Restricting $\theta_{ij} \in [0, \pi)$ makes the diagonal entries of B non-negative, and hence B is unique to which we can associate a $(k-1) \times (k-1)$ lower triangular matrix Θ with $k(k-1)/2$ angles:

$$
\Theta = \begin{bmatrix}
\theta_{21} & 0 & 0 & \dots & 0 \\
\theta_{31} & \theta_{32} & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\theta_{k1} & \theta_{k2} & \theta_{k3} & \dots & \theta_{k,k-1}
\end{bmatrix}.
$$

Note that the (i,j)-th element of Θ is denoted by $\theta_{i+1,j}$ so that θ_{ij} corresponds to the (i,j)-th element of R, we refer to Θ as the **angular matrix** associated to R. For further details, properties and applications of these angles, see Creal et al. (2011), Zhang et al. (2015) and Tsay and Pourahmadi (2017).

Given a correlation matrix R (symmetric and positive definite) and its Cholesky decomposition $R = BB^\top$ with entries b_{ij}, matching entries of both sides it follows that

$$
b_{11} = 1, \quad b_{i1} = r_{i1}, \quad i = 2, \dots, k.
\tag{2.2}
$$

Thus, the entries in the first columns of B and R are the same. The θ_{ij}s, the entries of Θ are computed recursively via

$$
\theta_{i1} = \arccos(b_{i1}) = \arccos(r_{i1}), \text{ for } i = 2, 3, \dots, k,
\tag{2.3}
$$

$$
\theta_{ij} = \arccos\left(\frac{b_{ij}}{\prod_{l=1}^{j-1} \sin(\theta_{il})}\right), \text{ for } 2 \leq j < i \leq k.
$$

Also, given a matrix Θ with entries $\theta_{ij} \in [0, \pi)$, construct the lower triangular matrix $B = (b_{ij})_{i,j \in \{1,2,\dots,k\}}$ via

$$
b_{ij} = \begin{cases}
\prod_{l=1}^{i-1} \sin(\theta_{il}), \text{ for } i = j, \\
\cos(\theta_{ij}) \prod_{l=1}^{j-1} \sin(\theta_{il}), \text{ for } 1 \leq j < i \leq k,
\end{cases}
\tag{2.4}
$$

and then the correlation matrix $R = BB^\top$.

We illustrate the nature of nonlinear relationship between the angles and correlations for a general correlation matrix using the following two simple examples ((a) and (b)) of increasing dimensions:

(a) For $k = 2$, there is only one angle θ_{21} and

$$B = \begin{bmatrix} 1 & 0 \\ \cos \theta_{21} & \sin \theta_{21} \end{bmatrix},$$

and from (2.2) we obtain the simple relation $r_{21} = \cos \theta_{21}$. Then, the statistical meaning of θ_{21} as the inverse cosine of r_{21} is fairly clear.

(b) For $k = 3$, we have

$$\Theta = \begin{bmatrix} \theta_{21} & 0 \\ \theta_{31} & \theta_{32} \end{bmatrix}.$$

The relationship between r_{ij}'s and θ_{ij}'s are

$$\theta_{21} = \arccos(r_{21}), \theta_{31} = \arccos(r_{31}), \theta_{32} = \arccos\left(\frac{r_{32} - \cos(\theta_{21})\cos(\theta_{31})}{\sin(\theta_{21})\sin(\theta_{31})}\right), \quad (2.5)$$

so that while the two angles in the first column are tied to the individual marginal correlations as in (2.2), this is not the case for θ_{32} in the second column. In fact, the situation gets more complicated for larger k's. In general, the entries of the first column of Θ are just the inverse cosine of the respective entries of the first column of R, but as one moves towards its last column the expression for θ_{ij} becomes more complicated and hence less interpretable as a function of the entries of R.

However, in case of some special correlation matrices, the angles are nicely structured. We present forms of Θ for AR(1) and banded correlation structure which are of special interest in longitudinal data in the next two examples and a block common correlation in example (e).

(c) **AR(1) correlation matrix:** For an AR(1) correlation specified by r, corresponding Θ is characterized by a single angle, $\theta_{21} = \arccos(r)$. This is called *pivotal angle* and other angles are function of it, called *implied angle*(Tsay and Pourahmadi, 2017). Moreover, it is instructive to note that the corresponding angle approaches to $\pi/2$ as the lag increases which is stated in the following proposition.

Proposition 1. *For a k-dimensional AR(1) correlation matrix R with $r_{ij} = r^{|i-j|}$, the angle $\theta_{ij} \to \pi/2$ $(i > j)$ as $(i - j) \to \infty$.*

Proof. We use induction on the column index j by noting that $(i - m) \to \infty$ for $1 \le m \le j$.

For $j = 1$, $r_{i1} = r^{(i-1)} = b_{i1}$ which goes to 0 under the assumption. Therefore, $\theta_{i1} \to \pi/2$. Suppose that the assertion holds for any integer $m < j$, i.e. $\theta_{im} \to \pi/2$. By construction of B, this implies $b_{im} \to 0$. Since $r_{ij} = b_{ij}b_{jj} + \sum_{l=1}^{j-1} b_{il}b_{jl}$ and diagonal entries of B are positive, it follows that $b_{ij} \to 0$ as $(i - j) \to \infty$, which further implies $\theta_{ij} \to \pi/2$. □

(d) **Banded correlation matrix:** A correlation matrix $R = (r_{ij})$ is called λ-banded if

$$r_{ij} = \begin{cases} r_{ij} \neq 0 & \text{if } |i - j| \leq \lambda, \\ 0 & \text{otherwise,} \end{cases}$$

where λ is an integer between 1 and k. Corresponding Θ is also banded by the following proposition.

Proposition 2. *A k-dimensional correlation matrix R is λ-banded if and only if the corresponding $\Theta = (\theta_{ij})$ satisfies*

$$\theta_{ij} = \begin{cases} \theta_{ij} \neq \pi/2 & \text{if } |i - j| \leq \lambda, \\ \pi/2 & \text{otherwise.} \end{cases}$$

Proof. The proof follows using the Cholesky decomposition and relating it to Θ. Consider $R = BB^\top$ where $B = ((b_{ij}))$ as given in (1). Suppose Θ satisfies $\theta_{ij} = \pi/2$ for $|i - j| > \lambda$. Then from (1), it follows that $b_{ij} = 0$ for $|i - j| > \lambda$.

Thus, for $|i - j| > \lambda$, we have

$$r_{ij} = \sum_{l=1}^{k} b_{il} b_{jl} = \sum_{l=1}^{j} b_{il} b_{jl} \quad \text{(assuming } i > j\text{)}.$$

Since $b_{ij} = 0$ for $|i - j| > \lambda$ and l runs over $1, 2, \ldots, j$ and $i > j$, i.e. contribution of b_{jl} for each summand is 0 which implies $r_{ij} = 0$ for $|i - j| > \lambda$.

For the only if part, $r_{ij} = 0$ for $|i - j| > \lambda$ where $1 \leq \lambda < k$ and the proof follows by induction. Note

$$r_{i1} = b_{i1} b_{11} \quad \text{(since } B \text{ is lower triangular matrix).}$$

By the construction of B, the diagonal entries are positive. Thus $r_{i1} = 0$ implies $b_{i1} = 0$ which in turn implies $\theta_{i1} = \pi/2$. For the induction step, assuming $j < i$ and $i - j > \lambda$, consider

$$r_{ij} = \sum_{l=1}^{j} b_{il} b_{jl} = b_{ij} b_{jj} + \sum_{l=1}^{j-1} b_{il} b_{jl}.$$

From the induction hypothesis, $\theta_{jl} = \pi/2$ for $l \leq j - 1$ implies the second summand is 0. Thus, we must have $b_{ij} = 0$ which implies $\theta_{ij} = \pi/2$ and this completes the proof. $\square$

(e) **Block common correlation matrices:** As a generalization of a compound symmetric (exchangeable) correlation matrix, consider a block common correlation matrix which is a blocked-matrix where the correlations within each block are equal and different across blocks. Such matrices arise in many applications due presence of common (latent) factors in different regions (aggregation of carbon dioxide sequestration storage assessment units Blondes et al. (2013)) and stock returns of different companies within the same industry (Liechty et al., 2004; Tsay and Pourahmadi, 2017). As an

illustration, we consider the following 6×6 correlation matrix,

$$R = \begin{pmatrix} 1 & r_1 & r_2 & r_2 & r_3 & r_3 \\ & 1 & r_2 & r_2 & r_3 & r_3 \\ & & 1 & r_4 & r_5 & r_5 \\ & & & 1 & r_5 & r_5 \\ & & & & 1 & r_6 \\ & & & & & 1 \end{pmatrix},$$

with 6 distinct correlations r_i, $i = 1, 2, \ldots, 6$, which is much smaller than 15, the number of distinct entries of a generic correlation matrix of this size. The corresponding matrix of angles Θ, is completely determined by six *pivotal angles* denoted by $(\theta_{21}, \theta_{31}, \theta_{51}, \theta_{43}, \theta_{53}, \theta_{65})$ (Tsay and Pourahmadi, 2017) where their subscripts indicate their locations in the partitioned matrix Θ.

In general, for a $k \times k$ correlation matrix R, if there is d common correlation blocks, then the pivotal angles consist of d angles in the range $[0, \pi)$, say $\theta_{pivotal} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$. The other angles in Θ matrix, called *implied angles*, are functions of pivotal angles and can be obtained using an algorithm in Tsay and Pourahmadi (2017). When the blocks are known, it is simple to determine the positions of pivotal angles. One can use prior only on the *pivotal angles* to shrink them to different target values, and this will reduce the dimension of the parameter space to d and finally use the algorithm in (Tsay and Pourahmadi, 2017, p. 11) to estimate the entire correlation matrix. It is worthwhile to note that AR(1) and banded correlation matrices often arise in the context of longitudinal data. Therefore, the characterization of Θ as discussed in examples (c) and (d) is useful to elicit prior in Bayesian estimation of such matrices.

2.1 The angles and semi-partial correlations

Statistical interpretation and plausible meaning of the angles as the new parameters of a correlation matrix are of interest when eliciting priors. This task is complicated by the nonlinearity of the relationships between the correlations and angles as seen in (2.5).

Here, we use a relatively dormant formula for b_{ij} stated without proof in (Cooke et al., 2011, Chapter 3) and identify the angles as the inverse cosine of the semi-partial correlations (SPCs) $\rho_{ji:1,2,\ldots,j-1}$ between the variables y_i and y_j $(i > j)$ conditioned on the previous variables, see Huber (1981), Eaves and Chang (1992), and Madar (2015). Surprisingly, the simplicity of the relations between the angles and SPCs is reminiscent of the relations in (2.2) between the entries of the first columns of Θ and R.

Theorem 1. *Let R be a general $k \times k$ positive-definite correlation matrix with the Cholesky decomposition $R = BB^\top$ where the Cholesky factor B is a lower triangular matrix. Then,*

(a) the entries of $B = (b_{ij})$ can be expressed in terms of the semi-partial correlations (SPCs) as

$$b_{i1} = r_{i1}, \quad b_{ii} = \sqrt{1 - \sum_{u=1}^{i-1} b_{iu}^2} \quad for \ i = 2, 3, \ldots, k, \tag{2.6}$$

$$b_{ij} = \rho_{ji:1,2,\ldots,j-1} \prod_{u=1}^{j-1} \sqrt{1 - \rho_{ui:1,2,\ldots,u-1}^2} \; for \; 2 \leq j < i \leq k. \qquad (2.7)$$

(b) the angles θ_{ij}'s are precisely the inverse cosine of the SPCs:

$$\rho_{ji:1,2,\ldots,j-1} = cos(\theta_{ij}) \; for \; 1 \leq j < i \leq k. \qquad (2.8)$$

The proof is given in Appendix A (Ghosh et al., 2020).

2.2 Distributions of the angles

Cholesky decomposition of a correlation matrix, and hence the concepts of the angles and semi-partial correlations depend on ordering or labeling the variables in R. Next, one may assign distributions to the angles so that the distribution of R is a power of its determinant and hence invariant to permutations of its rows and columns (Pourahmadi and Wang, 2015, Theorem 1).

Theorem 2. *For a k-dimensional random correlation matrix R with the corresponding matrix of angles Θ, let the random variables in the jth column of Θ be independent and identically distributed as*

$$\theta_{ij} \sim p_j(\theta) \propto (sin \, \theta)^{2\alpha+k-j} \; for \; j = 1, \cdots, k, \quad i = j+1, j+2, \ldots, k, \qquad (2.9)$$

where α is a constant, $\theta \in [0, \pi)$. Then

(a) the joint distribution of R is given by

$$p(R) = c_k(\alpha)[det(R)]^\alpha, \quad c_k(\alpha) = \prod_{j=1}^{k-1} \Big(\frac{\Gamma(\frac{2\alpha+j}{2}+1)}{\sqrt{\pi}\Gamma(\frac{2\alpha+j+1}{2})} \Big)^j, \qquad (2.10)$$

where $c_k(\alpha)$ is the normalizing constant.

(b) The marginal density of each r_{ij}, $1 \leq j < i \leq k$, of the correlation matrix R is proportional to $(1 - r_{ij}^2)^{\alpha+\frac{k}{2}-1}$, i.e., a shifted $Beta(\alpha + k/2, \alpha + k/2)$ distribution in $[-1, 1]$.

(c) The distribution is symmetric about $\pi/2$, hence its mean and median are equal to $\pi/2$.

Although the proof appears in (Pourahmadi and Wang, 2015, Theorem 1), for reader's convenience, we give an outline in Appendix B.

One can identify (2.9) as the distribution of axial data. In addition to non-negativity and density integrating to one, such a distribution must satisfy $p(\theta) = p(\theta+\pi)$ for all θ in $[0, \pi)$ (Arnold and SenGupta, 2006) which holds for (2.9) by noting that $p_j(\theta) = |p_j(\theta)|$ and $|sin(\theta + \pi)| = |sin(\theta)|$ for θ in $[0, \pi)$.

It turns out that these distributions on the angles reduce to the joint uniform prior of Barnard et al. (2000) on a correlation matrix for specific value of α.

Recall that the *joint uniform prior* assigns a uniform distribution to the set of all valid $k \times k$ correlation matrices

$$p_J(R) \propto 1, R \in \mathcal{R}^k. \tag{2.11}$$

Indeed, $\alpha = 0$ in (2.10) leads to p_J. As such this prior is noninformative and not suitable in longitudinal data analysis, since higher lag (auto)correlations tend to zero faster than those with smaller lags.

3   Prior specifications on angles

We propose shrinkage and selection priors on angles for ordered data guided by the fact that two variables far apart have correlation decaying to zero. Therefore, it is natural to expect that the semi-partial correlation between two variables y_i & y_j $(i > j)$ in the random vector Y given the preceding variables $y_1, y_2, \ldots, y_{j-1}$ decays to zero as the lag $(i - j)$ increases. In terms of angles, this essentially means that the corresponding θ_{ij} goes to $\pi/2$, since θ_{ij} is related to the corresponding semi-partial correlation only through the cosine function (2.8). A key role behind construction of our shrinkage prior is played by the (modified) shifted beta distribution with parameters α and β, denoted by SBeta(α, β), in $[-1, 1]$:

$$p(\rho) = \frac{(1+\rho)^{\alpha-1}(1-\rho)^{\beta-1}}{B(\alpha, \beta)2^{\alpha+\beta-1}} \ , \qquad \text{for } -1 \le y \le 1. \tag{3.1}$$

With $\alpha = \beta$, $p(\rho)$ symmetric around 0 having mean zero and variance $1/(2\alpha + 1)$.

We note that in Bayesian statistics, spike and slab priors have also been used in practice as a selection prior with a spike at a target value, say $\pi/2$. There is a vast literature on selection priors (Mitchell and Beauchamp, 1988; Ishwaran and Rao, 2005).

3.1   Selection prior

A way to motivate the formation of our selection prior is to note that when R is an identity matrix, all the entries of Θ are $\pi/2$. Thus, forming a selection prior as a mixture of a Dirac delta with mass at $\pi/2$ and a continuous density having support in $[0, \pi)$, is capable of selecting or centering the angles at $\pi/2$. In terms of the SPCs, this amounts to encouraging the semi-partial correlation between y_i and y_j given $y_1, y_2, \ldots, y_{j-1}$ to be centered at 0.

Similar to PAC framework where selection prior (Gaskins et al., 2014) on π_{ij} is constructed using a mixture of point mass at zero and a SBeta distribution in $[-1, 1]$, our selection prior on the angles, denoted by $p_{\theta,SP}$, assumes independent mixture distributions for individual θ_{ij}'s by

$$p_{\theta_{ij},SP}(\theta) \propto (1 - \eta_{ij})\delta_{\pi/2}(\theta) + \eta_{ij}\sin(\theta), \text{ where } \theta \in [0, \pi), \tag{3.2}$$

where, $\eta_{ij} = Pr(\theta_{ij} \neq \pi/2)$ for $1 \le i < j \le k$ and $\delta_{\pi/2}$ denotes a Dirac delta with mass at $\pi/2$. To make such priors more suitable for longitudinal data, we further parameterize

$\eta_{ij} = \eta_0|j-i|^{-\gamma}$ so that as the lag $|j-i|$ increases, the prior in (3.2) puts more weight at $\pi/2$. Since the angle θ_{ij} is related to the partial correlation $\rho_{ji:1,2,...,j-1}$ in (7) through $\cos(\theta_{ij})$, this implies that for variables which are far apart or having greater lag $|i-j|$, the corresponding θ_{ij}'s are closer to $\pi/2$. We further assume a Unif$(0,1)$ distribution for the hyper-parameter η_0 and a Gamma(a,a) distribution for the hyper-parameter γ so that γ has prior mean 1. In our simulation study, we choose $a = 5$ to make our results comparable to those in Gaskins et al. (2014).

3.2   Shrinkage prior

In Bayesian covariance estimation, shrinkage priors have been used to shrink the posterior estimate towards specific structures. For example, Liechty et al. (2004) considered priors to shrink the correlation matrix to certain group-structured targets, and Wang and Pillai (2013) considered scale mixture of uniform distributions to construct shrinkage priors for covariance matrix estimation.

The shrinkage prior in Gaskins et al. (2014) shrinks the PAC (π_{ij})'s towards 0 using SBeta$(\alpha_{ij}, \alpha_{ij})$ which has mean zero and variance $\xi_{ij} = 1/(2\alpha_{ij} + 1)$.

In the interest of parsimony, they parameterize $\xi_{ij} = \xi_0|i-j|^{-\gamma}$, for $\xi_0 \in (0,1), \gamma > 0$, so that for longitudinal data higher-lag terms are shrunk to 0 more aggressively.

To define an analogue of the above shrinkage prior on angles we exploit the interpretation of angles in (2.8) and assume that the semi-partial correlation $\rho_{ji:i,2,...,j-1}$ follows a SBeta$(\alpha_{ij}, \alpha_{ij})$ distribution. Using change of variable, we propose our shrinkage prior $p_{\theta_{ij},SH}$ on θ_{ij} by,

$$p_{\theta_{ij},SH}(\theta) \propto (\sin\theta)^{2\alpha_{ij}-1}, \quad \theta \in [0,\pi). \tag{3.3}$$

Note that the form (3.3) is very similar to (2.9) except the exponent. The mean of the distribution is $\pi/2$ [Theorem 2(c)] but the variance has no closed form expression. Using delta method, the variance can be approximated by $\eta_{ij} = 1/(2\alpha_{ij} + 1)$

As in selection prior, we further parameterize $\eta_{ij} = \eta_0|i-j|^{-\gamma}$ to make it dependent on lag and assume a Unif$(0,1)$ prior on η_0 and a Gamma(a,a) prior on γ.

4   Sampling from posterior distribution

We assume throughout that the data $y_1, y_2, \cdots, y_n$ follow a multivariate normal distribution of dimension k with mean zero vector and covariance R. Restricting attention to correlation matrices is natural, for example, in the analysis of multivariate probit model to circumvent the issue of identifiability (Chib and Greenberg, 1998). Denoting $Y = [y_1, y_2, \cdots, y_n]$ the likelihood function parameterized by Θ is given by

$$L(Y|\Theta) \propto \det(T(\Theta))^{-n/2}\exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}y_i^\top(T(\Theta))^{-1}y_i\Big\},$$

where T denotes the transformation from R to Θ.

Updating θ_{ij}:

Denote by $\Theta_{[-ij]}$, the Θ matrix after dropping its (i,j)-th element and $p_{\theta_{ij}}(\theta)$ the prior for θ_{ij}. For convenience, we often drop the subscripts SP and SH from $p_{\theta_{ij}}$ when the context is easily understood. The posterior distribution of θ_{ij} given others is:

$$p(\theta_{ij}|\Theta_{[-ij]}, \gamma, \eta_0, Y) \propto L(Y|\theta_{ij}, \Theta_{[-ij]})p_{\theta_{ij}}(\theta). \tag{4.1}$$

Note that the involvement of θ_{ij} in the likelihood makes the posterior non-conjugate. To sample from (4.1), we incorporate auxiliary variable z_{ij} (Damien et al., 1999; Neal et al., 2003; Gaskins et al., 2014), write

$$p(\theta_{ij}|\Theta_{[-ij]}, Y) = \int_0^\infty I\{z_{ij} < L(Y|\theta_{ij}, \Theta_{[-ij]})p_{\theta_{ij}}(\theta_{ij})\}dz_{ij} \tag{4.2}$$

and sample θ_{ij} in two steps. While the first step is common for both selection and shrinkage prior, the notable difference appears in second step owing to the presence of Dirac delta function in selection prior (3.2).

> Step 1. Sample z_{ij}^* from $\text{Unif}(0, L(Y|\theta_{ij}, \Theta_{[-ij]}))$ for an initial value of θ_{ij}.
>
> Step 2. For shrinkage prior (3.3), sample θ_{ij} uniformly from the set $\{\theta : z_{ij}^* < L(Y|\theta_{ij}, \Theta_{[-ij]})p_{\theta_{ij}}\}$.
>
> In case of selection prior (3.2), sample θ_{ij} from $p_{\theta_{ij}, SP}(\theta)$ restricted to the constrained set $\mathcal{C} = \{\theta : z_{ij}^* < L(Y|\theta_{ij}, \Theta_{[-ij]})\}$. To sample from $p_{\theta_{ij}, SP}(\theta)$ restricted to $\mathcal{C}$, let $F(\theta) = P(\theta_{ij} \leq \theta)$ be the cumulative distribution function of $p_{\theta_{ij}, SP}$. The expression of $F(\theta)$ is available in closed form. Next draw U uniformly over the set $F(\mathcal{C})$ and update θ_{ij} by $F^{-1}(U) = \inf\{\theta : F(\theta) \geq U\}$.

Updating η_0, γ:

For updating the parameters η_0, γ in selection prior, we incorporate dummy variable $\vartheta_{ij} = I\{\theta_{ij} \neq 0\} \sim Ber(\eta_{ij})$ so that $P(\theta_{ij} \neq 0) = \eta_{ij}$. This makes the distribution of η_0, γ dependent on Θ only through the variables ϑ_{ij}. Next we use two slice samplers to update them and similar is the case for shrinkage prior.

5 Simulations

5.1 Comparing priors on the angles and PACs

We perform a number of simulation studies to assess the performance of our selection and shrinkage priors on the angles relative to the selection and shrinkage priors of Gaskins et al. (2014) on partial autocorrelations. Since the selection prior performed better than the shrinkage prior in their simulation study, here we focus only on the selection prior and follow their simulation set-up as much as possible.

The frequentist risks of the posterior estimates are evaluated by averaging the loss over 60 simulation runs for the following two loss functions: The Kullback-Leibler loss function $L_1(R, \hat{R}) = \text{tr}(\hat{R}^{-1}R) - \log|\hat{R}^{-1}R| - k$, which is zero when $\hat{R} = R$. The second

loss function for estimating Θ is defined by $L_2(\Theta, \hat{\Theta}) = ||\hat{\Theta} - \Theta||_F^2 = \sum_{i<j}(\theta_{ij} - \hat{\theta}_{ij})^2$, where $\hat{R}$ and $\hat{\Theta}$ denote the posterior estimates.

We consider 4 different 6×6 correlation matrices: R_A (the identity matrix whose Θ a lower-triangular matrix of order 5 having all entries $\pi/2$), R_B (AR(1) matrix with correlation 0.7), where Θ_B is

$$\Theta_B = \begin{bmatrix} 0.795 & 0 & 0 & 0 & 0 \\ 1.059 & 0.960 & 0 & 0 & 0 \\ 1.221 & 1.189 & 0.960 & 0 & 0 \\ 1.328 & 1.316 & 1.189 & 0.960 & 0 \\ 1.402 & 1.396 & 1.316 & 1.189 & 0.960 \end{bmatrix},$$

and R_C, R_D constructed from the following Θ matrices:

$$\Theta_C = \begin{bmatrix} \pi/4(0.707) & 0.866 & 0 & 0 & 0 \\ \pi/6 & \pi/4(0.862) & 0.612 & 0 & 0 \\ \pi/2 & \pi/6 & \pi/4(0.431) & 0.306 & 0 \\ \pi/2 & \pi/2 & \pi/6 & \pi/4(0.431) & 0.306 \\ \pi/2 & \pi/2 & \pi/2 & \pi/6 & \pi/6(0.459) \end{bmatrix},$$

$$\Theta_D = \begin{bmatrix} \pi/4(0.707) & 0.866 & 0.707 & 0.866 & 0.809 \\ \pi/6 & \pi/4(0.862) & 0.933 & 0.789 & 0.866 \\ \pi/4 & \pi/6 & \pi/2(0.829) & 0.838 & 0.848 \\ \pi/6 & \pi/3 & \pi/2 & \pi/2(0.765) & 0.827 \\ \pi/5 & \pi/4 & \pi/2 & \pi/2 & \pi/2(0.805) \end{bmatrix}.$$

In the above notation, off-diagonal elements of the corresponding R are given in upper triangular matrix.

It can be seen that Θ_C leads to a banded correlation matrix and the entries in the rows of Θ_D decay to $\pi/2$.

For each of the 4 correlation matrices, we simulate 60 data-sets of sample sizes $n = 20, 200$ from a multivariate normal distribution having mean zero and covariance matrix equals to the chosen correlation matrix. For comparison of the risks, our competitor is $p_{\pi;SP}$ which performed the best in Gaskins et al. (2014). To compute the posterior for $p_{\pi;SP}$, we also obtain Π for each R above. We run an Markov chain Monte Carlo (MCMC) for 5000 iterations with a burn-in 1000 and retain every tenth iteration providing 500 outputs for each data-set. At each iteration of Θ, we retain the corresponding correlation matrix $R(\Theta)$ and at each iteration of Π, in addition to $R(\Pi)$, $\Theta(\Pi)$ or $\Theta(R(\Pi))$ is retained.

The posterior estimates of Θ, $R(\Theta)$, $R(\Pi)$ and $\Theta(\Pi)$ are obtained by taking the average of these samples after burn-in from each iteration. For each case, we gauged the performance by the risk estimates with respect to the two loss functions discussed earlier by taking average of these loss functions over 60 replications of the simulated data.

The results are summarized in Table 1, where we note that for the identity matrix (R_A) our selection prior outperforms all its competitors, but our shrinkage prior is

n	R	Loss	$p_{\pi;SP}$	$p_{\theta;SH}$	$p_{\theta;SP}$	p_J
20	R_A	$L_1(\hat{R}, R)$	0.025	0.032	0.018	0.967
		$L_2(\hat{\Theta}, \Theta)$	0.6027	0.595	0.1001	1.002
200	R_A	$L_1(\hat{R}, R)$	0.0014	0.0015	0.0008	0.7071
		$L_2(\hat{\Theta}, \Theta)$	0.0422	0.4247	0.0226	0.9961
20	R_B	$L_1(\hat{R}, R)$	0.34	0.45	0.35	1.065
		$L_2(\hat{\Theta}, \Theta)$	0.6113	0.7655	0.6133	1.0046
200	R_B	$L_1(\hat{R}, R)$	0.027	0.0355	0.028	0.9574
		$L_2(\hat{\Theta}, \Theta)$	0.1501	0.3553	0.1567	0.7654
20	R_C	$L_1(\hat{R}, R)$	2.095	0.5665	0.3500	1.985
		$L_2(\hat{\Theta}, \Theta)$	0.9532	0.3965	0.1905	0.9643
200	R_C	$L_1(\hat{R}, R)$	1.665	0.1025	0.0765	1.7855
		$L_2(\hat{\Theta}, \Theta)$	0.7562	0.2256	0.0985	0.7557
20	R_D	$L_1(\hat{R}, R)$	2.0035	0.8565	0.5567	2.6895
		$L_2(\hat{\Theta}, \Theta)$	0.9779	0.6312	0.4608	1.2575
200	R_D	$L_1(\hat{R}, R)$	1.5001	0.2542	0.1959	2.0698
		$L_2(\hat{\Theta}, \Theta)$	0.5672	0.2650	0.1065	0.9989

Table 1: Risks for our selection prior ($p_{\theta;SP}$), shrinkage prior ($p_{\theta;SH}$) and the selection prior in Gaskins et al. (2014).

outperformed by the selection prior in Gaskins et al. (2014). For the AR(1) (R_B), our selection prior and the selection prior of Gaskins et al. (2014) are comparable. For banded R_C and R_D, our selection prior and shrinkage prior are comparable and perform better than $p_{\pi;SP}$. The joint uniform prior performs poorly in all of these cases.

In summary, our selection prior and shrinkage prior show advantage over based on certain scenarios.

5.2   Computational advantages of angle parameterization

The computational challenges of using constrained priors like the joint uniform prior $p_J(R)$ are well-known, other notable examples are the common correlation priors in Liechty et al. (2004), priors for sparse R^{-1} in Wong et al., 2003; Pitt et al., 2006; Carter et al., 2011, which place a flat prior on the non-zero components for a given pattern of zeros. These methods usually require computing the normalizing constants related to volumes of certain subsets of $\mathcal{R}^k$ corresponding to patterns of zeros, and where the prior and posterior densities are supported on constrained sets. Due to the presence of the indicator function of $\mathcal{R}^k$ in the prior and posterior, in the Metropolis-Hastings scheme, the proposal density for updating r_{ij} has to be restricted to an interval $[l_{ij}, u_{ij}]$ where these bounds are functions of the rest of the entries of $R^{\pm 1}$ (Barnard et al., 2000;

Liechty et al., 2004). Of course, unconstrained parameterization resolves the tedious task of computing the normalizing constant in every update of the MCMC algorithm and consequently posterior computation is faster.

Next, we compare the time complexity of implementing the MCMC algorithm for the constrained prior $p_J(R)$ on the space of valid correlation matrices $\mathcal{R}^k$, and its two unconstrained reparameterizations on the spaces of angles Θ and partial autocorrelations $\mathbf{\Pi}$. The prior on the angles:

$$p_{\theta J}(\Theta) \propto \prod_{i>j}(\sin \theta_{ij})^{k-j}, \tag{5.1}$$

is obtained from (2.9) for $\alpha = 0$, and the prior on PACs is

$$p_{\pi J}(\mathbf{\Pi}) \propto \prod_{i<j}(1 - \pi_{ij}^2)^{1+[(k-1)-(j-i)]/2}, \tag{5.2}$$

where $\pi_{ij} \in [-1, 1]$, for more details see Gaskins et al. (2014).

We consider three different settings of (n, k), namely $(50, 5)$, $(100, 10)$ and $(500, 15)$ and simulate a sample of size n from a k-dimensional normal distribution having mean 0 and covariance matrices set to Identity, AR(1) with correlation 0.4 and a general correlation matrix, respectively. For posterior sampling, we use the slice sampling techniques similar to Section 4 for Θ and Gaskins et al. (2014) for Π with necessary adjustments needed the priors (5.1) and (5.2) and for the constrained case we use Metropolis-Hastings algorithm to sample individual correlations r_{ij} from the restricted set determined by rest of the entires of R as in (Barnard et al., 2000; Liechty et al., 2004).

In Figure 1, we present run times (in log scale of seconds) for 2000 independent MCMC samples satisfying "effectiveSize" function in R which gives effective number of MCMC samplers adjusting the autocorrelation. As expected the unconstrained priors outperform constrained method significantly in any dimension with respect to the execution-time. The simulations were run on a 2.6 GHz Intel Core i5 processor. The numerical results above may not be surprising by noting that the computational complexity of simulating a posterior of R based on priors on angles or generating general random correlation matrices (Pourahmadi and Wang, 2015) is $O(k^3)$ compared to $O(k^4)$ of the Joe (2006) proposal based on partial correlations, and $O(k^3)$ of the Lewandowski et al. (2009) method using the partial correlations defined on C-vines, respectively.

6   Data analysis

We analyze a data set (Gaskins et al., 2014) simulated based on first Commit to Quit (CTQ I) study of Marcus et al. (1999), a clinical trial designed to encourage women to stop smoking. The aim of the study was how exercise is effective to increase quit rate, as weight gain seems to be an influencing factor in a smoking cessation program. Providing an educational intervention of equal time for the control group, the study spans 12 weeks and the patients were encouraged to quit smoking at week 5.

The data is provided in the form of a 281×9 matrix, where rows correspond to patients and columns 2–9 correspond to weeks and first column corresponds to treatment
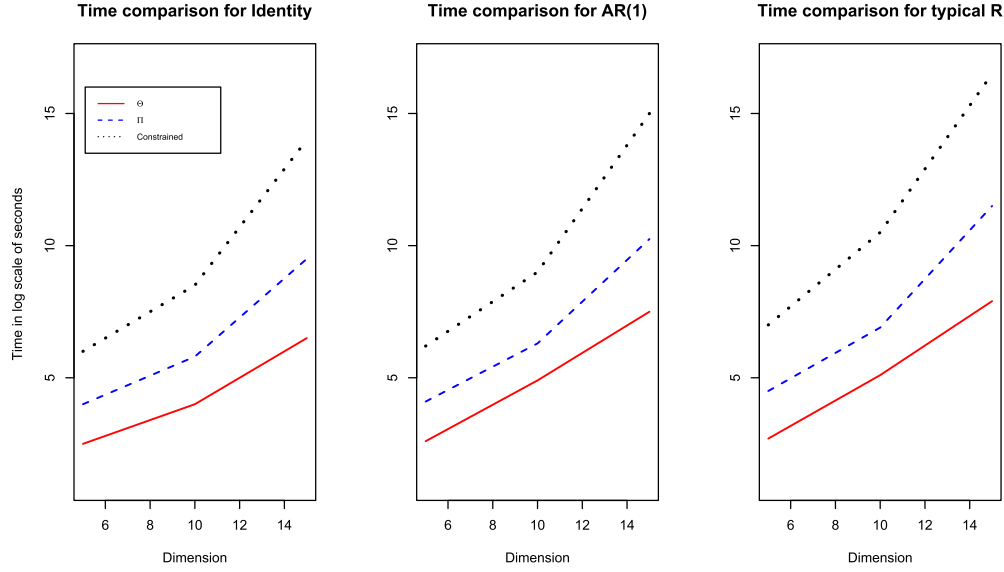
Figure 1: Time comparison in log scale of seconds for constrained vs unconstrained method for 2000 effectively independent MCMC samples for three indicated correlation matrices.

assignment (0 for control and 1 for exercise). For each patient, columns 2–9 denote the patient's smoking status from 5-th to 12-th week after they are asked to quit smoking. With $n = 281, k = 8$ (discarding first column), we associate an $n \times k$ matrix $Y = (y_{ij})$ to the data, whose entries take values $-1, 0, 1$; where 1 denotes success (i-th patient not smoking in j-th week), -1 denotes failure (still smoking in j-th week) and 0 denotes a missing observation. Introducing latent variables y_{ij}^*, we assume a multivariate probit model Chib and Greenberg (1998) where,

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > 0, \\ -1 & \text{if } y_{ij}^* < 0, \end{cases}$$

and if $y_{ij} = 0$, the sign of y_{ij}^* represents the (unobserved) quit status for the week.

Next, we assume $y_i^* = (y_{i1}^*, y_{i2}^*, \cdots, y_{ik}^*)^\top \sim N_k(\mu_i, R)$ for $i = 1, 2, \cdots, n$ and μ_i is parameterized as $\mu_i = X_i^\top \beta$; where X_i is a $q \times k$ matrix of covariates and β is a $q \times 1$ vector of regressors. To circumvent the identifiability issue, covariance matrix is restricted to be a correlation matrix. As in Gaskins et al. (2014), we consider two choices of X_i: time-varying which specifies a different μ_{it} for each time within each treatment group ($q = 2k$) and time-constant which gives the same μ_{it} across all times within treatment group ($q = 2$). With this set-up, we consider a flat prior on β and the priors on R are the selection and shrinkage priors in Gaskins et al. (2014) for PACs and the angle (Θ), respectively.

6.1   Posterior computation

For posterior computation, we run an MCMC chain for 12,000 iterations with a burn-in of 3000, retaining every tenth observation. The three sets of parameters appearing in the posterior are regression parameters, latent variables and correlation matrix.

1. Sampling β. The conditional posterior of β given latent variables $y_i^*, i = 1, 2, \ldots, n$ and R is multivariate normal.

2. Sampling R. For angle based priors, sampling scheme in Section 4 is used and R code provided in Gaskins et al. (2014) has been used for PAC based priors on the residuals $y_i^* - \mu_i$, for $i = 1, 2, \cdots, n$.

3. Sampling y_i^*s. For sampling latent variables, we use Proposition 1 of Liu et al. (2009) as in Gaskins et al. (2014).

For comparison we use deviance information criterion (DIC) which does not require counting the number of model parameters, making it an effective criterion for model selection when shrinkage or sparsity is concerned. DIC is defined as (Spiegelhalter et al., 2002) Dev $+ 2p_D$ where,

$$\text{Dev} = -2 \sum_{i=1}^{n} l(\hat{\beta}, \hat{R}|y_i), \tag{6.1}$$

$$p_D = E\{-2l(\beta, R|Y)\} - \text{Dev} \tag{6.2}$$

with l denoting log-likelihood function and expectation is taken with respect to the posterior distribution.

For the CTQ data, the posterior estimate $\hat{\beta}$ is the posterior mean, as for the posterior estimate of $\hat{R}$ we use the posterior median for angle-based priors and the one used by (Gaskins et al., 2014, pp.12) for PAC-based priors. The numerical results for various priors on the correlation matrix are reported in Table 2, where it can be seen that the DIC is smaller for the time constant mean structure in coherence with the findings of Gaskins et al. (2014). One can note that for time varying mean structure, the models are heavily penalized by p_D which deals with 14 extra parameters compared to time constant models. Our angle-based selection prior appears have a tendency of lower DIC value and hence preferred.

Our selection prior reports posterior mean of Θ to be

$$\hat{\Theta} = \begin{bmatrix} 0.81(0.69) & 0.63 & 0.49 & 0.48 & 0.45 & 0.27 & 0.02 \\ 0.89 & 0.82(0.72) & 0.66 & 0.64 & 0.61 & 0.29 & 0.09 \\ 1.05 & 0.85 & 0.86(0.85) & 0.79 & 0.72 & 0.41 & 0.09 \\ 1.06 & 1.02 & 0.92 & 0.86(0.76) & 0.72 & 0.62 & 0.19 \\ 1.10 & 1.09 & 1.05 & 1.02 & 1.01(0.81) & 0.78 & 0.25 \\ 1.30 & 1.18 & 1.12 & 1.10 & 1.09 & 1.05(0.54) & 0.20 \\ 1.55 & 1.54 & 1.48 & 1.30 & 1.25 & 1.24 & 1.21(0.05) \end{bmatrix}$$

and $\hat{R}$ is presented in upper triangular matrix.

Mean Structure	Prior	Dev	p_D	DIC
Time Constant	$p_{\pi;SH}$	1027	14	1055
Time Constant	$p_{\pi;SP}$	1045	13	1070
Time Constant	$p_{\theta;SH}$	1023	13	1049
Time Constant	$p_{\theta;SP}$	1026	11	1047
Time Varying	$p_{\pi;SH}$	1017	25	1069
Time Varying	$p_{\pi;SP}$	1037	29	1075
Time Varying	$p_{\theta;SH}$	1024	23	1069
Time Varying	$p_{\theta;SP}$	1019	22	1062

Table 2: DICs for various correlation priors for CTQ data.

7 Discussion

We have dealt with some computational challenges in Bayesian estimation of correlation matrices by using its Cholesky decomposition and the ensuing angles as the new parameters which vary freely in $[0, \pi)$. This reparameterization deals effectively with the positive-definiteness constraint on a correlation matrix and results in faster computation of the posteriors. At a first encounter, angles may not seem the most natural parameters in statistics. However, to our knowledge we have shown for the first time that the angles in the present context are simply the inverse cosine of the familiar semi-partial correlations, see Huber (1981), Eaves and Chang (1992), Cooke et al. (2011). Thus, the angles are statistically meaningful and the new connection opens up the possibility of using the wealth of distributions from directional statistics as potential priors for Bayesian analysis of correlation matrices. Through simulations and data analysis we have shown that the performance of our shrinkage and selection priors on the angles is better or comparable to those based on the PACs in Gaskins et al. (2014).

Supplementary Material

Supplementary Material: Bayesian Estimation of Correlation Matrices of Longitudinal Data (DOI: 10.1214/20-BA1237SUPP; .pdf).

References

Anderson, T. W. (2003). "An introduction to multivariate statistical analysis." *Wiley-Interscience*. MR1990662.   1039

Arnold, B. C. and SenGupta, A. (2006). "Probability distributions and statistical inference for axial data." *Environmental and Ecological Statistics*, 13(3): 271–285. MR2242189. doi: https://doi.org/10.1007/s10651-004-0011-8.   1046

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage." *Statistica Sinica*, 1281–1311. MR1804544.   1040, 1046, 1051, 1052

Blondes, M. S., Schuenemeyer, J. H., Olea, R. A., and Drew, L. J. (2013). "Aggregation of carbon dioxide sequestration storage assessment units." *Stochastic Environmental Research and Risk Assessment*, 27(8): 1839–1859. 1044

Carter, C. K., Wong, F., and Kohn, R. (2011). "Constructing priors based on model size for nondecomposable Gaussian graphical models: A simulation based approach." *Journal of Multivariate Analysis*, 102(5): 871–883. MR2776574. doi: `https://doi.org/10.1016/j.jmva.2011.01.001`. 1051

Chib, S. and Greenberg, E. (1998). "Analysis of multivariate probit models." *Biometrika*, 85(2): 347–361. 1048, 1053

Chiu, T. Y., Leonard, T., and Tsui, K.-W. (1996). "The matrix-logarithmic covariance model." *Journal of the American Statistical Association*, 91(433): 198–210. MR1394074. doi: `https://doi.org/10.2307/2291396`. 1039

Cooke, R. M., Joe, H., and Aas, K. (2011). "Vines arise." *Dependence Modeling: Vine Copula Handbook*, 37–71. MR2856968. 1045, 1055

Creal, D., Koopman, S. J., and Lucas, A. (2011). "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations." *Journal of Business & Economic Statistics*, 29(4): 552–563. MR2879242. doi: `https://doi.org/10.1198/jbes.2011.10070`. 1042

Damien, P., Wakefield, J., and Walker, S. (1999). "Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2): 331–344. MR1680334. doi: `https://doi.org/10.1111/1467-9868.00179`. 1049

Daniels, M. J. and Pourahmadi, M. (2009). "Modeling covariance matrices via partial autocorrelations." *Journal of Multivariate Analysis*, 100(10): 2352–2363. MR2560376. doi: `https://doi.org/10.1016/j.jmva.2009.04.015`. 1040

Eaves, D. and Chang, T. (1992). "Priors for ordered conditional variance and vector partial correlation." *Journal of Multivariate Analysis*, 41(1): 43–55. MR1156680. doi: `https://doi.org/10.1016/0047-259X(92)90056-L`. 1041, 1045, 1055

Fan, J., Fan, Y., and Lv, J. (2008). "High dimensional covariance matrix estimation using a factor model." *Journal of Econometrics*, 147(1): 186–197. MR2472991. doi: `https://doi.org/10.1016/j.jeconom.2008.09.017`. 1039

Gaskins, J., Daniels, M., and Marcus, B. (2014). "Sparsity inducing prior distributions for correlation matrices of longitudinal data." *Journal of Computational and Graphical Statistics*, 23(4): 966–984. MR3270706. doi: `https://doi.org/10.1080/10618600.2013.852553`. 1041, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055

Ghosh, R. P., Mallick, B., and Pourahmadi, M. (2020). "Supplementary Material: Bayesian Estimation of Correlation Matrices of Longitudinal Data." *Bayesian Analysis*. doi: `https://doi.org/10.1214/20-BA1237SUPP`. 1046

Huber, J. (1981). "Partial and Semipartial Correlation-A Vector Approach." *The Two-Year College Mathematics Journal*, 12(2): 151–153. 1041, 1045, 1055

Ishwaran, H. and Rao, J. S. (2005). "Spike and slab variable selection: frequentist and Bayesian strategies." *Annals of Statistics*, 730–773. MR2163158. doi: https://doi.org/10.1214/009053604000001147.   1047

Joe, H. (2006). "Generating random correlation matrices based on partial correlations." *Journal of Multivariate Analysis*, 97(10): 2177–2189. MR2301633. doi: https://doi.org/10.1016/j.jmva.2005.05.010.   1039, 1040, 1052

Kundu, S., Mallick, B. K., Baladandayuthapani, V., et al. (2019). "Efficient Bayesian regularization for graphical model selection." *Bayesian Analysis*, 14(2): 449–476. MR3934093. doi: https://doi.org/10.1214/17-BA1086.   1040

Kurowicka, D. and Cooke, R. (2003). "A parameterization of positive definite matrices in terms of partial correlation vines." *Linear Algebra and its Applications*, 372: 225–251. MR1999149. doi: https://doi.org/10.1016/S0024-3795(03)00507-X.   1040

Kurowicka, D. and Cooke, R. (2006). "Completion problem with partial correlation vines." *Linear Algebra and its Applications*, 418(1): 188–200. MR2257589. doi: https://doi.org/10.1016/j.laa.2006.01.031.   1040

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). "Generating random correlation matrices based on vines and extended onion method." *Journal of Multivariate Analysis*, 100(9): 1989–2001. MR2543081. doi: https://doi.org/10.1016/j.jmva.2009.04.008.   1052

Liechty, J. C., Liechty, M. W., and Müller, P. (2004). "Bayesian correlation estimation." *Biometrika*, 91(1): 1–14. MR2050456. doi: https://doi.org/10.1093/biomet/91.1.1.   1040, 1044, 1048, 1051, 1052

Liu, X., Daniels, M. J., and Marcus, B. (2009). "Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial." *Journal of the American Statistical Association*, 104(486): 429–438. MR2751428. doi: https://doi.org/10.1198/016214508000000904.   1054

Madar, V. (2015). "Direct formulation to Cholesky decomposition of a general nonsingular correlation matrix." *Statistics & Probability Letters*, 103: 142–147. MR3350874. doi: https://doi.org/10.1016/j.spl.2015.03.014.   1041, 1045

Marcus, B. H., Albrecht, A. E., King, T. K., Parisi, A. F., Pinto, B. M., Roberts, M., Niaura, R. S., and Abrams, D. B. (1999). "The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial." *Archives of Internal Medicine*, 159(11): 1229–1234.   1052

Mitchell, T. J. and Beauchamp, J. J. (1988). "Bayesian variable selection in linear regression." *Journal of the American Statistical Association*, 83(404): 1023–1032. MR0997578.   1047

Neal, R. M. et al. (2003). "Slice sampling." *The Annals of Statistics*, 31(3): 705–767. MR1994729. doi: https://doi.org/10.1214/aos/1056562461.   1049

Pinheiro, J. C. and Bates, D. M. (1996). "Unconstrained parametrizations for variance-covariance matrices." *Statistics and Computing*, 6(3): 289–296.   1039, 1040, 1041

Pitt, M., Chan, D., and Kohn, R. (2006). "Efficient Bayesian inference for Gaussian copula regression models." *Biometrika*, 93(3): 537–554. MR2261441. doi: https://doi.org/10.1093/biomet/93.3.537. 1051

Pourahmadi, M. (1999). "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation." *Biometrika*, 86(3): 677–690. MR1723786. doi: https://doi.org/10.1093/biomet/86.3.677. 1039

Pourahmadi, M. and Wang, X. (2015). "Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor." *Statistics & Probability Letters*, 106: 5–12. MR3389964. doi: https://doi.org/10.1016/j.spl.2015.06.015. 1040, 1046, 1052

Rapisarda, F., Brigo, D., and Mercurio, F. (2007). "Parameterizing correlations: a geometric interpretation." *IMA Journal of Management Mathematics*, 18(1): 55–73. MR2332847. doi: https://doi.org/10.1093/imaman/dpl010. 1039, 1041

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639. MR1979380. doi: https://doi.org/10.1111/1467-9868.00353. 1054

Tsay, R. S. and Pourahmadi, M. (2017). "Modelling structured correlation matrices." *Biometrika*, 104(1): 237–242. MR3626474. doi: https://doi.org/10.1093/biomet/asw061. 1040, 1042, 1043, 1044, 1045

Wang, H. and Pillai, N. S. (2013). "On a class of shrinkage priors for covariance matrix estimation." *Journal of Computational and Graphical Statistics*, 22(3): 689–707. MR3173737. doi: https://doi.org/10.1080/10618600.2013.785732. 1048

Wang, Y. and Daniels, M. (2014). "Computationally efficient banding of large covariance matrices for ordered data and connections to banding the inverse Cholesky factor." *Journal of Multivariate Analysis*, 130: 21–26. MR3229521. doi: https://doi.org/10.1016/j.jmva.2014.04.026. 1041

Wong, F., Carter, C. K., and Kohn, R. (2003). "Efficient estimation of covariance selection models." *Biometrika*, 90(4): 809–830. MR2024759. doi: https://doi.org/10.1093/biomet/90.4.809. 1051

Zhang, W., Leng, C., and Tang, C. Y. (2015). "A joint modelling approach for longitudinal studies." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1): 219–238. MR3299406. doi: https://doi.org/10.1111/rssb.12065. 1042