A New Method for Generating Random Correlation Matrices*

Ilya Archakov[a] and Peter Reinhard Hansen[b,c] and Yiyao Luo[b†]

[a] *University of Vienna*

[b] *University of North Carolina*

[c] *Copenhagen Business School*

Preliminary version. February 27, 2022

Abstract

We propose a new method for generating random correlation matrices that makes it simple to control both location and dispersion. The method is based on a vector parameterization, $\gamma = g(C)$, that shows that any distribution on $\mathbb{R}^{n(n-1)/2}$ translates to a distribution on the space of non-singular $n \times n$ correlation matrices. Correlation matrices with certain particular structures, such as block structures and strictly positive correlations are simple to generate. We compare the new method with existing methods.

Keywords: Covariance Modeling, Random Correlation Matrix, Fisher Transformation.

JEL Classification: C10; C22; C58

*Thanks...

†Address: University of North Carolina, Department of Economics, 107 Gardner Hall Chapel Hill, NC 27599-3305

1 Introduction

The correlation matrix plays a central role in many multivariate models. Random correlation matrices are commonly used in Bayesian analysis to specify priors and are used in frequentist analysis for the purpose of investigating properties of estimators, hypothesis tests, and other statistics. Generating random $n \times n$ correlation matrices can become onerous when particular features or structures are required for the correlation matrix. A variety of methods have been proposed in the literature to serve different needs, see Pourahmadi (2011) for a review. In this paper, we propose a novel method for generating random correlation matrices that is well-suited for a wide range of objectives. It can, in principle, be used to generate random correlation matrices with any distribution on the set on non-singular correlation matrices. Positive definite correlation matrices are guaranteed and it is simple to control both the location and dispersion of the correlation matrix. It is also possible to generate random correlation matrices with some specific structures, such as block structures and correlation matrices with the Perron-Frobenius property.

The paper is organized as follows. We introduced the new method for generating random correlation matrices in Section 2 and discuss several features and structures that can be generated with the new method. In Section 3, we review some existing methods for generating random correlation matrices and discuss their properties.

2 Random Correlation Matrices: A New Method

The proposed method for generating random correlation matrices is based on the following vector parameterization of non-singular correlation matrices,

$$\gamma = g(C) := \text{vecl}(\log C),$$

where the operator $\text{vecl}(\cdot)$ vectorizes the lower off-diagonal elements and $\log C$ is the matrix logarithm[1] of C. For an $n \times n$ correlation matrix the vector, $\gamma = g(C)$, has dimension $d = n(n-1)/2$. This parametrization was introduced in Archakov and Hansen (2021), who showed that g is a one-to-one correspondence between the set of $n \times n$ non-singular correlation matrices, denoted $\mathcal{C}_{n\times n}$, and $\mathbb{R}^d$. So, any vector, $\gamma \in \mathbb{R}^d$, corresponds to a unique correlation matrix $C(\gamma) \equiv g^{-1}(\gamma)$.

The new method for generating a random correlation matrices is simply to draw γ from a distribution

[1] The matrix logarithm for a non-singular correlation matrix with eigendecomposition, $C = Q\Lambda Q'$, is given by $\log C = Q \log \Lambda Q'$, where $\log \Lambda = \text{diag}(\log \lambda_1, \ldots, \log \lambda_n)$.

on $\mathbb{R}^d$ and compute $C(\gamma)$.[2] It follows that the mapping, $\gamma \mapsto C(\gamma)$, will induce a distribution on $\mathcal{C}_{n\times n}$ from any distribution on $\mathbb{R}^d$. For instance, the density, $f_\gamma(\gamma)$, on $\mathbb{R}^d$, will translate to the density

$$f_C(C) = f_\gamma(g(C))|\psi(C)|, \qquad \text{on } \mathcal{C}_{n\times n},$$

where $\psi(C)$ is the determinant of $\mathrm{d}\gamma/\mathrm{d}\varrho$ and $\varrho = \mathrm{vecl}C$ is the vector with the correlation coefficients in C. A simple example, for the case $n = 2$, is the logistic density, $f_\gamma(\gamma) = 2e^{-2\gamma}/(1 + e^{-2\gamma})^2$, that translates to a random 2×2 correlation matrix where the correlation coefficient is uniformly distribution on $[-1, 1]$, see Theorem 2 below.

The new method for generating random correlation matrices can be used to generate correlation matrices with some specific features that we discuss in the subsections below.

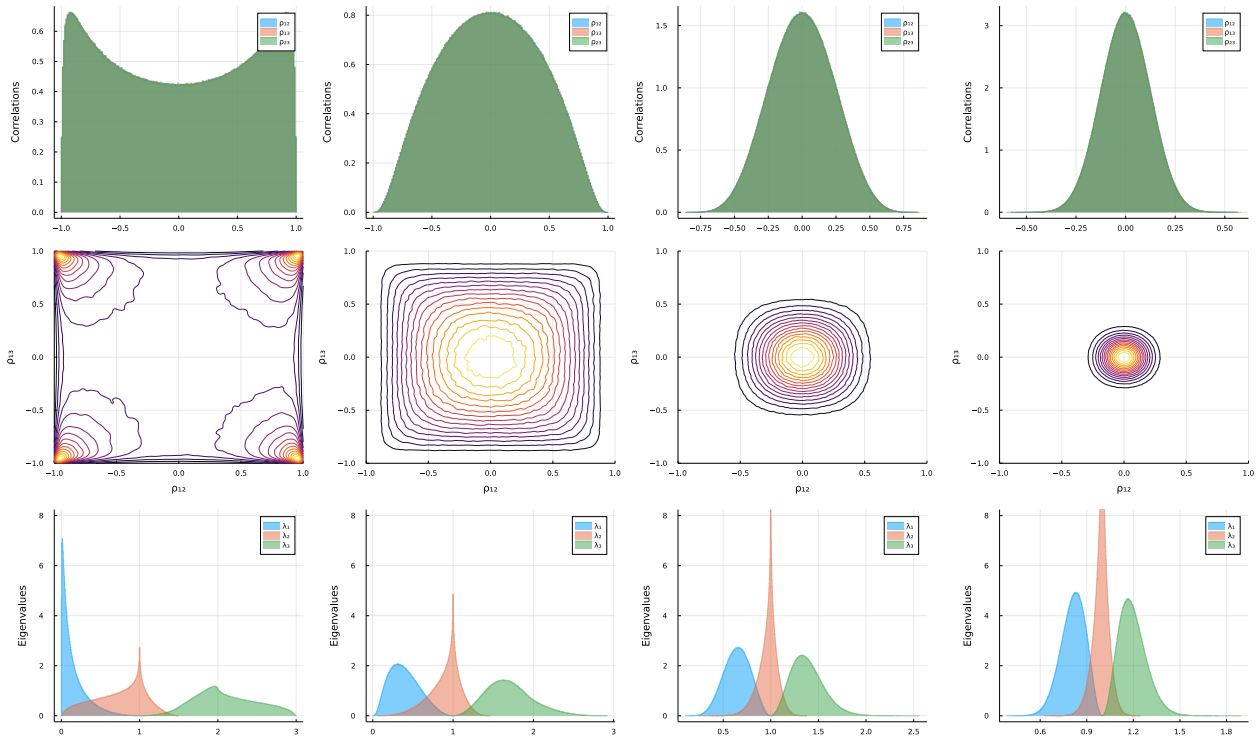
2.1 Correlation Coefficients with Identical Marginal Distributions

Some of the existing methods for generating random correlation matrices are carefully crafted to generate correlation coefficients with identical marginal distributions (typically a Beta-distributions).

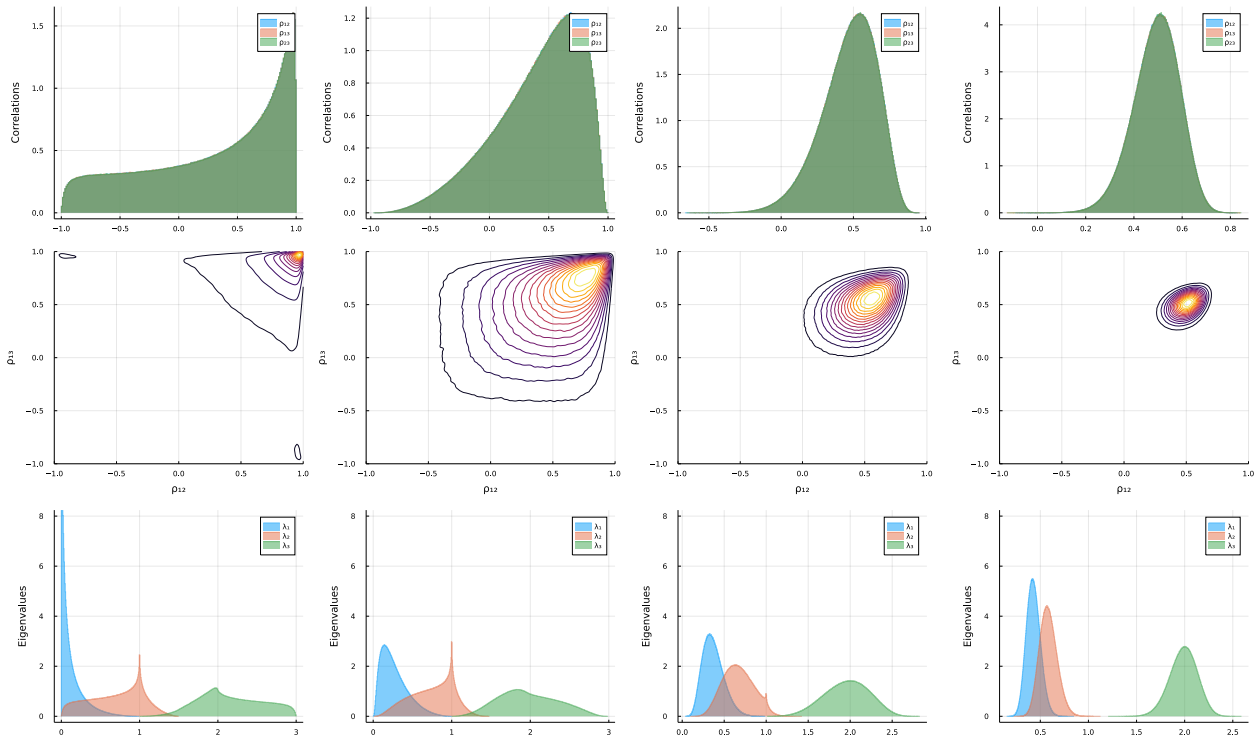
Theorem 1 (Permutation invariance)**.** *Let $P \in \mathbb{R}^{n\times n}$ be a permutation matrix and let $\gamma_i = \xi + \varepsilon_i$, where ε_i, $i = 1, \ldots, d$, are independent and identically distribution and independent of the common random variables, ξ. Then $C(\gamma)$ and $\tilde{C} = PC(\gamma)P'$ are identically distributed on $\mathcal{C}_{n\times n}$.*

An immediate implication of Theorem 1 is that all the marginal distributions of the correlations, ρ_{ij}, $i \neq j$ are identical, whenever the elements of γ are independent and identically distributed., i.e. the case were $\xi = 0$. More generally, the vector of correlations in the upper left principal submatrix, $\varrho = \mathrm{vecl}[C(\gamma)]_{i,j=1,\ldots,k}] \in \mathbb{R}^{k(k-1)/2}$ has the same distribution as the vector of correlations corresponding to any other principal submatrix, $\tilde{\varrho} = \mathrm{vecl}[C(\gamma)]_{i,j\in\{i_1,\ldots,i_k\}}]$, for some $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$. Existing methods for generating random correlation matrices can produce identically distributed correlation coefficients with a symmetric Beta distributions on $[-1, 1]$, as we detail in the Section 3. Theorem 1 shows that the new method makes it is possible to generate identically distributed correlation coefficients with a wide range of distributions, beyond Beta distributions. Theorem 1 can be generalized to the case where

[2] A simple algorithm for computing $C(\gamma)$ is given in Archakov and Hansen (2021).



(a) $\gamma \sim N_3(0, I)$ (b) $\gamma \sim N_3(0, \frac{1}{4}I)$ (c) $\gamma \sim N_3(0, \frac{1}{16}I)$ (d) $\gamma \sim N_3(0, \frac{1}{64}I)$



(e) $\gamma \sim N_3(\gamma_0, I)$ (f) $\gamma \sim N_3(\gamma_0, \frac{1}{4}I)$ (g) $\gamma \sim N_3(\gamma_0, \frac{1}{16}I)$ (h) $\gamma \sim N_3(\gamma_0, \frac{1}{64}I)$

Figure 1: Properties of random 3×3 correlation matrices generated with $\gamma \sim iidN(\mu_i, \omega_j^2)$, with $\mu_0 = 0$ (upper panels) $\mu_2 = \frac{1}{3}\log 4$ (lower panels) and $\omega_j^2 = 1, \frac{1}{4}, \frac{1}{16}$, and $\frac{1}{64}$ from left to right. In each panel we show: The marginal distribution of ρ_{ij}; Contour plot for the bivariate distribution of (ρ_{12}, ρ_{13}); and the densities of ordered eigenvalues of C.

We illustrate the new method for generating random correlation matrices based on a Gaussian distributed γ. Key features of the resulting random correlation matrices are shown in Figure 1 for the case where $n = 3$. Panels (a)-(d) corresponds to the case where $\gamma_i \sim iidN(0, \omega^2)$, $i = 1, 2, 3$, such that the random correlation matrices are located about $C = I_3$. Panels (e)-(h) are based on $\gamma_i \sim iidN(\frac{1}{3}\log 4, \omega^2)$. This leads to random correlation matrices about the equicorrelation matrix

$$\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix} = C(\gamma^*), \qquad \gamma^* = \frac{\log 4}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

We display the marginal distributions for the correlation coefficients (upper panels), contour plots for bivariate distributions (middle panels), and the distribution for the three eigenvalues (lower panels) for the cases $\omega^2 = 1$, $\frac{1}{4}$, $\frac{1}{16}$, and $\frac{1}{64}$. From Theorem 1 we know that the marginal distributions are identical when the elements of γ are independent and this can be seen from the simulated densities for ρ_{12}, ρ_{13}, and ρ_{23}, that are indistinguishable in all cases. The plots in the second row of each panel present bivariate distributions for a pair of correlations with contour plots. We present the simulated one for (ρ_{12}, ρ_{13}), but these bivariate distributions are identical for all pairs of correlation coefficients, as a consequence of Theorem 1.

When the variance of the elements of γ is relatively large, $\omega^2 = 1$, then $C(\gamma)$ tends to produce near-singular correlation matrices,. This is evident from the distribution of the smallest eigenvalue in Panels (a) and (e), and it can also be seen from the contour plots where the mass in concentrated in the corners. As the variance of γ_i is reduced, so is the variance of the resulting correlation coefficients. In Panels (a)-(d), the random correlation matrices become more concentrated about $C(0) = I$ and in Panels (e)-(h) the random correlations get more concentrated about $\frac{1}{2}$.

2.1.1 Heterogenous Marginal Distributions

In some applications it can be desirable to generate random correlation matrices where the dispersion of the correlation coefficients is heterogeneous. This situation could arise in a Bayesian context where there is strong prior knowledge about some correlation coefficients and less prior information about other correlations. The new method can accommodate this situation by using different variances over the elements in γ,. This follows because the Jacobian, $\mathrm{d}\varrho/\mathrm{d}\gamma$, is approximately diagonal.

2.2 Random Perturbation of Particular Correlation Matrix

When the objective is to generate random correlation matrices in the vicinity of a particular correlation matrix, C_0 say, then this can be achieved as follows: Let $\gamma_0 = g(C_0)$ be the vector that corresponds to C_0, then we can generate random correlation matrices in the vicinity of C_0 using $C(\gamma_0 + \varepsilon)$, where ε is a random vector centered about the zero-vector. The dispersion of the random correlation matrices about C_0 is controlled by the dispersion of ε.

2.3 Resembling the Distribution of Empirical Correlation Matrices

The method can be used to approximate the distribution of empirical correlation matrices. Let $\hat{C}$ be an empirical correlation matrix computed from T observations and consider $\hat{\gamma} = g(\hat{C})$. Under suitable regularity conditions, Archakov and Hansen (2021) showed that $\sqrt{T}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, V_\gamma)$ and derived an expression for V_γ. This asymptotic approximation works well in finite samples and the off-diagonal elements of V_γ tend to be close to zero, especially for high-dimensional correlation matrices, see Archakov and Hansen (2020). This suggests that the new method can be used to resemble the distribution of empirical correlation matrices by drawing γ from a suitable Gaussian distribution.

2.4 Equicorrelation Matrices

An equicorrelation matrix, C, is a correlation matrix where all the correlations are identical and the corresponding $\gamma = g(C)$ will be a vector whose elements are all identical. Let r denoted the common correlation coefficient in C and let z be the corresponding common element of γ, then the relationship between the two is given by,

$$z(r) = \frac{1}{n} \log\left(1 + n\tfrac{r}{1-r}\right),$$

and the inverse transformation is $r(z) = \frac{1-e^{-nz}}{1+(n-1)e^{-nz}}$, see e.g. Archakov and Hansen (2021). An equicorrelation matrix has two eigenvalues, $1 + r(n-1)$ and $1 - r$, where the latter has multiplicity $n - 1$, see Olkin and Pratt (1958). Thus, the $n \times n$ equicorrelation matrix is positive definite if and only if $r \in (-\frac{1}{n-1}, 1)$.

The following theorem shows how a random correlation matrix can be generated in such a way that r has a Beta distribution on this interval.

Theorem 2. *Let $\gamma = (z, \ldots, z)' \in \mathbb{R}^d$ with $d = n(n-1)/2$. Then $C(\gamma)$ is an equicorrelation matrix, where the common correlation coefficient, r, is confined to the interval $(-\frac{1}{n-1}, 1)$ for all $z \in \mathbb{R}$.*

Moreover, if z has density,

$$f_z(z) = \frac{1}{B(\alpha,\beta)} \frac{e^{-\beta\frac{z-\mu}{s}}}{s\left(1+e^{-\frac{z-\mu}{s}}\right)^{\alpha+\beta}}, \qquad z \in \mathbb{R}, \tag{1}$$

where $\mu = \frac{\log(n-1)}{n}$ and $s = \frac{1}{n}$, then r is Beta distributed, $B(\alpha,\beta)$, on the interval $(-\frac{1}{n-1}, 1)$.

The expression (1) is known as the Exponential Generalized Beta distribution of the second type and it is also known as the Generalized Logistic Distribution of Type IV.

If we set $\alpha = \beta = 1$, it follows immediately that r is uniformly distributed on $(-\frac{1}{n-1}, 1)$.

Corollary 1. *Let $\gamma = (z, \ldots, z)' \in \mathbb{R}^d$ with $d = n(n-1)/2$, and suppose that z is logistically distributed,*

$$f_z(z) = \frac{e^{-\frac{z-\mu}{s}}}{s\left(1+e^{-\frac{z-\mu}{s}}\right)^{2}}, \qquad z \in \mathbb{R}, \tag{2}$$

where $\mu = \frac{\log(n-1)}{n}$ and $s = \frac{1}{n}$, then r is uniformly distribution on the interval $(-\frac{1}{n-1}, 1)$.

In the special case where $n = 2$, we have $\mu = 0$ and $s = 1/2$ and the logistic distribution in (2) is also know as a Fisher z-distribution with degrees of freedom $(d_1, d_2) = (2, 2)$.[3]

Theorem 2 provides valuable insight about the dispersion of the elements in γ that one might expect as the dimension of the correlation matrix, n, increases, because $\mathrm{var}(z) = \frac{\pi^2}{3}n^{-2}$, in the distribution (2). This indicates that a scaling factor of $1/n$ should be used on the elements of γ to preserve similarly dispersion for the correlation coefficients in $C(\gamma)$ as n increases.

2.5 Block Correlation Matrices

If C has a block structure then $\log C$ and C^{-1} has the same block structure, see Archakov and Hansen (2022). This can be used to generate random correlation matrices or random precision matrices, C^{-1}, with a desired block structure, where positive definiteness is always guaranteed. A correlation matrix has a block structure if

$$C = \begin{bmatrix} C_{[1,1]} & C_{[1,2]} & \cdots & C_{[1,K]} \\ C_{[2,1]} & C_{[2,2]} & & \\ \vdots & & \ddots & \\ C_{[K,1]} & & & C_{[K,K]} \end{bmatrix} \in \mathbb{R}^{n\times n},$$

[3]Moreover, in this case where $Z \sim \mathrm{logistic}(0, \frac{1}{2})$ we also have that $\exp(2Z) \sim F(2,2)$, (the F-distribution with degrees of freedom $d_1 = d_2 = 2$).

where the diagonal blocks, $C_{[i,i]} \in \mathbb{R}^{n_i \times n_i}$, have ones along the diagonal and $\rho_{i,i} \in \mathbb{R}$ in all off-diagonal elements, and the off diagonal blocks, $C_{[i,j]} \in \mathbb{R}^{n_i \times n_j}$ have all elements equal to $\rho_{i,j} \in \mathbb{R}$, $i, j = 1, \ldots, K$ and $n_1 + \cdots + n_K = n$.

It is also possible to generate random correlation matrices in the vicinity of a particular block correlation matrix and the canonical representation of block matrices can also be used to generate high-dimensional correlation matrices by taking convex combinations of permutated random block matrices. The advantage of this approach is that each of the block matrices involve lower-dimensional objects. In Figure 2 provides an example of this. First we generated 20 random block correlation matrices, each having 3×3 blocks with a single common correlation coefficient within each block. The rows (and columns) are then shuffled with a random order, and their average define the random 200×200 correlation matrix, which is guaranteed to be positive definite. The smallest eigenvalue of the correlation matrix shown in Figure 2 is $\lambda_{\min} = 0.0726$ and

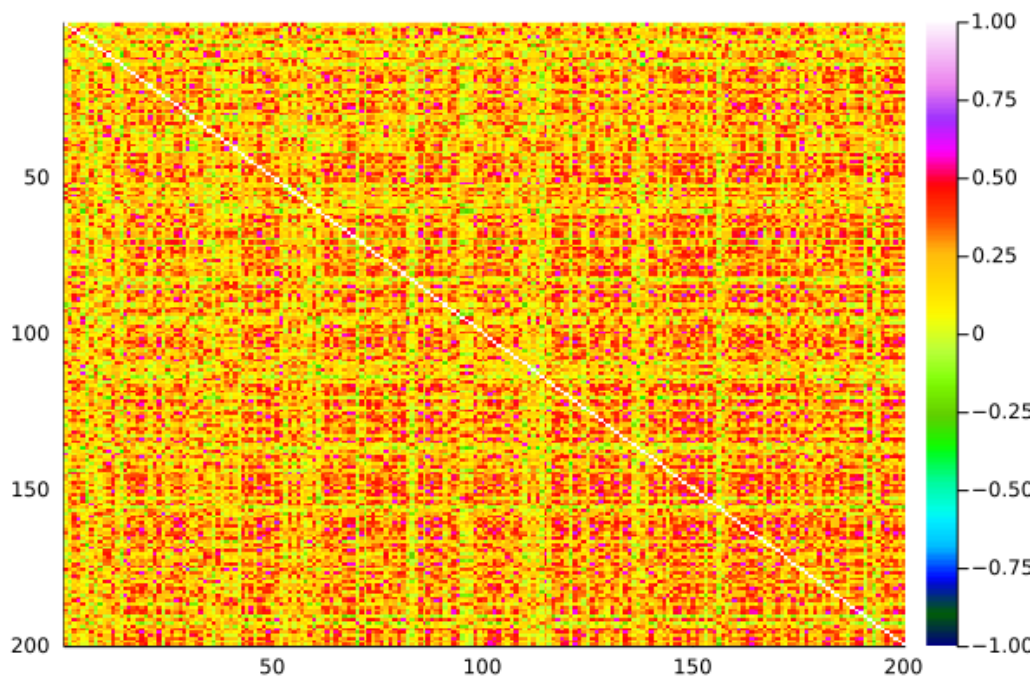


Figure 2: A random 200x200 correlation matrix, constructed as a the average of 20 random block matrices whose rows and column were subject to random permutations.

2.6 Positive Random Correlation Matrices and Monotonicity

A random correlation matrix with strictly positive correlation coefficients is guaranteed by drawing γ from a distribution on $\mathbb{R}^d_+$ (i.e. vectors with strictly positive elements). This would, for instance, guarantee that the Perron-Frobenius theorem is applicable to $C(\gamma)$. The less stringent requirement

that C has nonnegative coefficients is guaranteed by drawing γ from a distribution on $\mathbb{R}^d_{\geq 0}$ (i.e. vectors with non-negative elements). The latter follows from the fact that γ are the off-diagonal elements of $\log C$. A matrix with non-negative off-diagonal elements is a Metzler matrix and it is well known (from the literature on Markov processes) that the exponential of a Metzler matrix is a non-negative matrix. Thus, if the off-diagonal elements of $\log C$ are non-negative, the same is true for C.

Conjecture 1. *Let $\tilde{C} = g(\tilde{\gamma})$ and $C = g(\gamma)$ where $\tilde{\gamma} \geq \gamma \geq 0$, then $\tilde{C} \geq C$ where all inequalities are element-wise.*

2.7 A Bound for Smallest Eigenvalue of C

Conjecture 2. *Let $\gamma_{\max} = \max_k |\gamma_k|$ be the largest element of γ in absolute value. Then,*

$$-n\gamma_{\max} \leq \log \lambda_{\min} \leq -\gamma_{\max}.$$

So, if $\max_i |\gamma_i| \leq K$ for some constant, K, then the smallest eigenvalue of $C(\gamma)$ is bounded away from zero.

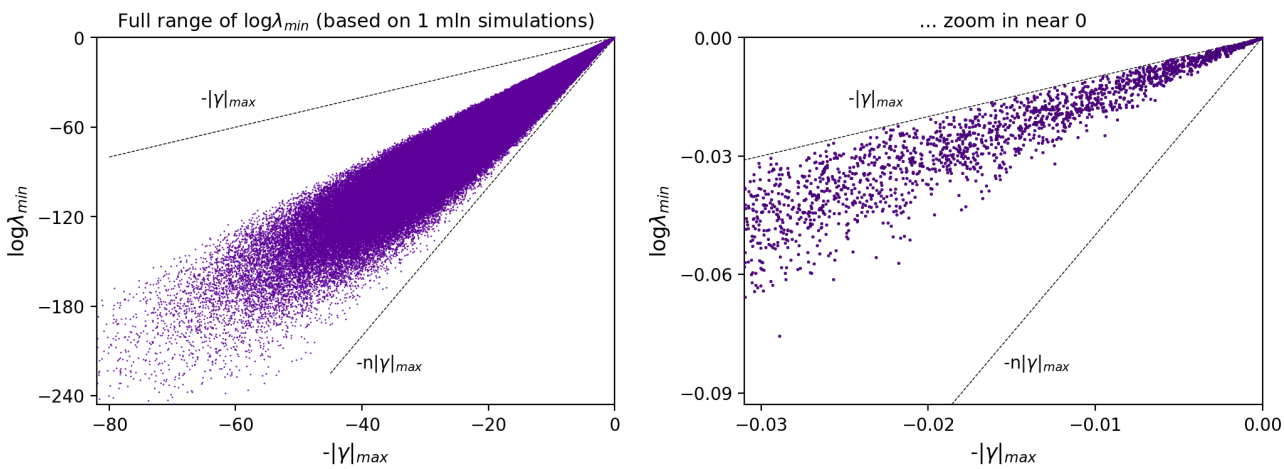


Figure 3: Scatter plots of $\log \lambda_{\min}$ against $-\gamma_{\max}$ for one million random 5×5 correlation matrices.

In Figure 3 we have shown $\log \lambda_{\min}$ plotted against $-\gamma_{\max}$ for one million random correlation matrices with $n = 5$ [add a higher dimension] along with the conjectured upper and lower bound for $\log \lambda_{\min}$. It appears that the lower bound is binding for very large values of $\gamma_{\max}$ whereas the upper bound only becomes binding for $\gamma_{\max} \simeq 0$ which is the case where $C \simeq I$.

3 Existing methods for Generating Random Correlation Matrices

Next, we discuss several existing methods for generating random correlation matrices. Several of these methods produce correlation matrices where the marginal distributions are Beta-distributions.

3.1 Random Gram Methods

A valid correlation matrix can be obtained from any matrix $m \times n$ matrix, $U = (u_1, \ldots, u_n)$, with normalized columns, $u_j' u_j = 1$, for $j = 1, \ldots, n$. It follows immediately that $C = U'U$ is positive semidefinite with ones along the diagonal. Moreover, if U has rank n, then $C = U'U$ is a non-singular correlation matrix. Several methods are based on this idea (typically with $m = n$), where a random correlation matrix is obtained from from random vectors, $u_1, \ldots, u_n$, on the unit sphere, $S_m = \{u \in \mathbb{R}^m, u'u = 1\}$. The random Gram method generates n vectors on S_n and the Gram matrix $C = U'U$ is the resulting random correlation matrix. The uniformly distribution on S_n was discussed in Marsaglia and Olkin (1984), see also Holmes (1991), and it generates a C where the marginal distributions of the correlation coefficients are Beta distributed, $B(\frac{1}{2}, \frac{n-1}{2})$, see Marsaglia and Olkin (1984). The vectors, u_j, $j = 1, \ldots, n$, can be drawn from other distributions, such as that proposed by Tuitman et al. (2020), which ensures that the average correlation coefficient is centered about a particular value.

3.2 Standard Angles Parameterization (SAP) Method

A variant of the Random Gram method is the case where U is a triangular matrix. This choice was discussed in Marsaglia and Olkin (1984) and a particular triangular form was proposed by Pinheiro and Bates (1996). Their choice for U is defined from the angels, $\theta_{ij} \in [0, \pi)$, for $1 \leq i < j \leq n$, such that

$$U = \begin{bmatrix} 1 & \cos\theta_{1,2} & \cos\theta_{1,3} & \cdots & \cos\theta_{1,n-1} & \cos\theta_{1,n} \\ 0 & \sin\theta_{1,2} & \cos\theta_{2,3}\sin\theta_{1,3} & \cdots & \cos\theta_{2,n-1}\sin\theta_{1,n-1} & \cos\theta_{2,n}\sin\theta_{1,n} \\ 0 & 0 & \Pi_{i=1}^{2}\sin\theta_{i,3} & & \cos\theta_{3,n-1}\Pi_{i=1}^{2}\sin\theta_{i,n-1} & \cos\theta_{3,n}\Pi_{i=1}^{2}\sin\theta_{i,n} \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & & \Pi_{i=1}^{n-2}\sin\theta_{i,n-1} & \cos\theta_{n-1,n}\Pi_{i=1}^{n-2}\sin\theta_{i,n} \\ 0 & 0 & 0 & \cdots & 0 & \Pi_{i=1}^{n-1}\sin\theta_{i,n} \end{bmatrix},$$

which is an upper triangular matrix. This requires $d = n(n-1)/2$ angles, θ_{ij}, and it follows that any distribution on $[0, \pi)^d$ will correspond to some distribution over the space of correlation matrices. If the angles are independent and uniformly distributed on $[0, \pi)$ then C has coefficients with very heterogeneous marginal distributions. If one instead specifies θ_{ij} to have the density

$$f_j(x; \alpha) = \frac{\sin^{2\alpha - j}(x)}{B(\alpha - \frac{j-1}{2}, \frac{1}{2})}, \qquad j = 1, \ldots, n-1,$$

where $\alpha \geq n/2$ is some parameter to be chosen, then marginal distributions of the correlation coefficients are identical and Beta distributed, Beta(α, α) on the interval $[-1, 1]$, see Pourahmadi and Wang (2015). This method is known as the *Standard Angles Parameterization* (SAP) method.

3.3 Eigendecomposition Method

A different approach is to use the eigendecomposition of the correlation matrix, $C = Q'\Lambda Q$ where $Q'Q = I$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. The approach has the advantage that one can control the distribution of the eigenvalues, which are confined to the simplex $\{(\lambda_1, \ldots, \lambda_n) : \sum_j \lambda_j = n,\ \lambda_j \geq 0\}$.

This eigendecomposition method generates the n random eigenvalues of C in the first step and then determines a valid set of eigenvectors in the second step. The second step is not trivial because the set of Q matrices that yield a valid correlation matrix for a given set of eigenvalues is a *shy* set in the set of all orthonormal matrices. For the pair (Λ, Q) to generate a valid correlation matrix, the following conditions must be satisfied.

1. The diagonal matrix, Λ, must satisfies $\lambda_j \geq 0$, $j = 1, \ldots, n$, and $\sum_{j=1}^n \lambda_j = n$.
2. The matrix $Q = (q_1, \ldots, q_n)$ must be orthonormal, $q_j' q_j = 1$ and $q_i' q_j = 0$ for all $i \neq j = 1, \ldots, n$.
3. Combined they must satisfy $q_j' \Lambda q_j = 1$, $j = 1, \ldots, n$.

The last condition is a cross restriction on Λ and Q. Given a particular Λ, almost all Q-matrices are precluded, such that this method requires a way to determine a valid Q-matrix. Algorithms for this were proposed by Chalmers (1975), Bendel and Mickey (1978), Marsaglia and Olkin (1984), and Davies and Higham (2000).[4] These methods begins with an initial (random) orthonormal matrix, Q_0, that is subjected to successive transformations until a valid Q-matrix is determined. The method by Davies and Higham (2000) is implemented in the MATLAB function gallery('`randcorr`').

[4] Holmes (1991) provides a comprehensive study of the statistical properties of spectral functions of correlation matrices generated by Bendel and Mickey's algorithm. For financial applications, Hüttner and Mai (2019) adapt the Bendel-Mickey Algorithm to generate correlation matrices with a Perron-Frobenius property.

3.4 Partial Correlations (PAC) Method

The partial correlation (PAC) method by Joe (2006) uses random partial correlations to generate random correlation matrices. Specifically the $n(n-1)/2$ partial correlations for the i-th and j-th variable conditional on the variables indexed between i and j. These are given by

$$\varrho_{ij} = \frac{C_{ij} - d_{ij}^{(i,j)}}{\sqrt{(1-d_{ii}^{(i,j)})(1-d_{jj}^{(i,j)})}}, \qquad \text{for} \quad 1 \leq i < j \leq n,$$

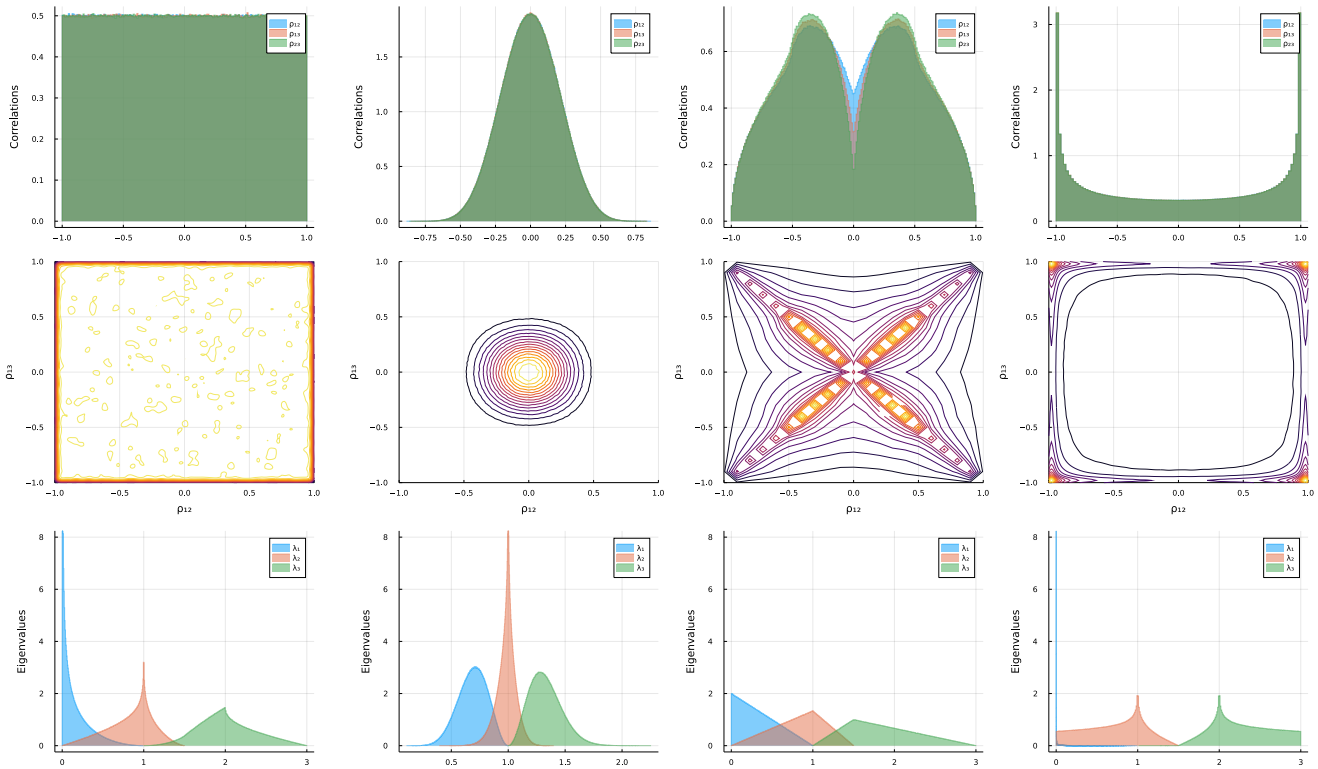
where $d_{ij}^{(i,j)} = C_{i,I_{ij}}[C_{I_{ij},I_{ij}}]^{-1}C_{I_{ij},j}$, and $C_{I_{ij},I_{ij}} = [C_{l,m}]_{i<l,m<j}$, $C_{i,I_{ij}} = [C_{i,m}]_{i<m<j}$, and $C_{I_{ij},j} = C'_{j,I_{ij}}$, are sub-matrices of C. When $j = i+1$ the partial correlation is simply the correlation, $\varrho_{i,i+1} = C_{i,i+1}$, otherwise it is the partial correlation between the i-th and j-th variables, conditional on all variables indexed between i and j. Clearly any correlation matrix, C, will map to $\{\varrho_{i,j}\}_{1\leq i<j\leq n}$ and any set of these partial correlation in $(-1,1)$, will translate to a valid correlation matrix. Interestingly, the determinant of C is given by $\det C = \prod_{1\leq i<j\leq n}(1-\varrho_{ij}^2)$, see Joe (2006, theorem 1).[5] The PAC method draws from a distribution on $(-1,1)^d$, with $d = n(n-1)/2$, and reconstructs the correlations from the partial correlations.

When the partial correlations, $\{\varrho_{i,j}\}_{1\leq i<j\leq n}$, are drawn independently from the Beta-distribution, $\text{Beta}(\alpha_{ij},\alpha_{ij})$ on $(-1,1)$, with $\alpha_{ij} = \alpha + (1-j+i)/2$, then the marginal distributions of the correlation coefficients are identical and given by $\text{Beta}(\alpha,\alpha)$, where $\alpha \geq (n-2)/2$, see Joe (2006). Moreover, the joint density of all correlations becomes proportional to the determinant of the correlation matrix to the power $\alpha - n/2$. [6]

The properties of some random correlation matrices, $n = 3$, are shown in Figure 4. Panel (a) is the Random Gram method where u_j, $j = 1,\ldots,3$ are independent and uniformly distributed on the sphere, S_3,. This choice yields uniformly distributed correlation coefficients. Panel (b) is the SAP method with $\alpha =$, Panel (c) is the eigendecomposition-based method and it produces rather bizarre marginal and joint distributions for the correlations. This suggests that the algorithm used to determine a valid orthonormal matrix, Q, has odd implications for the distributions of correlation coefficients. Panel (d) is the PAC method with $\alpha =$ (tiny value). Although SAP and PAC are different methods, they can produce the same marginal distributions for the correlation coefficients – the symmetric Beta-distributions, $\text{Beta}(\alpha,\alpha)$. However, the range for PAC, $\alpha > n/2-1$, is slightly larger than that of SAP, $\alpha > n/2$.

[5] We have here simplified the expression Joe (2006, theorem 1), which involved three products over three indices.

[6] The notation in Joe (2006) is $\alpha_{ij} = a + (n-1-j+i)/2$ and $\alpha = a + (n-2)/2$, which we have modified to make the resulting distribution directly comparable to the SAP method.



(a) Random Gram (b) SAP k=10 (c) Eigendecomposition (d) PAC (tiny)

Figure 4: Densities of pairwise correlations of dimension 3. The upper and bottom panels are corresponding to Bendel and Mickey's method with eigenvalues drawn from Exp(1) before normalization and random Gram matrix method. The left column shows the marginal distributions. The right column shows the bivariate densities of C_{21} and C_{31}.

4 Conclusion

We have proposed a new method for generating random correlation matrices, and compare it to some of the most commonly used methods.

The new methods provides a unified framework for generating random correlation matrices with a with range of properties.

References

Archakov, I. and Hansen, P. R. (2020), 'A generalized Fisher transformation for correlation matrices: A simulation study of its finite sample properties', *https://sites.google.com/site/peterreinhardhansen/* .

Archakov, I. and Hansen, P. R. (2021), 'A new parametrization of correlation matrices', *Econometrica* **89**, 1699–1715.

Archakov, I. and Hansen, P. R. (2022), 'A canonical representation of block matrices with applications to covariance and correlation matrices', *Forthcoming in Review of Economics and Statistics* .

Bendel, R. B. and Mickey, M. R. (1978), 'Population Correlation Matrices for Sampling Experiments', *Communications in Statistics - Simulation and Computation* **7**(2), 163–182.

Chalmers, C. (1975), 'Generation of correlation matrices with a given eigen–structure', *Journal of statistical computation and simulation* **4**(2), 133–139.

Davies, P. I. and Higham, N. J. (2000), 'Numerically stable generation of correlation matrices and their factors', *BIT Numerical Mathematics* **40**, 640–651.

Holmes, R. B. (1991), 'On Random Correlation Matrices', *SIAM Journal on Matrix Analysis and Applications* **12**(2), 239–272.

Hüttner, A. and Mai, J. F. (2019), 'Simulating realistic correlation matrices for financial applications: correlation matrices with the Perron-Frobenius property', *Journal of Statistical Computation and Simulation* **89**, 315–336.

Joe, H. (2006), 'Generating random correlation matrices based on partial correlations', *Journal of Multivariate Analysis* **97**, 2177–2189.

Linton, O. and McCrorie, J. R. (1995), 'Differentiation of an exponential matrix function: Solution', *Econometric Theory* **11**, 1182–1185.

Marsaglia, G. and Olkin, I. (1984), 'Generating correlation matrices', *SIAM Journal on Scientific Computing* **5**(2), 470–476.

Olkin, I. and Pratt, J. W. (1958), 'Unbiased estimation of certain correlation coefficients', *The Annals of Mathematical Statistics* **29**(1), 201 – 211.

Pinheiro, J. C. and Bates, D. M. (1996), 'Unconstrained parametrizations for variance-covariance matrices', *Statistics and Computing* **6**, 289–296.

Pourahmadi, M. (2011), 'Covariance estimation: The GLM and regularization perspectives', *Statistical Science* **26**, 369–387.

Pourahmadi, M. and Wang, X. (2015), 'Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor', *Statistics and Probability Letters* **106**, 5–12.
URL: *http://dx.doi.org/10.1016/j.spl.2015.06.015*

Tuitman, J., Vanduffel, S. and Yao, J. (2020), 'Correlation matrices with average constraints', *Statistics and Probability Letters* **165**, 1–13.

A Appendix of Proofs

Proof of Theorem 1. Let $G = \log C(\gamma)$ such that $\gamma = \mathrm{vecl}(G)$. If the elements of γ are given by $\gamma_i = \xi + \varepsilon_i$, with ε_i, $i = 1, \ldots, d$, independent and identically distribution and independent of ξ, then

γ and $\tilde{\gamma} = \text{vecl}(\tilde{G})$ have the same distribution where $\tilde{G} = \text{vecl}(PGP)$ for any permutation matrix, P. Consequently, $C = C(\gamma)$ and $\tilde{C} = C(\tilde{\gamma})$ have the same distribution, since the latter is given by,

$$\tilde{C} = \exp(\tilde{G}) = \exp(PGP') = P\exp(G)P' = PCP'.$$

Here we used that $\exp(P\log CP') = \exp(PQ\log\Lambda Q'P') = PQ\exp(\log\Lambda)Q'P'$ and that $Q'P'PQ = I$. □

Proof of Theorem 2. Since $r(z) = \frac{1-e^{-nz}}{1+(n-1)e^{-nz}}$ it follows that $r(z) \in (-\frac{1}{n-1}, 1)$. Next, we determine the expression for $f_z(z) = \left|\frac{\partial r(z)}{\partial z}\right| f_r(r(z))$, where

$$f_r(r) = \frac{1}{B(\alpha,\beta)} \frac{\left(\frac{1}{n-1}+r\right)^{\alpha-1}(1-r)^{\beta-1}}{\left(\frac{n}{n-1}\right)^{\alpha+\beta-1}} \times 1_{\{-\frac{1}{n-1}<r<1\}}.$$

Since $(n-1)e^{-nz} = e^{-\frac{z-\mu}{s}} = e^{-\zeta}$ where $\zeta = \frac{z-\mu}{s}$, we can write

$$r(z) = \frac{1-e^{-nz}}{1+(n-1)e^{-nz}} = \frac{1-\frac{1}{n-1}e^{-\zeta}}{1+e^{-\zeta}},$$

and

$$\begin{aligned}
\tfrac{1}{n-1} + r(z) &= \frac{1+e^{-\zeta}+(n-1)(1-\frac{1}{n-1}e^{-\zeta})}{(n-1)(1+e^{-\zeta})} = \frac{n}{n-1}\frac{1}{1+e^{-\zeta}}, \\
1 - r(z) &= \frac{1+e^{-\zeta}-(1-\frac{1}{n-1}e^{-\zeta})}{1+e^{-\zeta}} = \frac{e^{-\zeta}(1+\frac{1}{n-1})}{1+e^{-\zeta}} = \frac{n}{n-1}\frac{e^{-\zeta}}{1+e^{-\zeta}}.
\end{aligned}$$

Since $r(z) \in (-\frac{1}{n-1}, 1)$ is guaranteed, it follows that

$$f_r(r(z)) = \frac{1}{B(\alpha,\beta)} \frac{\left(\frac{1}{1+e^{-\zeta}}\right)^{\alpha-1}\left(\frac{e^{-\zeta}}{1+e^{-\zeta}}\right)^{\beta-1}}{\left(\frac{n}{n-1}\right)^{+1}} = \frac{1}{B(\alpha,\beta)} \frac{e^{-\beta\zeta}\left(e^{-\zeta}\right)^{-1}}{\frac{n}{n-1}\left(1+e^{-\zeta}\right)^{\alpha+\beta-2}}.$$

Next, the derivative is given by

$$\tfrac{\partial r(z)}{\partial z} = n^2 \frac{e^{-nz}}{(1+(n-1)e^{-nz})^2} = n\frac{n}{n-1}\frac{e^{-\zeta}}{1+e^{-\zeta}},$$

such that

$$
\begin{aligned}
f_z(z) &= n\frac{n}{n-1}\frac{e^{-\frac{z-\mu}{s}}}{1+e^{-\frac{z-\mu}{s}}}\frac{1}{B(\alpha,\beta)}\frac{e^{-\beta\frac{z-\mu}{s}}\left(e^{-\frac{z-\mu}{s}}\right)^{-1}}{\frac{n}{n-1}\left(1+e^{-\frac{z-\mu}{s}}\right)^{\alpha+\beta-2}} \\
&= \frac{1}{B(\alpha,\beta)}\frac{1}{s}\frac{e^{-\beta\frac{z-\mu}{s}}}{\left(1+e^{-\frac{z-\mu}{s}}\right)^{\alpha+\beta-1}},
\end{aligned}
$$

as stated. This completes the proof. □

Proof of Corollary 1. Follow from Theorem 2 by setting $\alpha = \beta = 1$, and it can also be verified directly that $f_z(z) = \frac{n-1}{n}\left|\frac{\partial r(z)}{\partial z}\right| = n(n-1)\frac{e^{-nz}}{(1+(n-1)e^{-nz})^2}$. □

B Expression for Determinant

We seek $\psi(C) = \det\frac{\partial\gamma}{\partial\varrho}$. Let $C = Q\Lambda Q'$ let $E_l \in \mathbb{R}^{d\times n^2}$, $E_u \in \mathbb{R}^{d\times n^2}$ and $E_d \in \mathbb{R}^{n\times n^2}$ be the elimination matrices that extract the lower-triangle, upper-triangle, or diagonal elements of an $n\times n$ matrix, i.e. $\mathrm{vecl}M = E_l\mathrm{vec}M$, $\mathrm{vecl}M' = E_u\mathrm{vec}M$ and $\mathrm{diag}M = E_d\mathrm{vec}M$ for any $M \in \mathbb{R}^{n\times n}$. From Archakov and Hansen (2021, proposition 3) we have that $\frac{\partial\varrho}{\partial\gamma} = E_l\Big(I - A_C E_d'\Big(E_d A_C E_d'\Big)^{-1}E_d\Big)A_C(E_l+E_u)'$, were $A_C = (Q\otimes Q)\Xi(Q\otimes Q)'$ and Ξ is the $n^2\times n^2$ diagonal matrix whose elements are given by

$$
\Xi_{(i-1)n+j,(i-1)n+j} = \xi_{ij} = \begin{cases} \lambda_i, & \text{if} \quad \lambda_i = \lambda_j, \\ \frac{\lambda_i-\lambda_j}{\log\lambda_i - \log\lambda_j}, & \text{if} \quad \lambda_i \neq \lambda_j, \end{cases} \tag{3}
$$

for $i = 1,\ldots,n$ and $j = 1,\ldots,n$. Note that A_C is symmetric and positive definite, because $\xi_{ij} > 0$ for all i,j. Here we have adapted the expression $\frac{\mathrm{dvec}\exp X}{\mathrm{dvec}X}$ in Linton and McCrorie (1995) to our context where $A_C = \frac{\mathrm{dvec}C}{\mathrm{dvec}\log C}$. It follows that

$$
\psi(C) = \frac{1}{\det\Big(E_l\Big(I - A_C E_d'\Big(E_d A_C E_d'\Big)^{-1}E_d\Big)A_C(E_l+E_u)'\Big)}.
$$