# A New Parametrization of Correlation Matrices*

**Ilya Archakov**[a] **and Peter Reinhard Hansen**[b†]

[a] *University of Vienna*

[b] *University of North Carolina & Copenhagen Business School*

*January 14, 2019*

**Abstract**

For the modeling of covariance matrices, the literature has proposed a variety of methods to enforce the positive (semi) definiteness. In this paper, we propose a method that is based on a novel parametrization of the correlation matrix, specifically the off-diagonal elements of the matrix logarithmic transformed correlations. This parametrization has many attractive properties, a wide range of applications, and may be viewed as a multivariate generalization of Fisher's $Z$-transformation of a single correlation.

*Keywords:* Covariance Modeling, Covariance Regularization, Fisher Transformation, Multivariate GARCH, Stochastic Volatility.

*JEL Classification:* C10; C22; C58

*We are grateful for many valuable comments made by Immanuel Bomze, Bo Honoré, Ulrich Müller, Georg Pflug, Werner Ploberger, Rogier Quaedvlieg, and Christopher Sims, as well as many conference and seminar participants.

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

# 1 Introduction

In the modeling of covariance matrices, it is often advantageous to parametrize the model so that the parameters are unrestricted. The literature has proposed several methods to this end, and most of these methods impose additional restrictions beyond the positivity requirement. Only a handful of methods ensure positive definiteness without imposing additional restrictions on the covariance matrix, see Pinheiro and Bates (1996) for a discussion of five parameterizations of this kind.

In this paper, we propose a new way to parametrize the covariance matrix that ensures positive definiteness without imposing additional restrictions, and we show that this parametrization has many attractive properties. The central element of the parametrization is the matrix logarithmic transformation of the correlations matrix, $\log C$, specifically the lower (or equivalently upper) off-diagonal elements of this matrix, which we denote by $\gamma = \gamma(C)$. We show that this transformation is an isomorphic mapping between the set of $n \times n$ non-singular correlation matrices and $\mathbb{R}^{n(n-1)/2}$, and we propose a fast algorithm for the computation of the inverse mapping. In the bivariate case $\gamma(C)$ is identical to the Fisher transformation, and we document that $\gamma(C)$ also has attractive properties in higher dimensional cases, so that $\gamma(C)$ may be viewed as the generalization of Fisher's $Z$-transformation beyond the bivariate setting.

Our theoretical results have many applications for the modeling of covariance matrices. First, a non-singular $n \times n$ covariance matrix can be expressed as a unique vector in $\mathbb{R}^{n(n+1)/2}$, which consists of the $n$ log-variances and $\gamma$. This facilitates the modeling of covariance matrices in terms of an unrestricted vector in $\mathbb{R}^{n(n+1)/2}$, where additional structure can be imposed, if so desired. Second, for models with dynamic covariance matrices, such as multivariate GARCH models and stochastic volatility models, the parametrization offers a new way to structure multivariate volatility models. Third, the representation of the correlation matrix facilitates a novel approach to regularizing large covariance matrices. We find, in a classical setting, that $\sqrt{T}[\gamma(\hat{C}) - \gamma(C)]$ is approximately normally distributed with nearly uncorrelated elements. This is a desirable property for several regularization methods, including James-Stein shrinkage, and this property could be valuable for hypothesis testing. Fourth, the representation is also convenient in applications where random draws of a covariance matrix are needed. For instance, the transformation yields a new and simple way to formulate Bayesian priors for correlation matrices.

It is convenient to reparametrize a covariance matrix as an unrestricted vector for several of reasons. For instance, the distributional properties of estimates obtained with constrained optimization tend to be more involved than those of unrestricted estimates, and constrained optimization tend to be computationally burdensome. The literature has therefore proposed a number of ways to parameterize a

covariance matrix in terms of an unrestricted vector. These methods include the Cholesky decomposition, the spherical trigonometric transformation, transformations based on partial correlation vines, see Kurowicka and Cooke (2003), and various related methods. For instance, that of Pourahmadi (1999), which is based on the Cholesky decomposition and the matrix logarithmic transformation of the covariance matrix which is related to the spectral representation. The matrix logarithm has been used in the modeling of covariance matrices in Leonard and Hsu (1992) and Chiu et al. (1996). In dynamic volatility models it was used in Kawakatsu (2006) (multivariate GARCH), Ishihara et al. (2016) (stochastic volatility), and Asai and So (2015) (Dynamic Conditional Correlations). Moreover, Bauer and Vorkink (2011) used the matrix logarithm for modeling and forecasting of realized covariance matrices. The transformation also emerges as a special case of Box-Cox transformations, see Weigand (2014) for the application to realized covariance matrices.

We do not apply the matrix logarithm to the covariance matrix. Instead we will apply it to the correlation matrix, while the variances can be modeled separately, for instance by taking the logarithm to each of the variances. A key variable in our analysis is the vector $\gamma = \gamma(C)$, which consists of the lower off-diagonal elements of $\log C$, where $C$ is the non-singular correlation matrix. We will show that mapping from $C$ to $\gamma$ is one-to-one, so that the set of non-singular correlation $n \times n$ matrices is isomorphic with $\mathbb{R}^{n(n-1)/2}$. We show that the correlation matrix can be reconstructed from $\gamma$ alone, and propose a fast algorithm to this end.[1] In the special case where $C$ is a $2 \times 2$ matrix, the off-diagonal element of $\log C$ is the Fisher transformation of a single correlation, and for higher dimensional correlation matrices, the finite sample distribution of the vector $\gamma(\hat{C})$ is well approximated by a Gaussian distribution under standard regularity conditions.

Modeling the correlation matrix separately from the individual variances is a common approach, and the new parametrization may be useful in this context. This modeling approach is quite common in multivariate GARCH models, and is the underlying structure in the Constant Conditional Correlations model by Bollerslev (1990), the Dynamic Conditional Correlations model by Engle (2002), see also Tse and Tsui (2002), and the Dynamic Equicorrelation model of Engle and Kelly (2011). The new parametrization can be used to define a new family of multivariate GARCH models, that need not impose additional restrictions beyond positivity. However, structure can be imposed, if so desired, and we detail some examples of this kind in Section 5.

Our results can also be used in dynamic models of multivariate volatility that make use of realized measures of volatility. Models in this area include those of Liu (2009) and Chiriac and Voev (2011), that

[1] Code for this algorithm (Matlab, Ox, and Python) is available in the Web Appendix.

models the elements of realized covariance matrix with simple time series models, the more sophisticated variants by Golosnoy et al. (2012) and Bauwens et al. (2012), as well as the models by Noureldin et al. (2012), Hansen et al. (2014), Dumitrescu and Hansen (2017) and Gorgi et al. (2018), that jointly model the vector of returns and their corresponding realized covariance measures.

The paper is organized as follows. We introduce and motivate the new parametrization of correlation matrices, $\gamma(C)$, in Section 2, using the simplest case with a $2\times 2$ covariance matrix as the starting point. We present the main theoretical results in Section 3, and in Section 4 we highlight useful distributional properties of the new parametrization. We derive the asymptotic distribution of $\gamma(\hat{C})$, and simulations show that the finite sample distribution of $\gamma(\hat{C})$ is well approximated by a Gaussian distribution with weakly correlated elements. We present several auxiliary results in Section 5. For instance, we show that certain structures in correlation matrices result in interesting structures in $\gamma(C)$, and how the parametrization can be used to specify priors on correlation matrices. We present the algorithm for computing $\gamma^{-1}$ in Section 6 and study its speed of convergence. We conclude and summarize in Section 7. All proofs are given in the Appendix, and additional results are collected in a Web Appendix, see Archakov and Hansen (2018).

## 2 Motivation

In this section, we motivate the proposed method by considering the simple case with a $2\times 2$ covariance matrix. In this setting, there is a convenient way to represent the covariance matrix, which involves the Fisher transformed correlation. Having explored the $2\times 2$ case we proceed with a discussion of two distinct ways of generalizing this methods to covariance matrices of higher dimensions, and highlight the advantages of the transformation proposed in this paper.

Consider a non-singular $2\times 2$ covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \bullet & \sigma_2^2 \end{pmatrix},$$

with correlation $\rho = \sigma_{12}/(\sigma_1\sigma_2) \in (-1,1)$. The mapping of $\Sigma$, $v = \nu(\Sigma) = (\log\sigma_1, \log\sigma_2, \frac{1}{2}\log\frac{1+\rho}{1-\rho})'$, is an example of a vector-representation of $\Sigma$, for which any covariance matrix maps to a unique vector in $\mathbb{R}^3$, and any vector $v\in\mathbb{R}^3$ maps to a unique non-singular covariance matrix using $(\sigma_1,\sigma_2,\rho) = (e^{v_1}, e^{v_2}, (e^{2v_3}-1)/(e^{2v_3}+1))$. This establishes that the set of non-singular $2\times 2$ covariance matrices is isomorphic with $\mathbb{R}^3$. Moreover, each element of the vector $v$ is easily interpretable because they relate

directly to the original parameters, $\sigma_1$, $\sigma_2$, and $\rho$. This approach therefore circumvents the need for parameter restrictions, as would be required if one instead modeled $(\sigma_1^2, \sigma_2^2, \sigma_{12})$ or $(\sigma_1^2, \sigma_2^2, \rho)$. Notice that the last element, $v_3$, is the well known Fisher transformation of the correlation, $F(\rho) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$. This approach to modeling a bivariate process was used in Hansen et al. (2014) who proposed a multivariate Realized GARCH model. In their model, the vector $v$ is driven by a vector autoregressive model, and the inverse transformation guarantees a non-singular (conditional) covariance matrix in every period.

For a parametrization of a covariance matrix, the following properties are desireble:

1. Any covariance matrix, $\Sigma$, maps to a unique vector $v = \nu(\Sigma) \in \mathbb{R}^d$
2. Any vector $v \in \mathbb{R}^d$ maps to a unique covariance matrix $\Sigma = \nu^{-1}(v)$.
3. The parametrization, $v = \nu(\Sigma)$, is "invariant" to the ordering of the variables that define $\Sigma$.

For practical implementation it will also be desirable that both $\nu(\cdot)$ and $\nu^{-1}(v)$ are computationally "simple", ideally in closed-form, and that the elements of $v$ are easily interpretable.

For a $2 \times 2$ covariance matrix the parametrization, $v = (\log\sigma_1, \log\sigma_2, \frac{1}{2}\log\frac{1+\rho}{1-\rho})'$, has all of the above properties, which is not the case for other methods. For instance, the Cholesky representation is not invariant to the ordering of variables, nor are the resulting elements easily interpretable. The transformation based on the matrix logarithm of the covariance matrix, $\log\Sigma$, has the three properties listed above, but the resulting elements are more difficult to interpret, because they are non-linear functions of all variances and all covariances. An interesting question is whether the method for $2 \times 2$ matrices, which is based on the Fisher transformation, can be generalized to higher-dimensional covariance matrices. And if so, how many of the above properties can be preserved.

As a way to generalize the transformation beyond $2 \times 2$ matrices, we first consider the simple approach, where the Fisher transformation is applied to each of the correlations. This idea was explored in Dumitrescu and Hansen (2017). A drawback of this approach is that the element-wise Fisher transformed correlations do not vary freely in $\mathbb{R}^d$ as can be demonstrated with a $3 \times 3$ with correlation matrix,

$$C = \begin{pmatrix} 1 & \bullet & \bullet \\ \rho_{21} & 1 & \bullet \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}.$$

The three Fisher transformed correlations can be represented by the vector $\phi = (\phi_1, \phi_2, \phi_2)' = (F(\rho_{21}), F(\rho_{31}), F(\rho_{32}))'$, but these do not vary freely in $\mathbb{R}^3$. This follows from the simple observation that

$\tilde{\phi} = (-2, 0, \frac{1}{2})$ does not correspond to a valid correlation matrix, for the simple reason that

$$\det \begin{pmatrix} 1 & \bullet & \bullet \\ F^{-1}(-2) & 1 & \bullet \\ F^{-1}(0) & F^{-1}(\frac{1}{2}) & 1 \end{pmatrix} = \det \begin{pmatrix} 1 & \bullet & \bullet \\ -0.964 & 1 & \bullet \\ 0 & 0.462 & 1 \end{pmatrix} \simeq -0.143 < 0.$$

So the element-wise Fisher transformation fails to meet the second objective listed above.

Returning to the case with a $2 \times 2$ correlation matrix. We observe that the Fisher transformation appears as the off-diagonal elements when we take the matrix-logarithm to a $2 \times 2$ correlation matrix:

$$\log \begin{pmatrix} 1 & \bullet \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\log(1-\rho^2) & \bullet \\ \frac{1}{2}\log\frac{1+\rho}{1-\rho} & \frac{1}{2}\log(1-\rho^2) \end{pmatrix}.$$

In this paper, we will argue that a natural extension of the Fisher transformation is defined by the off-diagonal elements of the matrix-logarithm of the correlation matrix, $G = \log C$. For an $n \times n$ correlation matrix this will result in $n(n-1)/2$ distinct off-diagonal elements. In Section 4, we study the asymptotic and finite sample properties of these off-diagonal elements, and find that the elements of this transformation are only weakly correlated, and well approximated by a Gaussian distribution in finite samples.

The transformation of an $n \times n$ covariance matrix, $\Sigma$, into $n$ log-variances and the $n(n-1)/2$ off-diagonal elements of $G = \log C$ satisfies the three objectives stated above, as we will show in Section 3. The computational aspect of this transformation is simplified by an algorithm that converges quickly, as we demonstrate in Section 6. Some elements of the transformation are one-to-one with the individual variances, and these elements are therefore easily interpretable. The remaining elements of the transformation – the off-diagonal elements of $G$ – are not simple to interpret individually, because they depend on all correlations in a nonlinear manner. However, interesting structures are preserved in $G = \log C$, for certain types of models, including covariance stationary time-series, where $C$ is a symmetric Toeplitz matrix, and models where $C$ has a (block) equi-correlation structure. We detail this in Section 5.1.

## 3 Theoretical Framework

In this Section, we present the main theoretical results. The most challenging step is to show the existence and uniqueness in Theorem 1, which establishes that the set of covariance matrices is isomorphic

with $\mathbb{R}^{n(n+1)/2}$. First we introduce the necessary notation.

### 3.1 Notation

The operator, diag$(\cdot)$, is used in two ways. When the argument is a vector, $v = (v_1, \ldots, v_n)'$, then diag$(v)$ denotes the $n \times n$ diagonal matrix with $v_1, \ldots, v_n$ along the diagonal, and when the argument is a square matrix, $A \in \mathbb{R}^{n \times n}$, then diag$(A)$ extracts the diagonal of $A$ and returns it as a column vector, i.e. diag$(A) = (a_{11}, \ldots, a_{nn})' \in \mathbb{R}^n$.

The matrix exponential is defined by $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$ for any matrix $A$. For a symmetric matrix, $A$, with eigendecomposition, $A = Q\Lambda Q'$, we have $e^A = Q\text{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})Q'$. Here $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, while $Q$ is an orthonormal matrix, i.e. $Q'Q = I$. The general definition of the matrix logarithm is more involved, but for a symmetric positive definite matrix, we have that $\log A = Q \log \Lambda Q'$, where $\log \Lambda = \text{diag}(\log \lambda_1, \ldots, \log \lambda_n)$.

We use vecl$(A)$ to denote the vectorization operator of the lower off-diagonal elements of $A$ (so this operator excludes the diagonal elements unlike the related vech$(\cdot)$ operator). We will use this operator to extract the off-diagonal elements of matrices. For a non-singular correlation matrix, $C$, we let $G = \log C$ denote the logarithmically transformed correlation matrix, and let $F$ be the matrix of element-wise Fisher transformed correlations (whose diagonal is unspecified). The vector of correlation is denoted by $\varrho = \text{vecl}C$, and the corresponding elements of $G$ and $F$ are denoted by $\gamma = \text{vecl}G$ and $\phi = \text{vecl}F$, respectively.

**Definition 1** (New Parametrization of Correlation Matrices)**.** For a non-singular correlation matrix, $C$, we introduce the following parametrization:

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

To illustrate the mapping, consider the case with a $3 \times 3$ matrix, where we have

$$G = \log C = \begin{pmatrix} G_{11} & G_{12} & G_{13} \\ G_{12} & G_{22} & G_{23} \\ G_{13} & G_{23} & G_{33} \end{pmatrix}, \quad \text{and} \qquad \gamma(C) = \begin{pmatrix} G_{12} \\ G_{13} \\ G_{23} \end{pmatrix}.$$

So $\gamma(C)$ discards the diagonal elements of $\log C$. A relevant question is whether one can reconstruct a correlation matrix from an arbitrary vector $\gamma$, and, if affirmative, whether the reconstructed correlation matrix is unique. In the situation with a $3 \times 3$ correlation matrix, the question is whether there, for

any vector $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$, exists a vector, $x(\gamma) = (x_1, x_2, x_3)'$, so that

$$\exp \begin{pmatrix} x_1 & \gamma_1 & \gamma_2 \\ \gamma_1 & x_2 & \gamma_3 \\ \gamma_2 & \gamma_3 & x_3 \end{pmatrix},$$

is a correlation matrix, and the second question is whether the solution, $x(\gamma)$, is unique.

To formalize the problem of reconstructing a correlation matrix from an arbitrary vector $\gamma$, we introduce the following operator that replaces the diagonal of a matrix. For an $n \times n$ matrix, $A$, and any vector $x \in \mathbb{R}^n$ we let $A[x]$ denote the matrix $A$ where $x$ has replaced its diagonal. So it follows that $\text{vecl}(A) = \text{vecl}(A[x])$ and that $x = \text{diag}(A[x])$. For instance, in the $3 \times 3$ case we have

$$A[x] = \begin{pmatrix} x_1 & A_{12} & A_{13} \\ A_{21} & x_2 & A_{23} \\ A_{31} & A_{32} & x_3 \end{pmatrix}.$$

The interesting case, in the present context, is the case where the off-diagonal elements of $A$ are given by the elements of $\gamma$.

### 3.2 Main Theoretical Results

First we state the key result that $\gamma(C)$ is an isomorphic mapping between $\mathcal{C}_n$ and $\mathbb{R}^{n(n-1)/2}$, where $\mathcal{C}_n$ denotes the set of non-singular correlation matrices. An implication is that any vector in $\mathbb{R}^{n(n-1)/2}$ is mapped into a unique correlation matrix. This is a desirable property, because it eliminates concerns about positive definiteness, which is automatically guaranteed.

**Theorem 1.** *For any real symmetric matrix, $A \in \mathbb{R}^{n \times n}$, there exists a unique vector, $x^* \in \mathbb{R}^n$, such that $e^{A[x^*]}$ is a correlation matrix.*

So any vector in $\mathbb{R}^{n(n-1)/2}$ maps to a unique correlation matrix, and since any correlation matrix maps to $\gamma = \gamma(C) \in \mathbb{R}^{n(n-1)/2}$, we have establishes a one-to-one correspondence between $\mathcal{C}_n$ and $\mathbb{R}^{n(n-1)/2}$. It is possible that the results could be generalized to cover singular correlation matrices by extending the domain of $\gamma$ to include $\pm\infty$, but we do not explore this possibility in this paper.

The proof of Theorem 1 is given in the Appendix, but we outline the structure of the proof here, because it provides the intuition behind the algorithm that is used to reconstruct a correlation matrix

from an arbitrary vector, $\gamma$, of proper dimension.[2]

From the properties of the matrix exponential, it follows that $e^{A[x]}$ will be positive definite for any vector, $x$. A solution is therefore characterized by the identity

$$\operatorname{diag}(e^{A[x^*]}) = \iota, \tag{1}$$

where $\iota = (1, \ldots, 1)'$ is the vector of ones. Since (1) is a nonlinear matrix equation, involving $n$ equations with $n$ unknowns (the elements of $x^*$), it may have a unique solution. This is, however, not self-evident, because (1) is a system of nonlinear equations. Theorem 1 settles this issue and informs us that for any symmetric matrix $A$, a solution exists and the solution is unique.

The proof is based on the following mapping $g : \mathbb{R}^n \curvearrowright \mathbb{R}^n$,

$$g(x) = x - \log \operatorname{diag}(e^{A[x]}),$$

where the vector $\log \operatorname{diag}(e^{A[x]})$ is a vector of zeros if $e^{A[x]}$ is a correlation matrix. So the requirement is that $g(x^*) = x^*$, so that $x^*$ is a fixed-point for $g$. Hence, Theorem 1 is equivalent to the statement that a fixed-point exists and is unique for any matrix $A$. This, in turns, follows by showing the following result and applying Banach fixed-point theorem.

**Lemma 1.** *The mapping $g$ is a contraction for any symmetric matrix $A$.*

The proof of Lemma 1, which is given in the appendix, entails deriving the Jacobian for $g$ and showing that all its eigenvalues are less than one in absolute value.

### 3.3 Invariance to Reordering of Variables

The mapping, $\gamma(C)$, is invariant to the reordering of variables that define $C$, in the sense that a permutation of the variables that define $C$ will merely result in a permutation of the elements of $\gamma$. The formal statement is as follows.

**Proposition 1.** *Suppose that $C_x = \operatorname{corr}(X)$ and $C_y = \operatorname{corr}(Y)$, where the elements of $X$ is a permutation of the elements of $Y$. Then the elements of $\gamma_x = \gamma(C_x)$ is a permutation of the elements of $\gamma_y = \gamma(C_y)$.*

[2] If $\gamma \in \mathbb{R}^d$ then we need $d = n(n-1)/2$ for some integer $n$, i.e., we need $\frac{1}{2}(1 + \sqrt{1 + 8d})$ to be an integer.

### 3.4 An Algorithm for Reconstructing the Correlation Matrix

Evidently, the solution, $x^*$, must be such that $\log \operatorname{diag}(e^{A[x^*]}) = 0 \in \mathbb{R}^n$. This observation motivates the following iterative procedure for determining $x^*$:

**Corollary 1.** *Consider the sequence,*

$$x_{(k+1)} = x_{(k)} - \log \operatorname{diag}(e^{A[x_{(k)}]}), \qquad k = 0, 1, 2, \ldots$$

*with an arbitrary initial vector $x_{(0)} \in \mathbb{R}^n$. Then $x_{(k)} \to x^*$, where $x^*$ is the solution in Theorem 1.*

It is possible that the algorithm could be improved further by using the Jacobian for $g$, which is derived in the Appendix. But in practice we find that the simple algorithm, proposed in Corollary 1, converges very fast. This is demonstrated in Section 6 for matrices with dimension up to $n = 100$.

## 4 $\hat{\gamma}$ is Approximately Gaussian with Weakly Correlated Elements

In this section, we study the asymptotic and finite sample properties of $\hat{\gamma} = \gamma(\hat{C})$ in a classical setting and compare them with those of $\hat{\varrho}$ and $\hat{\phi}$. First we derive their limit distributions in a classical setting, then we study the finite sample properties using simulation methods. The overall conclusion is that $\hat{\gamma}$ is well approximated by a Gaussian distribution with weakly correlated elements.

In our analysis we can take the correlation matrix $C$ to be the covariance matrix, because of a certain scale invariance. Suppose that $\Sigma = \operatorname{var}(X)$ and $C = \operatorname{corr}(X)$ then $C$ is also the correlation matrix of $DX$, if $D = \operatorname{diag}(d_1, \ldots, d_n)$, $d_j \neq 0$ for all $j = 1, \ldots, n$. Consequently, $\hat{C}$, and hence $\hat{\varrho}$, $\hat{\phi}$, and $\hat{\gamma}$, have the same scale invariance, and we can therefore, without loss of generality, focus on the case where the diagonal elements of $\Sigma$ are all one, i.e. $\Sigma = C$.

To illustrate certain results we will often make use the following Toeplitz structure for the correlation matrix,

$$C = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}, \tag{2}$$

where $\rho \in (-1, 1)$ regulates the degree of correlation. The variables are uncorrelated when $\rho = 0$, and

the pairwise correlation increases as $\rho$ increases from 0 to 1. The asymptotic distributions derived next do not rely on any particular structure for $C$.

### 4.1 Asymptotic Properties

In this section we derive the asymptotic distributions of $\hat{\phi}$ and $\hat{\gamma}$ by deducing them from those of $\hat{C}$. Thus we consider a situation where the empirical correlation matrix is such that $\sqrt{T}(\hat{C}-C) \xrightarrow{d} N(0,\Omega)$, as $T \to \infty$, for some asymptotic covariance matrix, $\Omega = \text{avar}(\text{vec}(\hat{C}))$. Evidently, $\Omega$ is a reduced rank matrix, because $\hat{C}$ is symmetric and has ones along the diagonal. A convenient closed-form expression for $\Omega$ can be obtained under certain assumptions, such as that in Neudecker and Wesselman (1990), see also Nel (1985) and Browne and Shapiro (1986).

For $\hat{\varrho} = \text{vecl}(\hat{C})$ it follows directly that

$$\sqrt{T}\Big(\hat{\varrho} - \varrho\Big) \xrightarrow{d} N\Big(0,\ E_l \Omega E_l'\Big), \quad \text{as} \quad T \to \infty, \tag{3}$$

where $E_l$ is an elimination matrix, characterized by $\text{vecl}[M] = E_l \text{vec}[M]$ for any square matrix $M$ (of dimension $n \times n$).

The asymptotic distributions of the element-wise Fisher transformed correlations, $\hat{\phi}$, and the new parametrization, $\hat{\gamma} = \gamma(\hat{C})$, can be derived using (3) and the delta method.

For the element-wise Fisher transform, the asymptotic distribution reads

$$\sqrt{T}\Big(\hat{\phi} - \phi\Big) \xrightarrow{d} N\Big(0,\ E_l D_c \Omega D_c E_l'\Big), \tag{4}$$

where $D_c = \text{diag}\Big(\frac{1}{1-c_i^2}, \frac{1}{1-c_2^2}, \ldots, \frac{1}{1-c_{n^2}^2}\Big)$ and $c_i$ is an $i$-th element of $c = \text{vec}(C)$, whereas the asymptotic distribution of the new parametrization of $C$ takes the following form,

$$\sqrt{T}\Big(\hat{\gamma} - \gamma\Big) \xrightarrow{d} N\Big(0,\ E_l A^{-1} \Omega A^{-1} E_l'\Big), \tag{5}$$

where $A$ is a Jacobian matrix, such that $d\text{vec}(C) = A\, d\text{vec}\big(\log(C)\big)$. The expression for $A$ is given in the Appendix, see (A.3)-(A.4), and is taken from Linton and McCrorie (1995).

In a classical setting where $\hat{C}$ is computed from i.i.d. Gaussian distributed random variables, the diagonal elements of the asymptotic variance matrix in (4) will all be one, as this is one of the characteristic of the Fisher transformation. This is not quite the case for the asymptotic variance matrix of $\hat{\gamma}$, but the expression (5) defines ways to modify $\hat{\gamma}$, for instance by scaling the elements of $\hat{\gamma}$ to have unit variance. We do not pursue this possibility in this paper.

The two expressions for the asymptotic variances, $\text{avar}(\hat{\phi}) = E_l D_c \Omega D_c E_l'$ and $\text{avar}(\hat{\gamma}) = E_l A^{-1} \Omega A^{-1} E_l'$, are not easily compared in general, but once $\Omega$ is specified it becomes straight forward. Here we will compare them in the situation where $\widehat{C}$ is computed from $X_t \sim \text{iid} N_3(0, C)$, and $C$ has the Toeplitz structure in (2). We present the asymptotic variances and correlations of the vectors, $\hat{\varrho}$, $\hat{\phi}$ and $\hat{\gamma}$, for the cases where $\rho = \{0, 0.5, 0.9, 0.99\}$.

| $C$ | $\text{avar}(\hat{\varrho})$ | $\text{avar}(\hat{\phi}) =$ $\text{acorr}(\hat{\varrho}) = \text{acorr}(\hat{\phi})$ | $\text{avar}(\hat{\gamma})$ | $\text{acorr}(\hat{\gamma})$ |
|---|---|---|---|---|
| $\begin{pmatrix} 1 & \bullet & \bullet \\ 0 & 1 & \bullet \\ 0 & 0 & 1 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0 & 1.000 & \bullet \\ 0 & 0 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0 & 1.000 & \bullet \\ 0 & 0 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0 & 1.000 & \bullet \\ 0 & 0 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0 & 1.000 & \bullet \\ 0 & 0 & 1.000 \end{pmatrix}$ |
| $\begin{pmatrix} 1 & \bullet & \bullet \\ 0.5 & 1 & \bullet \\ 0.25 & 0.5 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0.562 & \bullet & \bullet \\ 0.316 & 0.879 & \bullet \\ 0.070 & 0.316 & 0.562 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.450 & 1.000 & \bullet \\ 0.125 & 0.450 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 0.966 & \bullet & \bullet \\ 0.018 & 0.962 & \bullet \\ 0.021 & 0.018 & 0.966 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.018 & 1.000 & \bullet \\ 0.021 & 0.018 & 1.000 \end{pmatrix}$ |
| $\begin{pmatrix} 1 & \bullet & \bullet \\ 0.9 & 1 & \bullet \\ 0.81 & 0.9 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0.036 & \bullet & \bullet \\ 0.046 & 0.118 & \bullet \\ 0.015 & 0.046 & 0.036 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.698 & 1.000 & \bullet \\ 0.405 & 0.698 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 0.817 & \bullet & \bullet \\ 0.081 & 0.860 & \bullet \\ 0.093 & 0.081 & 0.817 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.097 & 1.000 & \bullet \\ 0.114 & 0.097 & 1.000 \end{pmatrix}$ |
| $\begin{pmatrix} 1 & \bullet & \bullet \\ 0.99 & 1 & \bullet \\ 0.98 & 0.99 & 1 \end{pmatrix}$ | $\frac{1}{10}\begin{pmatrix} 0.004 & \bullet & \bullet \\ 0.006 & 0.016 & \bullet \\ 0.002 & 0.006 & 0.004 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.745 & 1.000 & \bullet \\ 0.490 & 0.745 & 1.000 \end{pmatrix}$ | $\begin{pmatrix} 0.756 & \bullet & \bullet \\ 0.106 & 0.793 & \bullet \\ 0.134 & 0.106 & 0.756 \end{pmatrix}$ | $\begin{pmatrix} 1.000 & \bullet & \bullet \\ 0.137 & 1.000 & \bullet \\ 0.178 & 0.137 & 1.000 \end{pmatrix}$ |

Table 1: Asymptotic covariance and correlation matrices for $\hat{\varrho}$, $\hat{\phi}$ and $\hat{\gamma}$, (3)-(5) for the case where $C$ has the Toeplitz structure in (2) with $\rho = 0$, $\rho = 0.5$, $\rho = 0.9$, and $\rho = 0.99$. The diagonal elements of the asymptotic variance matrix for $\hat{\phi}$ are all one, so it is also the asymptotic correlation matrix for $\hat{\phi}$. Because $\hat{\phi}$ is an element-by-element transformation of the corresponding elements of $\hat{\phi}$, it is also the asymptotic correlation matrix for $\hat{\varrho}$.

The asymptotic variance and correlation matrices are reported in Table 1. The asymptotic variance of the correlation coefficient, $\hat{\varrho}_j$, is $(1 - \varrho_j^2)^2$, which defines the diagonal elements of $\text{avar}(\hat{\varrho})$, and the element-wise Fisher transformation ensures that $\text{avar}(\hat{\phi}_j) = 1$ for all $j = 1, \ldots, n$. However, we observe a high degree of correlation across the elements of $\hat{\phi}$. The asymptotic correlation matrix for $\hat{\phi}$ is, in fact, identical to that of the empirical correlations, $\hat{\varrho}$, because the transformation is an element-by-element (Fisher) transformation, causing the Jacobian $D_c = \text{d}\phi/\text{d}\varrho$ to be a diagonal matrix. Consequently the asymptotic correlations are unaffected by this transformation. While the diagonal elements of $\text{avar}(\hat{\phi})$ are invariant to $C$, this is not quite the case for the diagonal elements of $\text{avar}(\hat{\gamma})$. However, the diagonal elements are not nearly as sensitive as is the case for $\text{avar}(\hat{\varrho})$, and the main advantage of $\hat{\gamma}$ is that the asymptotic covariances across elements are relatively small, in particular when $|\rho| \leq 0.5$.

### 4.2 Finite Sample Properties

We study the finite sample properties in a classical setting where $\hat{C}$ is computed from the sample covariance matrix, $\hat{\Sigma} = \frac{1}{T}\sum_t^T (X_t - \bar{X})(X_t - \bar{X})'$, with $\bar{X} = \frac{1}{T}\sum_{t=1}^T X_t$, and where $X_t \sim iidN(0, \Sigma)$. As discussed earlier, we can without loss of generality consider the case where the diagonal elements of $\Sigma$ are all one, i.e. $\Sigma = C$.

The Fisher transformation is also known as Fisher's Z-transform. In standard problems, where $\hat{\rho}$ is the empirical correlation coefficient, the Fisher transformation is, in part, motivated by $F(\hat{\rho}) - F(\rho)$ being better approximated by a Gaussian distribution than is $\hat{\rho} - \rho$.

In this section, we undertake a simulation study, where we analyze the finite sample properties of $\hat{\gamma}$, and compare them to those of the element-wise Fisher transformed correlations. First, we simulate a trivariate system, where $X_i \sim iidN_3(0, \Sigma)$, $i = 1, \ldots, T$ where $\Sigma = C$ is given as in (2). Here we will present results for the case where $T = 40$ and $\rho = 0.9$. Additional results, based on different designs, are presented in the Web Appendix. In this design, $\hat{\varrho}$, $\hat{\phi}$, and $\hat{\gamma}$ are 3-dimensional, because the marginal distributions of the three elements within each of the vectors are very similar, we only present results for the first element of these vectors, $\hat{\varrho}_1$, $\hat{\phi}_1$, and $\hat{\gamma}_1$. Results for all three elements are presented in the Web Appendix.

Figure 1 shows the finite sample distributions for $\hat{\varrho}_1$, $\hat{\phi}_1$, and $\hat{\gamma}_1$, for the case where $T = 40$, and $C$ is given from (2) with $\rho = 0.9$. The left panels illustrates the well known results that the finite sample distribution of the empirical correlation is poorly approximated by a Gaussian distribution, in particular in a situation such as this one, where the population value, $\varrho_1 = 0.9$, is relatively close to one. The middle panels present the analogous results for the element-wise Fisher transformed correlation, $\hat{\phi}_1$, and the right panels the results for $\hat{\gamma}_1$. The finite sample distribution of both $\hat{\phi}_1$ and $\hat{\gamma}_1$ are well approximated by the Gaussian distribution. This was clearly expected for the Fisher transformation, $\hat{\phi}_1$, but somewhat unexpected for $\hat{\gamma}_1$. In fact, the QQ-plots, including several reported in the Web Appendix, indicate that the Gaussian approximation is slightly better for $\hat{\gamma}$ than for $\hat{\phi}$. A plausible explanation for this discrepancy is that the individual Fisher transformed correlations are subject to cross restrictions.

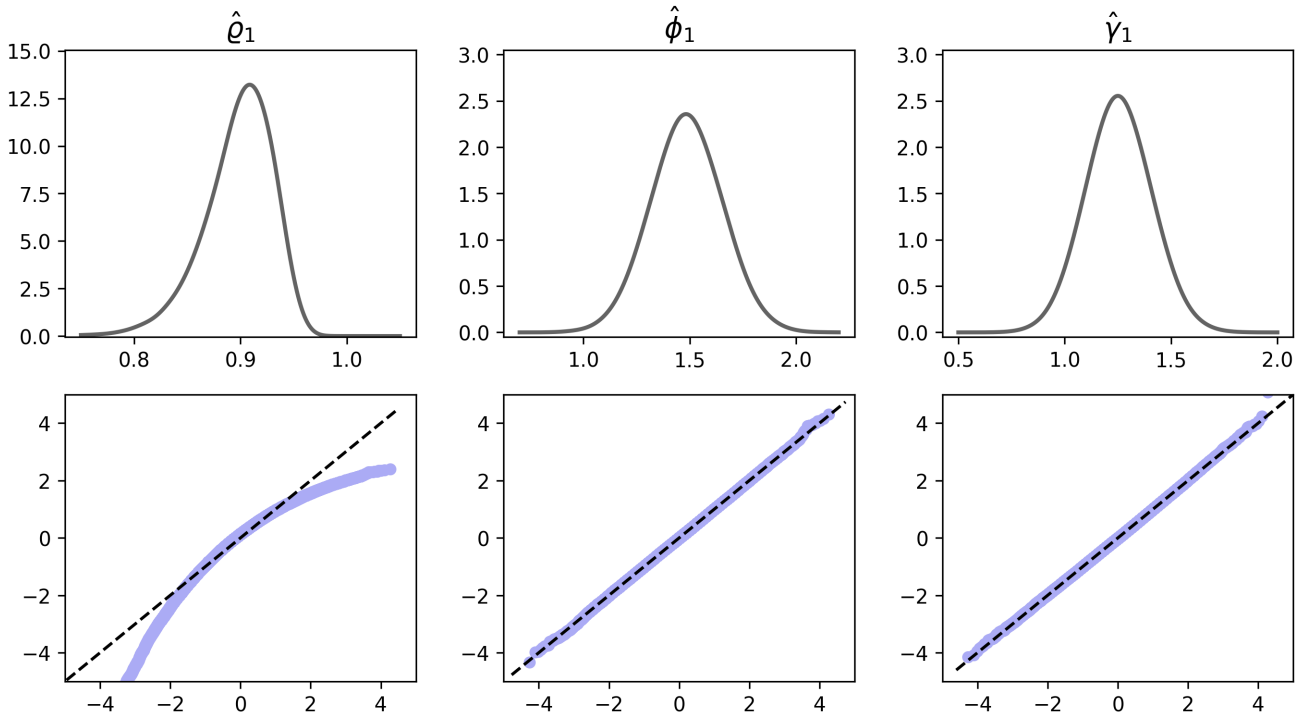


Figure 1: The finite sample distribution (upper) and QQ-plots (lower) of the first elements of the vectors $\hat{\varrho}$, $\hat{\phi}$, and $\hat{\gamma}$. Results for the marginal distributions of $\hat{\varrho}_1$, $\hat{\phi}_1$, and $\hat{\gamma}_1$, are in the left, middle, and right panels, respectively. The simulation design has $\hat{C}$ computed from $T = 40$ independent vectors that are distributed as $N_3(0, C)$, where $C$ has the structure in (2) with $\rho = 0.9$. The results are based on 100,000 simulations. The QQ-plots are the quantiles of the standardized empirical distribution plotted against quantiles of the standard normal distribution.

Having established that the marginal distribution of $\hat{\gamma}$ is well approximated by a Gaussian distribution, in line with the properties of $\hat{\phi}$, we turn our attention to features of the joint distributions. Using the same simulation design as in Figure 1 we now present the three bivariate distributions that emerge from the three empirical correlations, the corresponding three Fisher transformed correlations, and the three elements of $\hat{\gamma}$. Contour plots (non-parametrically estimated) for these bivariate distributions are given in Figure 2. The dependence between empirical correlations is evident, and this cross dependence carries over to the element-wise Fisher transformed correlations. In contrast, the dependence between the elements of $\hat{\gamma}$ is much weaker. The contour plots for the bivariate distributions of $\hat{\gamma}$ resemble those of a bivariate Gaussian distribution with little correlation between the two variables. This highlights an unexpected benefit of the proposed transformation which appears to hold more generally, as suggested by results from additional simulation designs that are presented in the Web Appendix. In the Web Appendix, we show that $\hat{\gamma}$ greatly reduces the finite sample skewness, as is the case for the marginal distributions of the elements of $\hat{\phi}$.

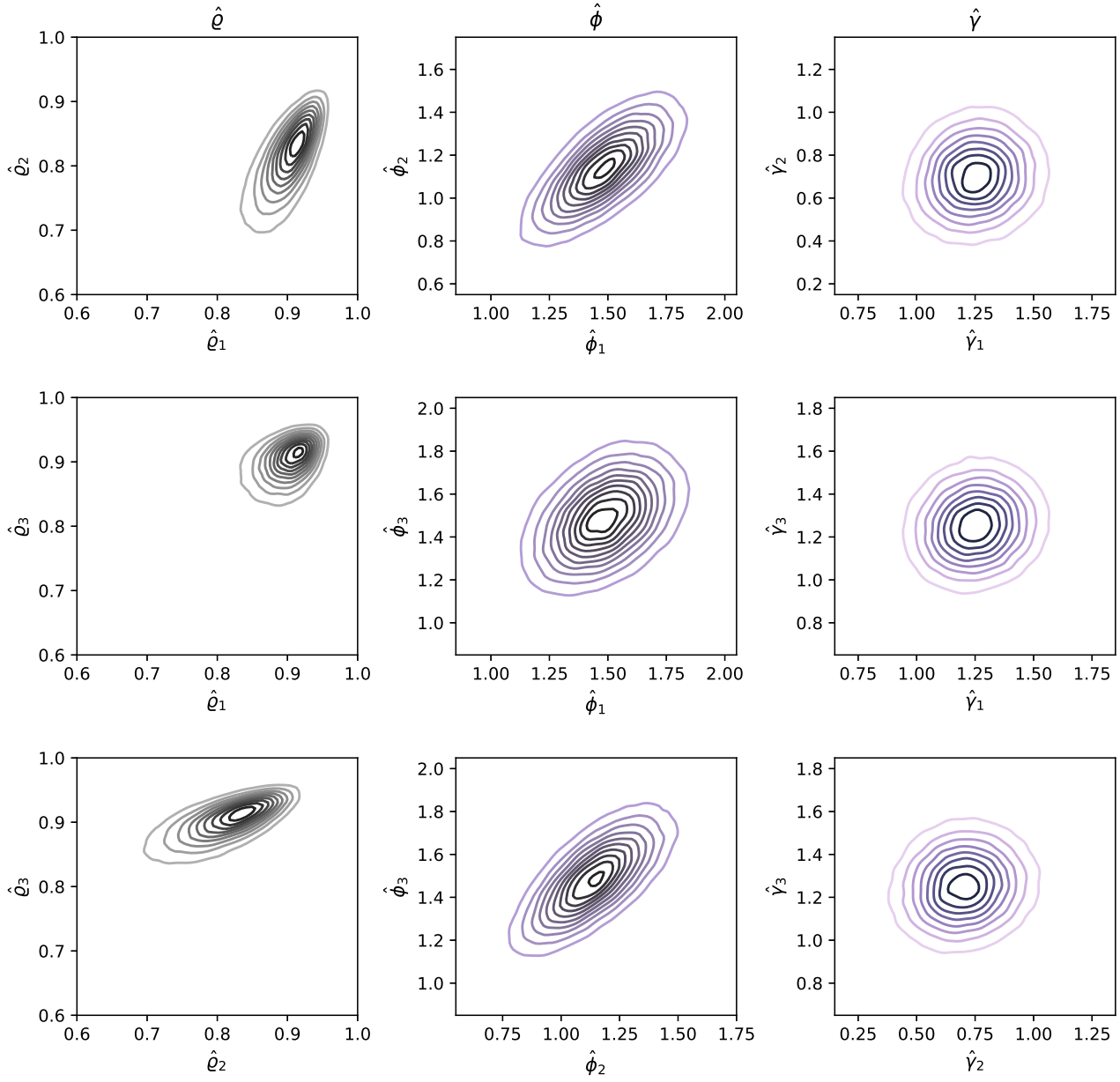


Figure 2: Contour plots of the bivariate distributions. The left panels are those for the empirical correlations, $\hat{\varrho}$, the middle panels are for the element-wise Fisher transformed correlations, $\hat{\phi}$, and the right panels are those for new parametrization of the correlation matrix, $\hat{\gamma}$. The contour plots are based on 100,000 random draws, where each of the correlation matrices were based on $T = 40$ i.i.d. variables distributed as $N_3(0, C)$, where $C$ has the structure in (2) with $\rho = 0.9$.

So far we have focused on the case where $C$ is an $3 \times 3$ matrix. Next we study the finite sample dependence across elements of $\hat{\phi}$ and $\hat{\gamma}$ for matrices of higher dimensions, using the Toeplitz structure in (2), where the degree of dependence is controlled by the single parameter, $\rho$. We consider the finite

sample correlation matrices for $\hat{\phi}$ and $\hat{\gamma}$,

$$R_{\phi,T}(\rho) = \text{corr}(\hat{\phi}) \qquad \text{and} \qquad R_{\gamma,T}(\rho) = \text{corr}(\hat{\gamma}),$$

and will compute measures of dependence from $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$. One measure of dependence is the largest eigenvalue of the correlation matrix, $R$, because $\lambda_{\max}(R) = 1$ if and only if $R = I$. A natural measure of dependence for an $d \times d$ correlation matrix is given by

$$\psi(R) = (\lambda_{\max}(R) - 1)/(d - 1),$$

because $\lambda_{\max}(R) \simeq 1+(d-1)\bar{r}$, where $\bar{r}$ is the average correlation of $R$, see Morrison (1967), who derives the approximation when the correlations are non-negative and similar in value, and see Friedman and Weisberg (1981) for a more general interpretation.

We study the finite sample properties when $T = 100$ for correlation matrices of dimensions, $n = 10$, $n = 20$, and $n = 40$. In our simulation study we draw $T = 100$ observations from $N(0, C)$, where $C$ has the structure in (2) with $\rho$ ranging from 0 to 0.99. In each simulation we compute $\hat{\phi}$ and $\hat{\gamma}$, and compute their corresponding correlation matrices, $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$, from 10,000 simulations. The dimension of these correlation matrices are for $n = 10$, 20, and 40 given by $45 \times 45$, $190 \times 190$, and $780 \times 780$, respectively, since $d = n(n-1)/2$.

Besides reporting $\psi(R_{\phi,T}(\rho))$ and $\psi(R_{\gamma,T}(\rho))$ we also report their smallest and largest off-diagonal element, and the 10% and 90%-quantiles of the many correlations, to get a sense of the dispersion of the many elements in $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$.

The results are reported in Figure 3. Results for the element-wise Fisher transforms, $\hat{\phi}$, are in the left panels, and the corresponding results for the new parametrization, $\hat{\gamma}$, are in the right panels. The solid lines present the eigenvalue-based measure of dependence, $\psi(R_{\phi,T}(\rho))$ and $\psi(R_{\gamma,T}(\rho))$, as a function of $\rho$. The lightly shaded regions display the range of correlations in $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$ from smallest to largest, whereas the darker shaded regions are defined by the 10%-quantile and 90%-quantile of the off-diagonal elements in $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$.

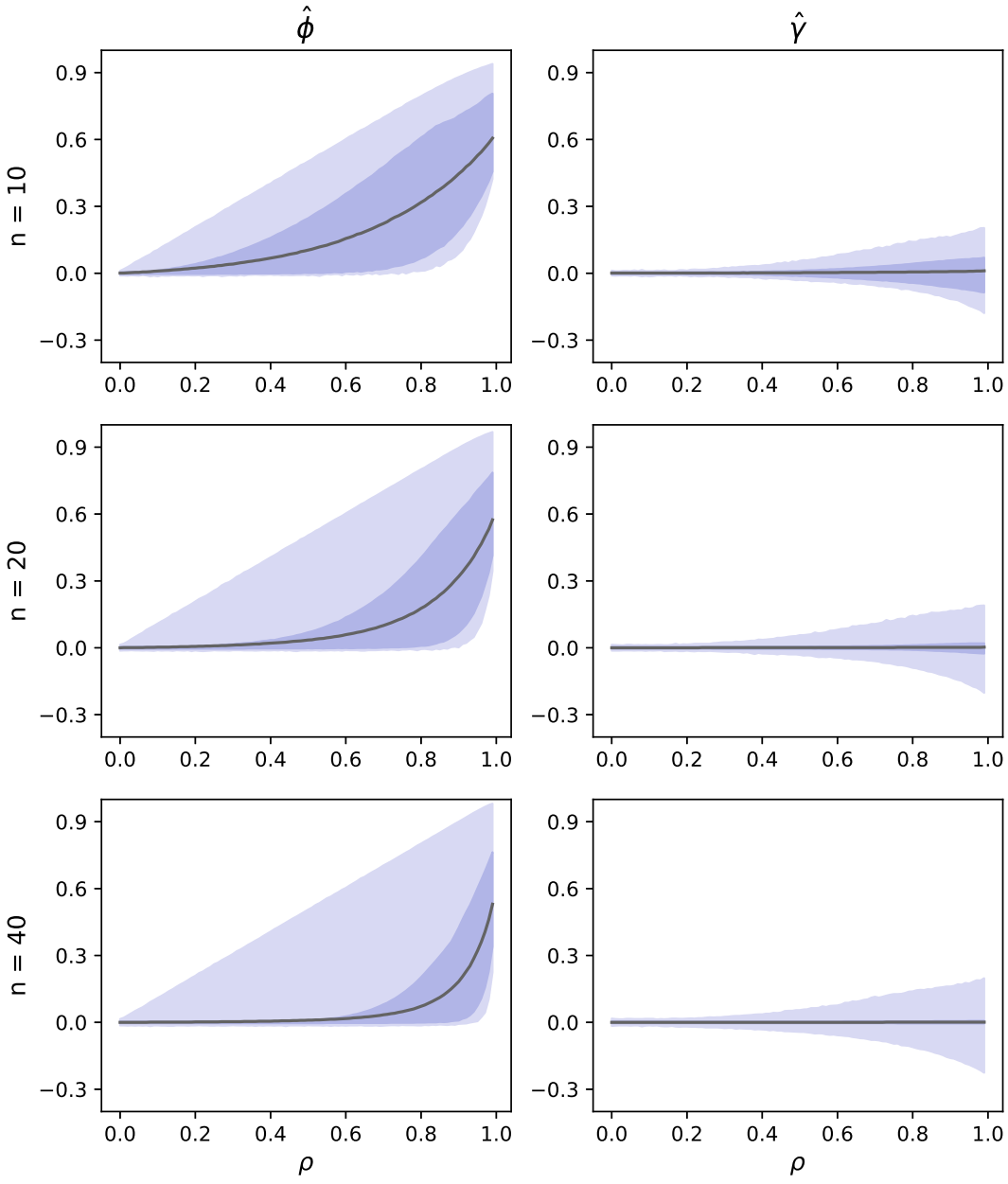


Figure 3: The solid line is the eigenvalue-based measure of dependence in $R_{\phi,T}(\rho)$ (left panels) and in $R_{\gamma,T}(\rho)$ (right panels), as a function of $\rho$. The upper, middle, and lower panels are for the case where $\hat{C}$ is an $10 \times 10$, $20 \times 20$, and $40 \times 40$ respectively. The lightly shaded regions display the full range of correlations, whereas the darker shaded regions are defined by the 10%-quantile and 90%-quantile of the off-diagonal elements in $R_{\phi,T}(\rho)$ and $R_{\gamma,T}(\rho)$. Evidently, the elements of $\hat{\gamma}$ are far less correlated than are the element-wise Fisher correlations, $\hat{\phi}$.

Figure 3 also shows that there is far less dependence across the elements of $\hat{\gamma}$ than is the case for the element-wise Fisher transformed variables, $\hat{\phi}$, for all values of $\rho$. While the correlations between elements of $\hat{\phi}$ are substantially different from zero when $\rho$ approaches one, the correlations between elements of $\hat{\gamma}$ are small and centered about zero. The darker shaded regions include 80% of the

correlations, and Figure 3 shows that the large majority of correlations for $\hat{\gamma}$ are very close to zero, even as $\rho$ approaches 1, where as the dispersions for correlations related to $\hat{\phi}$ is far larger. Interestingly, for this design, we see that the correlations between elements of $\hat{\gamma}$ are increasingly concentrated near zero, as the dimension of $C$ increases. The darker shaded region is barely visible for $\hat{\gamma}$ when $n = 40$. For both $\hat{\phi}$ and $\hat{\gamma}$ we observe that the extreme correlations, (the smallest and largest), are similar for $n = 10$, $n = 20$, and $n = 40$.

A potential use of the weak correlation between the elements of $\hat{\gamma}$ is to the problem of regularization of large covariance matrices, see Pourahmadi (2011) for a review. The observation that the finite sample distribution of $\hat{\gamma}$ seems well approximated by a vector of nearly independent Gaussian random variables, paves the way for applying shrinkage methods, such as James-Stein shrinkage to $\hat{\gamma}$. This approach has been applied to the element-wise Fisher transformations, $\hat{\phi}$, Lin and Perlman (1985). Unreported simulation results indicate that the elements of $\log \hat{\Sigma}$ have similar properties in the homoskedastic case, where the diagonal elements of $\Sigma$ are identical, so that regularization may also be applied directly to $\operatorname{vech} \log \hat{\Sigma}$, as explored in Deng and Tsui (2013).

# 5 Auxiliary Results and Properties

## 5.1 Structure for Certain Correlation Matrices

While the elements of $\gamma$ depend on the correlation matrix in a nonlinear way, there are some interesting correlation structures that do carry over to the matrix $G = \log C$, and hence $\gamma$. First, we consider the case with an equicorrelation matrix and a block-equicorrelation matrix.

**Proposition 2.** *Suppose $C$ is a positive definite equicorrelation matrix with correlation parameter $\rho$. Then, all the off-diagonal elements of matrix $G = \log C$ are identical and equal to*

$$\gamma_c = -\frac{1}{n} \log\left( \frac{1-\rho}{1+(n-1)\rho} \right) \in \mathbb{R}, \tag{6}$$

*so that $\gamma = \gamma_c \iota$, where $\iota \in \mathbb{R}^{n(n-1)/2}$ is the vector of ones.*

This result, in conjunction with Theorem 1, establishes that $\gamma_c$ is an isomorphic mapping from the set of equicorrelation matrices to the real line, $\mathbb{R}$, and the inverse mapping (to the common correlation) is given in closed-form by

$$\rho(\gamma_c, n) = \frac{1 - e^{-n\gamma_c}}{1 + (n-1)e^{-n\gamma_c}}.$$

Note that this yields a correlation coefficient, $\rho(\gamma_c, n)$, that is confined to the interval $(-\frac{1}{n-1}, 1)$, which is to be expected since this is the range for $\rho$ that produce a positive definite equicorrelation matrix.

It is easy to verify that if $C$ is a block diagonal matrix, with equicorrelation diagonal blocks and zero correlation across blocks, then $G = \log C$ will have the same block structure and (6) can be used to compute the elements in $\gamma$. In the more general case where $C$ is a block equicorrelation matrix, see e.g. Engle and Kelly (2011), then it can be shown that the logarithmic transformation preserves the block structure, so that $A$ has the same block structure with the same coefficient in each of the blocks. So the transformation provides a simple way to model block equicorrelation matrices. We illustrate this with the following example

$$C = \begin{pmatrix} 1.0 & 0.4 & 0.4 & 0.2 & 0.2 & 0.2 \\ 0.4 & 1.0 & 0.4 & 0.2 & 0.2 & 0.2 \\ 0.4 & 0.4 & 1.0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1.0 & 0.6 & 0.6 \\ 0.2 & 0.2 & 0.2 & 0.6 & 1.0 & 0.6 \\ 0.2 & 0.2 & 0.2 & 0.6 & 0.6 & 1.0 \end{pmatrix} \Leftrightarrow \log C = \begin{pmatrix} -.16 & .349 & .349 & .104 & .104 & .104 \\ .349 & -.16 & .349 & .104 & .104 & .104 \\ .349 & .349 & -.16 & .104 & .104 & .104 \\ .104 & .104 & .104 & -.36 & .553 & .553 \\ .104 & .104 & .104 & .553 & -.36 & .553 \\ .104 & .104 & .104 & .553 & .553 & -.36 \end{pmatrix}.$$

Another interesting class of correlation matrices are the Toeplitz-correlation matrices, which arises in various models, such as autoregressive time series models. In this case, $\log C$ is a bisymmetric matrix. To take an example:

$$C = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix} \Leftrightarrow \log C = \begin{pmatrix} * & a & b & c \\ a & * & d & b \\ b & d & * & a \\ c & b & a & * \end{pmatrix}.$$

For $k+1 \times k+1$ Toeplitz correlation matrices, one can represent the matrix using just $k$ off-diagonal elements of $\log C$. In the example above, that would amount to $\gamma = (a, b, c)'$, and a correlation matrix with the proper Toeplitz structure is then obtain by selecting an appropriate value of $d$, in addition to the diagonal elements.

### 5.2 Bayesian Priors on Correlation and Covariance Matrices

The literature has proposed a number of methods to formulate Bayesian priors on correlation matrices, see Pourahmadi (2011, section 4) for a review. The new parametrization of correlation matrices provides

a new method for formulating priors over correlation matrices. The results in Section 4 could motivate the use of a Gaussian prior on $\gamma$, which induces a distribution over $C(\gamma)$.

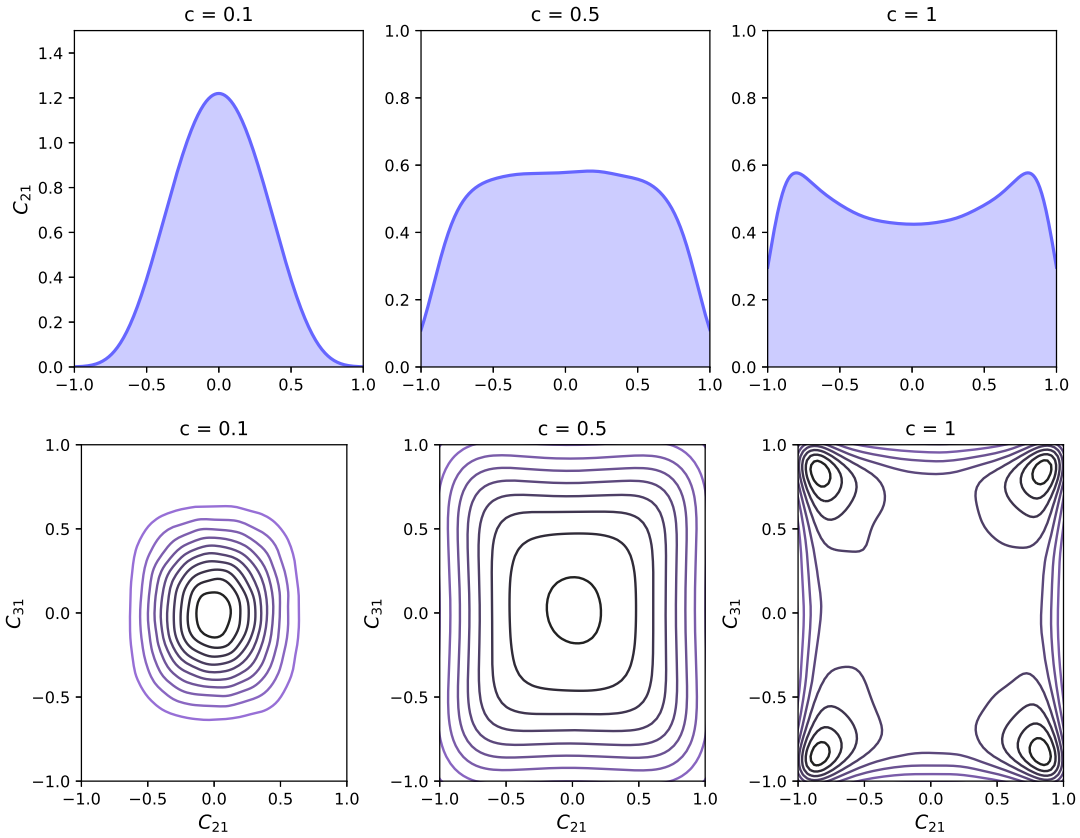


Figure 4: The upper panel displays the marginal distribution of correlation coefficients in a $3 \times 3$ matrix, $C(\gamma)$, as induced by $\gamma \sim N_3(0, cI)$, for $c = 0.1$, $c = 0.5$, and $c = 1$.

### 5.3 The Inverse and other Powers of the Correlation Matrix

While $G = \log C$ is directly tied to the correlation matrix, $C$, it is straight forward to obtain other powers of $C$ from $A$, since $C^\alpha = e^{\alpha G}$. This can, for example, be used to obtain the inverse covariance matrix,

$$\Sigma^{-1} = \Lambda^{-1} e^{-G} \Lambda^{-1},$$

where $\Lambda = \text{diag}(\sigma_1, \ldots, \sigma_n)$. The inverse is, for instance, of interest in portfolio choice problems and $\Sigma^{-1}$ also yields the partial correlation coefficients.

### 5.4 Useful Structure for Inference and Dynamic Models

Next we establish a result that shows that $d\text{vecl}[C]/d\text{vecl}[G]$ has a relatively simple expression. This is a convenient for inference, such as computation of standard errors, and for the construction of dynamic

GARCH-type models, such as a score-driven model for $\gamma = \text{vecl}G$, see Creal et al. (2013), and for the construction of parameter stability tests, such as that of Nyblom (1989).

**Proposition 3.** *Derivatives of the correlations from $C$ with respect to the off-diagonal elements of the log-transformed correlation matrix $G = \log C$ are*

$$\frac{d\text{vecl}[C]}{d\text{vecl}[G]} = E_l\Big(I - AE_d'\Big(E_dAE_d'\Big)^{-1}E_d\Big)A(E_l + E_u)',$$

*where $A = d\text{vec}C/d\text{vec}G$ and the matrices $E_l$, $E_u$ and $E_d$ are elimination matrices, such that* $\text{vecl}M = E_l\text{vec}M$, $\text{vecl}M' = E_u\text{vec}M$ *and* $\text{diag}M = E_d\text{vec}M$ *for any square matrix $M$ of the same size as $C$ and $G$.*

Here $A$ is the same matrix that appears in the asymptotic distribution for $\hat{\gamma}$, see (5), and its expression is given in the Appendix, see (A.3)-(A.4).

# 6 Algorithm for $\gamma^{-1}$

In this section, we study how fast the iterative algorithm converges to the solution. The algorithm reconstructs the correlation matrix $C$ from $\gamma = \text{vecl}G$, by determining the diagonal elements of $G$, for which $e^G$ is a correlation matrix. We find that the algorithm converges fast, even for high dimensional matrices. Interestingly, the number of iterations required for convergence depends mainly on the correlation structure, and to a lesser extent on the matrix size.

We have the following helpful result that informs us that none of the diagonal elements of $G$ are positive.

**Lemma 2.** *The diagonal elements of $G = \log C$ are non-positive.*

Given $n$ and $\rho$ we construct $C$ as in (2) and compute $G = \log C$. The diagonal of $G$ is then replaced with some starting value, $x_{(0)}$. The algorithm recovers $G$, starting from the matrix $G[x_{(0)}]$ by applying the iterative algorithm described in Corollary 1. In our simulation experiment we set $x_{(0)}$ to be random with each element being drawn independently and set to $-|Z|$ where $Z \sim N(0, 100)$. So the starting value will often far from the solution. The algorithm has converges once all diagonal elements of $e^{G[x_{(k)}]}$ are sufficiently close to zero. Since $-\log\text{diag}(e^{G[x_{(k-1)}]}) = x_{(k)} - x_{(k-1)}$ we can simply define the algorithm to have converged (at iteration $k$) once $||x_{(k)} - x_{(k-1)}||$ is less than the selected tolerance threshold, where $\| z \| = \sqrt{\frac{z'z}{n}}$. We set the tolerance threshold to $10^{-8}$. For each value of $\rho$ and $n$, we run the algorithm with 1000 distinct (random) starting values.

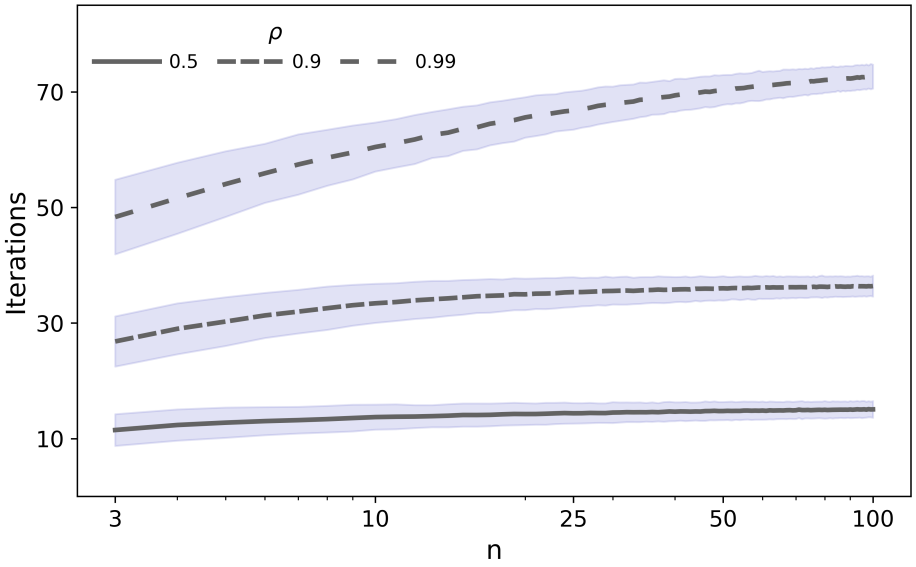


Figure 5: Convergence of the iterative algorithm with a tolerance threshold $10^{-8}$. The black lines correspond to the average number of iterations required for convergence. The bands correspond to $\pm 2$ standard deviations.

The average numbers of iterations required for convergence are depicted in Figure 5. Interestingly, the number of iterations needed for convergence, only increases modestly as the dimension of the matrix increases. The shaded bands depict the dispersion in the number of iterations needed for convergence (average $\pm 2$ standard deviations). The dispersion is rather modest, which shows that the algorithm is relatively insensitive to the choice of an initial vector, $x^{(0)}$. The number of iterations is more influenced by the actual correlation structure. For correlation matrices with highly correlated variables ($\rho = 0.99$), it takes 50-70 iterations for the algorithm to converge. In contrast, for moderately correlated structures ($\rho = 0.5$), the algorithm converges in 10-15 iterations.

In the Web Appendix, we present convergence results for the same design with weaker convergence tolerance thresholds. For instance, using the threshold of $10^{-4}$ (instead of $10^{-8}$) reduces the number of required iterations by a factor of about two.

Thus, the proposed algorithm appears to converge sufficiently fast for most practical problems even if the dimension is large. On average, it took about 0.07 seconds for the algorithm to converge when $C$ is a $100 \times 100$ matrix with the Toeplitz structure, where $\rho = 0.99$ and tolerance requirement is $10^{-8}$. For a less exotic correlation structure ($\rho = 0.5$) and a relatively small-scale matrix ($n = 5$), the average convergence time is only about $3.7 \cdot 10^{-4}$ seconds.[3]

The algorithm has been implemented in Matlab, Python, and Ox.

[3]The execution times have been obtained using MATLAB R2017a on a computer with Intel Core i7-6700 (3.40GHz).

# 7 Concluding Remarks

In this paper, we have shown that the set of non-singular $n \times n$ correlation matrices is isomorphic with $\mathbb{R}^{n(n-1)/2}$, so that one can represent a covariance matrix in terms of the $n$ (log-)variances and $n(n-1)/2$ off-diagonal elements of $\log C$, where $C$ is the correlation matrix, which we denoted by $\gamma$. The latter is closely related to the Fisher transformation of a single correlation. The reason is that the two coincide when $C$ is an $2 \times 2$ matrix, and the finite sample properties of $\hat{\gamma}$ are well approximated by a Gaussian distribution. Unlike the element-wise Fisher transformations, $\phi$, the new parametrization produces a vector, $\gamma$, with unrestricted range.

The new parametrization of the correlation matrix, presented in this paper, adds an item to the menu of methods for modeling correlation matrices, where the new method has a unique set of properties. Key advantages include a variation-free parametrization, the ability to model variances and correlations separately, and attractive distributional properties of the transformed correlation matrix, $\gamma(\hat{C})$. The main drawback of the new method is that $\gamma^{-1}$ is not given in closed-form, albeit this is mitigated by the proposed algorithm that converges very fast, even for high-dimensional matrices. In the paper, we have highlighted many possible application of the new parametrization, such as the construction of multivariate volatility models, inference, regularization of correlation matrices, and specifying Bayesian priors over correlation matrices.

# References


Archakov, I. and Hansen, P. R. (2018), 'Web-appendix for 'a new parametrization of correlation matrices", *https://sites.google.com/site/peterreinhardhansen/* .

Asai, M. and So, M. (2015), 'Long memory and asymmetry for matrix-exponential dynamic correlation processes', *Journal of Time Series Econometrics* **7**, 69–74.

Bauer, G. H. and Vorkink, K. (2011), 'Forecasting multivariate realized stock market volatility', *Journal of Econometrics* **160**, 93–101.

Bauwens, L., Storti, G. and Violante, F. (2012), 'Dynamic conditional correlation models for realized covariance matrices', *Working Paper* (2012060).

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model', *The Review of Economics and Statistics* **72**, 498–505.

Browne, M. and Shapiro, A. (1986), 'The asymptotic covariance matrix of sample correlation coefficients under general conditions', *Linear Algebra and its Applications* **82**, 169 – 176.

Chiriac, R. and Voev, V. (2011), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* **26**, 922–947.

Chiu, T., Leonard, T. and Tsui, K.-W. (1996), 'The matrix-logarithmic covariance model', *Journal of the American Statistical Association* **91**, 198–210.

Creal, D. D., Koopman, S. J. and Lucas, A. (2013), 'Generalized autoregressive score models with applications', *Journal of Applied Econometrics* **28**, 777–795.

Deng, X. and Tsui, K.-W. (2013), 'Penalized covariance matrix estimation using a matrix-logarithm transformation', *Journal of Computational and Graphical Statistics* **22**, 494–512.

Dumitrescu, E. and Hansen, P. (2017), 'Forecasting exchange rate volatility: Multivariate realized GARCH framework', *Working Paper* .

Engle, R. (2002), 'Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. and Kelly, B. (2011), 'Dynamic Equicorrelation', *Journal of Business & Economic Statistics* **30**(2), 212–228.

Friedman, S. and Weisberg, H. F. (1981), 'Interpreting the first eigenvalue of a correlation matrix', *Educational and Psychological Measurement* **41**, 11–21.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012), 'The conditional autoregressive Wishart model for multivariate stock market volatility', *Journal of Econometrics* **167**(1), 211–223.

Gorgi, P., Hansen, P. R., Janus, P. and Koopman, S. J. (2018), 'Realized Wishart-GARCH: A score-driven multi-asset volatility model', *Journal of Financial Econometrics* p. forthcoming.

Hansen, P. R., Lunde, A. and Voev, V. (2014), 'Realized beta GARCH: A multivariate GARCH model with realized measures of volatility', *Journal of Applied Econometrics* **29**, 774–799.

Ishihara, T., Omori, Y. and Asai, M. (2016), 'Matrix exponential stochastic volatility with cross leverage', *Computational Statistics & Data Analysis* **100**, 331–350.

Kawakatsu, H. (2006), 'Matrix exponential GARCH', *Journal of Econometrics* **134**, 95–128.

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.

Leonard, T. and Hsu, J. S. J. (1992), 'Bayesian inference for a covariance matrix', *Annals of Statistics* **20**, 1669–1696.

Lin, S. P. and Perlman, M. D. (1985), A monte carlo comparison of four estimators of a covariance matrix, *in* 'Multivariate Analysis VI', North-Holland, Amsterdam, pp. 411–429.

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

Liu, Q. (2009), 'On portfolio optimization: How and when do we benefit from high-frequency data?', *Journal of Applied Econometrics* **24**(4), 560–582.

Morrison, D. F. (1967), *Multivariate Statistical Methods*, McGraw-Hill, New York.

Nel, D. (1985), 'A matrix derivation of the asymptotic covariance matrix of sample correlation coefficients', *Linear Algebra and its Applications* **67**, 137 – 145.

Neudecker, H. and Wesselman, A. (1990), 'The asymptotic variance matrix of the sample correlation matrix', *Linear Algebra and its Applications* **127**, 589 – 599.

Noureldin, D., Shephard, N. and Sheppard, K. (2012), 'Multivariate high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics* **27**, 907–933.

Nyblom, J. (1989), 'Testing for the constancy of parameters over time', *Journal of the American Statistical Association* **84**, 223–230.

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

Pourahmadi, M. (1999), 'Joint mean-covariance models with applications to longitudinal data', *Biometrika* **86**, 677–690.

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

Tse, Y. K. and Tsui, A. K. C. (2002), 'A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business and Economic Statistics* **20**, 351–362.

Weigand, R. (2014), 'Matrix Box-Cox models for multivariate realized volatility', *Working Paper* .

# Appendix of Proofs

We prove $g$ is a contraction by deriving its Jacobian, $J(x)$, and showing that all its eigenvalues are less than one in absolute value. To this end we observe that

$$g(x) = x - \log \delta(x), \qquad \text{where} \quad \delta(x) = \operatorname{diag}(e^{G[x]}). \tag{A.1}$$

Hence, an intermediate step towards the Jacobian for $g$, is to derive the Jacobian for $\delta(x)$.

To simplify notation, we will sometimes suppress the dependence on $x$ for some terms. For instance, we will write $\delta_i$ to denote the $i$-th element of the vector $\delta(x)$.

**The Jacobian of $g(x)$**

We seek the Jacobian, $\mathrm{d}g(x) = J(x)\mathrm{d}x$, and from (A.1) it follows that

$$[J(x)]_{i,j} = \frac{\partial [g(x)]_i}{\partial x_j} = 1_{\{i=j\}} - \frac{1}{\delta_i}\frac{\partial [\delta(x)]_i}{\partial x_j},$$

so that

$$J(x) = I - [D(x)]^{-1}H(x), \tag{A.2}$$

where $D = \mathrm{diag}(\delta_1, \ldots, \delta_n)$ is a diagonal matrix and $H(x)$ is the Jacobian matrix of $\delta(x)$, that we derive below.

Let $G[x] = Q\Lambda Q'$, where $\Lambda$ is the diagonal matrix with the eigenvalues, $\lambda_1, \ldots, \lambda_n$, of $G[x]$ and $Q$ is an orthonormal matrix (i.e. $Q' = Q^{-1}$) with the corresponding eigenvectors. It is well known, see e.g. Linton and McCrorie (1995), that

$$\mathrm{dvec}\, e^{G[x]} = A(x)\, \mathrm{dvec} G[x],$$

where

$$A(x) = (Q \otimes Q)\Xi(Q \otimes Q)', \tag{A.3}$$

is and $n^2 \times n^2$ matrix and $\Xi$ is the $n^2 \times n^2$ diagonal matrix with elements given by

$$\xi_{ij} = \Xi_{(i-1)n+j,(i-1)n+j} = \begin{cases} e^{\lambda_i}, & \text{if} \quad \lambda_i = \lambda_j \\ \frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j}, & \text{if} \quad \lambda_i \neq \lambda_j \end{cases} \tag{A.4}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n$. Evidently we have $\xi_{ij} = \xi_{ji}$, for all $i$ and $j$. Importantly, it also follows that $A(x)$ is a symmetric positive definite matrix, because all the diagonal elements of $\Xi$ are strictly positive.

Our interest concerns $\mathrm{diag}[e^{G[x]}]$ which is only a subset of the elements of $\mathrm{vec}[e^{G[x]}]$. So we seek, the matrix $H(x)$,

$$\mathrm{d}\,\delta(x) = H(x)\,\mathrm{d}x,$$

where $H(x)$ is a principal sub-matrix of $A(x)$, which is obtained by preserving $n$ rows and columns of $A$ with symmetric indices $(i-1)n+i$, $i = 1, \ldots, n$. A generic element of matrix $H(x)$ is therefore given by:

$$
\begin{aligned}
h_{ij} = \frac{\partial \delta(x)_i}{\partial x_j} &= (e_i \otimes e_i)'(Q \otimes Q)\Xi(Q \otimes Q)'(e_j \otimes e_j) \\
&= (e_i'Q \otimes e_i'Q)\Xi(Q'e_j \otimes Q'e_j) = (Q_{i,\cdot} \otimes Q_{i,\cdot})\Xi(Q_{j,\cdot} \otimes Q_{j,\cdot})' \\
&= \sum_{k=1}^{n}\sum_{l=1}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl},
\end{aligned} \tag{A.5}
$$

where $e_i$ is a $n \times 1$ unit vector with one at the $i$-th position and zeroes otherwise and $Q_{i,\cdot}$ denotes the $i$-th row of $Q$.

An interesting property of $J(x)$ is that $J(x)\iota = 0$, so that the vector of ones, $\iota$, is an eigenvector of $J(x)$ associated with the eigenvalue 0, i.e. $J(x)$ has reduced rank. Because, the $i$-th row of $J(x)$ times $\iota$ reads

$$
1 - \sum_{j=1}^{n}\frac{1}{\delta_i}\sum_{k=1}^{n}\sum_{l=1}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl} = 1 - \frac{1}{\delta_i}\sum_{k=1}^{n}\sum_{l=1}^{n} q_{ik}q_{il}\xi_{kl}\sum_{j=1}^{n} q_{jk}q_{jl} = 1 - \frac{1}{\delta_i}\sum_{k=1}^{n} q_{ik}^2\xi_{kk} = 0,
$$

due to $\sum_{k=1}^{n} q_{ik}q_{jk} = 1_{\{i=j\}}$.

## Proof that $g$ is a Contraction: Lemma 1

We now want to prove that the mapping $g(x)$ is a contraction. In order to show this, it is sufficient to demonstrate that all eigenvalues of the corresponding Jacobian matrix $J(x)$ are below one in absolute values for any real vector $x$. First we establish a number of intermediate results.

**Lemma A.1.** *(i)* $e^y - y - 1 > 0$ *for all* $y \neq 0$*, and (ii)* $1 + e^y - \frac{2}{y}(e^y - 1) > 0$ *for* $y \neq 0$*.*

*Proof.* The first and second derivatives of $f(y) = e^y - y - 1$ show that $f$ is strictly convex with unique minimum, $f(0) = 0$, which proves $(i)$. Next we prove $(ii)$. Now let $f(y) = 1 + e^y - \frac{2}{y}(e^y - 1)$. Its first derivative is given by $f'(y) = e^y y^{-2} g(y)$, where $g(y) = y^2 - 2y + 2 - 2e^{-y}$, so that $f'(y) < 0$ for $y < 0$ and $f'(y) > 0$ for $y > 0$. Since $\lim_{y\to 0} f(y) = 0$ (by l'Hospital's rule) the result follows. □

From the definition, (A.4), it follows that $\xi_{ij} = \xi_{ii} = \xi_{jj}$ whenever $\lambda_i = \lambda_j$. When $\lambda_i \neq \lambda_j$ we have the following results for the elements of $\Xi$:

**Lemma A.2.** *If* $\lambda_i < \lambda_j$*, then* $\xi_{ii} < \xi_{ij} < \xi_{jj}$ *and*

$$
2\xi_{ij} < \xi_{ii} + \xi_{jj}.
$$

*Proof.* From the definition, (A.4), we have

$$\xi_{ij} - \xi_{ii} = \frac{e^{\lambda_j} - e^{\lambda_i}}{\lambda_j - \lambda_i} - e^{\lambda_i} = e^{\lambda_i}\left(\frac{e^{\lambda_j - \lambda_i} - 1}{\lambda_j - \lambda_i} - 1\right) = e^{\lambda_i}\frac{e^{\lambda_j - \lambda_i} - 1 - (\lambda_j - \lambda_i)}{\lambda_j - \lambda_i} > 0,$$

because the numerator is positive by Lemma A.1.*i*, and so are $e^{\lambda_i}$ and $\lambda_j - \lambda_i$. This proves $\xi_{ij} > \xi_{ii}$. Analogously,

$$\xi_{jj} - \xi_{ij} = e^{\lambda_j} - \frac{e^{\lambda_j} - e^{\lambda_i}}{\lambda_j - \lambda_i} = e^{\lambda_j}(1 - \tfrac{1 - e^{\lambda_i - \lambda_j}}{\lambda_j - \lambda_i}) = e^{\lambda_j}\frac{-(\lambda_i - \lambda_j) - 1 + e^{\lambda_i - \lambda_j}}{\lambda_j - \lambda_i} > 0,$$

because all terms are positive, where we again used Lemma A.1.*i*.

Next, consider

$$\xi_{ii} + \xi_{jj} - 2\xi_{ij} = e^{\lambda_i} + e^{\lambda_j} - 2\frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} = e^{\lambda_i}\left(1 + e^{\lambda_j - \lambda_i} - 2\frac{e^{\lambda_j - \lambda_i} - 1}{\lambda_j - \lambda_i}\right) > 0,$$

where the inequality follows by Lemma A.1.*ii*, because $\lambda_i \neq \lambda_j$. □

**Lemma A.3.** *$J(x)$ and $\tilde{J}(x) = I - D^{-\frac{1}{2}}HD^{-\frac{1}{2}}$ have the same eigenvalues, and $\tilde{J}(x)$ can be expressed as*

$$\tilde{J} = \sum_{k=1}^{n-1}\sum_{l=k}^{n}\varphi_{kl}\Big(D^{-\frac{1}{2}}u_{kl}u_{kl}'D^{-\frac{1}{2}}\Big),$$

*where $u_{kl} = Q_{\cdot,k} \odot Q_{\cdot,l} \in \mathbb{R}^n$ and $\varphi_{kl} = \xi_{kk} + \xi_{ll} - 2\xi_{kl}$.*

*Proof.* For a vector $y$ and a scalar $\nu$, we have that

$$Jy = \nu y \Leftrightarrow \tilde{J}w = \nu w,$$

where $y = D^{-\frac{1}{2}}w$. This follows from

$$\nu y = D^{-\frac{1}{2}}(\nu w) = D^{-\frac{1}{2}}(\tilde{J}w) = D^{-\frac{1}{2}}(I - D^{-\frac{1}{2}}HD^{-\frac{1}{2}})w = (D^{-\frac{1}{2}} - D^{-1}HD^{-\frac{1}{2}})w = JD^{-\frac{1}{2}}w = \nu y.$$

Next, we derive the expression for $\tilde{J}$. Note that

$$\sum_{k=1}^{n} q_{ik}^2\xi_{kk} = \sum_{k=1}^{n} q_{ik}^2 e^{\lambda_k} = Q_{i,\cdot}e^{\Lambda}Q_{i,\cdot}' = [e^{Q\Lambda Q'}]_{ii} = [e^{G}]_{ii} = \delta_i.$$

Thus for a diagonal element, $\tilde{J}_{ii}$, we have (using the expression for $h_{ii}$) that

$$\begin{aligned}\tilde{J}_{ii} = 1 - \frac{h_{ii}}{\delta_i} &= \frac{1}{\delta_i}\Big(\sum_{k=1}^{n} q_{ik}^2\xi_{kk} - \sum_{k=1}^{n}\sum_{l=1}^{n} q_{ik}^2 q_{il}^2\xi_{kl}\Big) = \frac{1}{\delta_i}\Big(\sum_{k=1}^{n} q_{ik}^2\xi_{kk} - \sum_{k=1}^{n} q_{ik}^2 q_{ik}^2\xi_{kk} - 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2\xi_{kl}\Big) \\ &= \frac{1}{\delta_i}\Big(\sum_{k=1}^{n} q_{ik}^2\xi_{kk}(1-q_{ik}^2) - 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2\xi_{kl}\Big) = \frac{1}{\delta_i}\Big(\sum_{k=1}^{n} q_{ik}^2\xi_{kk}\sum_{\substack{l=1\\l\neq k}}^{n} q_{il}^2 - 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2\xi_{kl}\Big) \\ &= \frac{1}{\delta_i}\Big(\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2(\xi_{kk}+\xi_{ll}) - 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2\xi_{kl}\Big) = \frac{1}{\delta_i}\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}^2 q_{il}^2\varphi_{kl}.\end{aligned}$$

Similarly for the off-diagonal elements we have

$$\begin{aligned}\tilde{J}_{ij} = -\frac{h_{ij}}{\sqrt{\delta_i\delta_j}} &= -\frac{1}{\sqrt{\delta_i\delta_j}}\sum_{k=1}^{n}\sum_{l=1}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl} = -\frac{1}{\sqrt{\delta_i\delta_j}}\Big(\sum_{k=1}^{n} q_{ik}^2 q_{jk}^2\xi_{kk} + 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl}\Big) \\ &= -\frac{1}{\sqrt{\delta_i\delta_j}}\Big(\sum_{k=1}^{n} q_{ik}q_{jk}\Big(-\sum_{\substack{l=1\\l\neq k}}^{n} q_{il}q_{jl}\Big)\xi_{kk} + 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl}\Big) \\ &= -\frac{1}{\sqrt{\delta_i\delta_j}}\Big(-\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}q_{jk}q_{il}q_{jl}(\xi_{kk}+\xi_{ll}) + 2\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}q_{jk}q_{il}q_{jl}\xi_{kl}\Big) \\ &= \frac{1}{\sqrt{\delta_i\delta_j}}\sum_{k=1}^{n-1}\sum_{l=k}^{n} q_{ik}q_{jk}q_{il}q_{jl}\varphi_{kl}.\end{aligned}$$

In the derivations above we used that $\sum_{k=1}^{n} q_{ik}q_{jk} = 1_{\{i=j\}}$, since $Q'Q = QQ' = I$. □

**Proof of Lemma 1.** Because $G(x)$ is symmetric and positive definite, then so is the principal submatrix, $H(x)$. Consequently, $M = D^{-\frac{1}{2}}H(x)D^{-\frac{1}{2}}$ is symmetric and positive definite. Thus any eigenvalue, $\mu$ of $M$ is strictly positive. So if $\nu$ is an eigenvalues of $\tilde{J}(x) = I - D^{-\frac{1}{2}}HD^{-\frac{1}{2}}$, then $\nu = 1-\mu$ where $\mu$ is an eigenvalue of $M$, from which it follows that all eigenvalues of $\tilde{J}$ are strictly less than 1.

Consider a quadratic form of $\tilde{J}$ with an arbitrary vector $z \in \mathbb{R}^n$. Using Lemma (A.3), it follows that any quadratic form is bounded from below by

$$z'\tilde{J}z = \sum_{k=1}^{n-1}\sum_{l=k}^{n}\varphi_{kl}\Big(z'D^{-\frac{1}{2}}u_{kl}u_{kl}'D^{-\frac{1}{2}}z\Big) = \sum_{k=1}^{n-1}\sum_{l=k}^{n}\varphi_{kl}\Big(z'D^{-\frac{1}{2}}u_{kl}\Big)^2 \geq 0,$$

because $\varphi_{kl} > 0$ by Lemma A.2. Hence, $\tilde{J}$ is positive semi-definite and $\nu_i \geq 0$, for all $i = 1,\ldots,n$.

Finally, since $J(x)$ and $\tilde{J}(x)$ have the same eigenvalues, it follows that all eigenvalues of $J(x)$ lie within the interval $[0,1)$, which proves that $g(x)$ is a contraction. □

**Proof of Theorem 1.** The Theorem is equivalent to the statement that for any symmetric matrix $G$, there always exists a unique solution to $g(x) = x$. This follows from Lemma 1 and Banach's fixed

point theorem. □

**Proof of Proposition 1.** We have $Y = PX$, for some permutation matrix, $P$, so that $C_y = PC_xP'$. Since a permutation matrix is such that $P'P = PP' = I$, it follows that

$$(C_y - I)^k = (PC_xP' - I)^k = (P[C_x - I]P')^k = P'(C_x - I)^k P'.$$

Consequently,

$$\begin{aligned} G_y = \log C_y &= (C_y - I) - \tfrac{1}{2}(C_y - I)^2 + \tfrac{1}{3}(C_y - I)^3 + \cdots \\ &= P[(C_x - I) - \tfrac{1}{2}(C_x - I)^2 + \tfrac{1}{3}(C_x - I)^3 + \cdots]P' \\ &= P[\log C_x]P' = PG_xP'. \end{aligned}$$

Suppose that the $i$-th and $j$-th rows of $P$ are the $r$-th and $s$-th unit vectors, $e_r'$ and $e_s'$, respectively, then we have $[G_y]_{ij} = [G_x]_{rs}$, and by symmetry

$$[G_y]_{ij} = [G_y]_{ji} = [G_x]_{rs} = [G_x]_{sr},$$

which shows that $\gamma_y$ is simply a permutation of the elements in $\gamma_x$. □

**Proof of Proposition 2.** An equi-correlation matrix can be written as $C = (1 - \rho)I_n + \rho U_n$, where $I_n \in \mathbb{R}^{n \times n}$ is identity matrix and $U_n \in \mathbb{R}^{n \times n}$ is a matrix of ones. Using the Sherman–Morrison formula, we can obtain the inverse,

$$C^{-1} = \frac{1}{1 - \rho}\left(I_n - \frac{\rho}{1 + (n-1)\rho}U_n\right). \tag{A.6}$$

Using (A.6), we can alternatively define $G$ as

$$G = -\log(C^{-1}) = -\log\left(\frac{1}{1 - \rho}I_n\right) - \log\left(I - \frac{\rho}{1 + (n-1)\rho}U_n\right). \tag{A.7}$$

Since the first term in (A.7) is a diagonal matrix, the off-diagonal elements of $G$ are determined only by the second term. The second term in (A.7) can be rewritten as follows,

$$\log(I_n - \varphi U_n) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{(-\varphi U_n)^k}{k} = \left(\frac{1}{n}\sum_{k=1}^{\infty}(-1)^{k+1}\frac{(-n\varphi)^k}{k}\right)U_n = \frac{1}{n}\log(1 - n\varphi)U_n, \tag{A.8}$$

where we denote $\varphi = \rho/(1 + (n-1)\rho)$ and use the fact that $U_n^k = n^{k-1}U_n$.

From (A.7) and (A.8), it follows that all the off-diagonal elements of $G$ are identical and take the value

$$G_{ij} = -\frac{1}{n}\log\left(1 - \frac{n\rho}{1+(n-1)\rho}\right)$$

for all $i$ and $j$, such that $i \neq j$. A closed-form result for $\gamma(C)$ follows immediately since $\gamma(C) = \text{vecl}(G)$. □

**Proof of Proposition 3.** From Theorem 1 it follows that the diagonal, $x = \text{diag}G$, is fully characterized by the off-diagonal elements, $y = \text{vecl}G = \text{vecl}G'$, and we may write $x = x(y)$. For the off-diagonal elements of the correlation matrix, $z = \text{vecl}C = \text{vecl}C'$, we have $z = z(x, y) = z(x(y), y)$, since $C = e^G$, and it follows that

$$\frac{dz(x,y)}{dy} = \frac{\partial z(x,y)}{\partial x}\frac{dx(y)}{dy} + \frac{\partial z(x,y)}{\partial y}. \tag{A.9}$$

With $A(x, y) = d\text{vec}C/d\text{vec}G$ and the definitions of $E_l$ and $E_u$, the second term is given by:

$$\frac{\partial z(x,y)}{\partial y} = E_l A(x,y) E_l' + E_l A(x,y) E_u'. \tag{A.10}$$

The expression has two terms because a change in an element of $y$ affects two symmetric entries in the matrix $G$. Similarly, for the first part of the first term in (A.9) we have,

$$\frac{\partial z(x,y)}{\partial x} = E_l A(x,y) E_d', \tag{A.11}$$

and what remains is to determine $\frac{dx(y)}{dy}$. For this purpose we introduce $D(x, y) = \text{diag}[e^{G(x,y)}] - \iota$ which implicitly defines the relation between $x$ and $y$. The requirement that $e^G$ is a correlation matrix, is equivalent to $D(x, y) = 0$. Next, let $\frac{\partial D}{\partial x}$ and $\frac{\partial D}{\partial y}$ denote the Jacobian matrices of $D(x, y)$ with respect to $x$ and $y$, respectively. These Jacobian matrices have dimensions $n \times n$ and $n \times n(n-1)/2$, respectively, and can also be expressed in terms of matrix $A(x, y)$, as follows

$$\frac{\partial D}{\partial x} = E_d A(x,y) E_d', \qquad \frac{\partial D}{\partial y} = E_d A(x,y) E_l' + E_d A(x,y) E_u'.$$

Note that $\frac{\partial D}{\partial x}$ is a principal sub-matrix of positive definite matrix $A$ and, hence, is an invertible matrix. Therefore, from the Implicit Function Theorem it follows

$$\frac{dx(y)}{dy} = -\Big(\frac{\partial D}{\partial x}\Big)^{-1}\frac{\partial D}{\partial y} = -\Big(E_d A(x,y) E_d'\Big)^{-1}\Big(E_d A(x,y) E_l' + E_d A(x,y) E_u'\Big). \tag{A.12}$$

The results now follows by inserting (A.10), (A.11) and (A.12) into (A.9). □

**Proof of Lemma 2.** We have $G = Q \log \Lambda Q'$, where $C = Q\Lambda Q'$ is the spectral decomposition of the correlation matrix. Thus a generic element of $G$ can be written as

$$G_{ij} = \sum_{k=1}^{n} q_{ik} q_{jk} \log \lambda_k.$$

From the Jensen's inequality it follows that

$$G_{ii} = \sum_{k=1}^{n} q_{ik}^2 \log \lambda_k \leq \log \left( \sum_{k=1}^{n} q_{ik}^2 \lambda_k \right),$$

where we used that $\sum_{k=1}^{n} q_{ik} q_{jk} = 1_{\{i=j\}}$, because $Q'Q = I$. Finally, since $\sum_{k=1}^{n} q_{ik}^2 \lambda_k = C_{ii} = 1$, it follows that $G_{ii} \leq \log 1 = 0$. □