Separation-based Parameterization Strategies for Estimation of Restricted Covariance Matrices in Multivariate Model Systems

Shobhit Saxena
Department of Civil Engineering
Indian Institute of Science (IISc), Bengaluru 560012, India
Email: shobhits@iisc.ac.in

Chandra R. Bhat (Corresponding author)
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Tel: 1-512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abdul Rawoof Pinjari
Department of Civil Engineering
Centre for Infrastructure, Sustainable Transportation, and Urban Planning (C*i*STUP)
Indian Institute of Science (IISc), Bengaluru 560012, India
Tel: +91-80-2293-2043
Email: abdul@iisc.ac.in

ABSTRACT
Many multivariate model systems involve the estimation of a covariance matrix that must be positive-definite. A common strategy to ensure positive definiteness of the covariance matrix is through the use of a Cholesky parameterization of the covariance matrix. However, several model systems require imposing restrictions on the elements of the covariance elements. For instance, modelling systems may require fixing some (or all) of the diagonal elements in the covariance matrix to unity due to identification considerations. However, imposing such restrictions using the traditional Cholesky decomposition approach is not feasible and requires the additional parameterization of the Cholesky elements.

In this paper, we explore a separation-based strategy with spherical parameterization of the Cholesky matrix to impose restrictions on the covariance matrix. Importantly, using this separation-based parameterization strategy, we also explore the possibility of restricting some covariance (or correlation) terms to zero. The effectiveness of the proposed strategy is assessed through extensive simulation experiments. The results from the simulation experiments highlight better performance of the separation-based strategy in terms of recovery of model parameters – particularly those in the covariance matrix, than the traditional Cholesky parameterization approach. Finally, the proposed strategy is implemented in a joint multivariate binary probit ordered probit model system to analyze the usage (and the extent of use) of non-private modes of transportation in Bengaluru, India. In doing so, the proposed strategy is implemented to restrict several correlations to zero, thus avoiding the estimation of a profligate correlation matrix and substantially easing the estimation process.

Keywords: Covariance matrix, Cholesky decomposition, separation-based strategy, spherical parametrization, restrictions on the correlation matrix.

1 INTRODUCTION

The estimation of statistical and econometric models that include multiple outcome variables (that is, a multivariate dependent variable model) has increased over the years, thanks to the ability to generate multivariate distributions through the use of relatively flexible copula-based methods and/or the use of effective factorization techniques for the parsimonious estimation of covariance matrices (see, for example, Rana et al., 2010, Bhat, 2015, Müller and Czado, 2018, Jiryaie and Khodadadi, 2019, and Ong et al., 2018). Besides, the alternative of simply ignoring the dependence and estimating separate models, while computationally appealing, is inefficient in estimating covariate effects for each outcome because it fails to borrow information on other outcomes, and is limiting in its ability to answer intrinsically multivariate questions such as the effect of a covariate on a multidimensional outcome (Teixeira-Pinto and Harezlak, 2013). Perhaps, more importantly, when the intention is to consider the potential effect of one outcome (say outcome A) on another outcome (say outcome B), the simple introduction of outcome A as an independent exogenous variable in the modeling of outcome B immediately triggers possible endogeneity bias effects because of the inter-relationship between the two outcomes due to common unobserved effects. On the other hand, the explicit consideration of the possibility that the outcomes may be co-determined controls for this possible endogeneity bias.

An important consideration in multivariate models (or even in univariate models with multiple random coefficients on exogenous variables with correlations across the coefficients) is the estimation of a covariance matrix (sometimes also referred to as a variance-covariance matrix, though we will use the shorter label "covariance matrix" that includes both the diagonal and non-diagonal elements). This covariance matrix, along with other model parameters, is estimated using Bayesian or frequentist methods and will generally involve optimization methods that extensively search over a large parameter space.[1] In these estimation methods, a challenge is to maintain positive definiteness of the covariance matrix all through the search process (equivalently, to ensure that the covariance matrix does not become nonpositive definite).[2] This can be achieved in one of two ways. The first method is to impose restrictions on the covariance elements to ensure positive definiteness, which will generally lead to constrained optimization methods. Unfortunately, such constrained optimization methods collapse to multiple unconstrained problem estimations with some trial-and-error or to solving a complex non-linear equation system that is itself difficult to formulate (Dennis and Schanbel, 1989; Pinheiro and Bates, 1996). In particular, each unconstrained estimation typically tends to be undertaken using a simple one-level Cholesky decomposition schemes that write the Cholesky elements in a form conforming to the unit diagonal

[1] When the covariance matrix to be estimated pertains to observed quantities, one can ensure positive semi-definiteness using a sample covariance matrix estimator. However, such an estimator is not possible in most model situations, because the covariance matrix pertains to conditional unobserved error quantities.

[2] A covariance matrix is said to be semi-positive definite if none of its diagonal elements are negative (so, the diagonal elements can be zero or positive), while it is said to be positive definite if none of its diagonal elements are zero or negative (so, the diagonal elements are all strictly positive). In an estimation context, a semi-positive, but not positive definite, covariance matrix would imply that the variance of one or more components is zero, which is a degenerate case and implies lack of stochasticity for that component. Thus, we will restrict our attention to positive definite covariance matrices here.

vector in the correlation matrix. The problem with such an approach (see Srinivasan and Bhat, 2005) is that the estimation can break down, unless the code imposes a steep penalty if any of the diagonal Cholesky elements turn out to be complex (imaginary) during the search process. While a reasonable strategy, such estimations also typically entail the construction of a "nearest" valid correlation matrix when positive definiteness fails (for example, by replacing the negative eigenvalue components in the correlation matrix with a small positive value, or by adding a sufficiently high positive value to the diagonals of a matrix and normalizing to obtain a correlation matrix; see Rebonato and Jäckel, 2000, Higham, 2009, and Schöttle and Werner, 2004 for detailed discussions of these and other adjusting schemes; a review of these techniques is beyond the scope of this paper). And even then, there is no guarantee that the correlation matrix at "convergence" will be positive definite. Thus, it is almost always the case that a second method that involves a reparameterization of the covariance matrix in a way that renders the resulting estimation process completely unconstrained (while also enforcing the positive definiteness condition) is the preferred method.

Many different reparameterization approaches for the unconstrained estimation of covariance matrices have been proposed in the literature, including a spectral decomposition method, a matrix logarithm method, the typically used Cholesky decomposition approach for the covariance matrix (which only guarantees semi-positive definiteness rather than positive definiteness), modified Cholesky decomposition methods for the covariance matrix (such as the so-called LDLT decomposition and the log-Cholesky decomposition), the Cholesky decomposition of the inverse of the covariance matrix, factor-analytic approaches for the covariance matrix, and a spherical parameterization approach for the covariance matrix that combines the Cholesky decomposition with a specific spherical parameterization of the Cholesky matrix (see Lindstrom and Bates, 1988, Pinheiro and Bates, 1996, Pourahmadi, 2000, Leonard and Hsu, 1992, and McNeish and Bauer, 2022 for details and reviews of these methods). However, these decompositions, except for the spherical parameterization, are not immediately suitable for estimation when correlation matrices are the focus rather than covariance matrices (this is because these decompositions do not adhere to the additional restriction of unit diagonals of a correlation matrix). For instance, correlation matrices are the focus in a multivariate binary choice model system or a multivariate ordered response system (see Bhat et al., 2010, Dias et al., 2020, and Bhat and Mondal, 2021), where the scale of the latent variables underlying the limited dependent outcomes have to be normalized.[3]

Even in cases where a covariance matrix is to be estimated, there may be substantial value in breaking down the covariance matrix into a scale (standard deviation) matrix and a correlation

[3] As indicated by Bhat and Mondal (2021), the scale normalization itself may be achieved in one of two ways. The first is to normalize the error term scale (leading to the case of correlation matrices being estimated rather than covariance matrices), and the second is to set a coefficient affecting the latent variable to a fixed constant, while estimating a covariance matrix. The second approach may seem more attractive, but the first approach leads to far fewer cases of numerical estimation instability than the latter, because it fixes the overall scale of the conditional unobserved error terms to be congruent across the dependent outcomes (see also Kohli et al., 2019 and McNeish and Bauer, 2022 for a related discussion).

matrix and estimating these separately. Barnard et al. (2000) refer to such an approach as a "separation strategy", while we also see this as a "divide and conquer" strategy. Doing so has both specifications as well as estimation advantages. *On the specification side*, it is more natural for analysts to think about standard deviations (which are scale-related) and correlations (which are scale-free) in expressing (and imposing) a priori judgements about the relationship among a set of variables, whether in a Frequentist setting or a Bayesian setting. For example, consider a set of random coefficients in an inter-city travel mode choice model on such level-of-service (LOS) explanatory variables as in-vehicle travel time, out-of-vehicle travel time, travel cost, and service frequency. If there is a belief that income and gender play a role in trade-offs, one can specify different mean coefficients on the LOS variables for each of the four segments. But rather than specify the same covariance matrix for these random parameters across the four segments (which could be very restrictive), or allow a separate covariance matrix for each of the segments (which would lead to a profligate model in parameters, potentially not estimable with the sample size for estimation), one could settle for an intermediate specification that specifies the scale (the standard deviation matrix, or spread of values) of the random coefficients to be different across the segments, but maintains the same dependence structure (correlation matrix, or multivariate dependency shape) for the random coefficients across the segments. Similarly, in the case of a multivariate mixed dependent outcome model, the analyst may reasonably assume (in terms of the best balance between parsimony and behavioral realism) that, in a mixed model system of residential choice (say in 3-4 broad categories of downtown, urban, suburban, and rural living), bicycle ownership, car ownership, commute mode choice, and the number of leisure trips per time period, the correlations across the underlying latent variables for the latter four dimensions are the same across residential locations, but the scales are different. Similar specification structures of identical correlation patterns, but different scales, are commonly used for model coefficients or for the relationship across multiple outcomes when a single model is estimated from different sources of data (such as revealed and stated preference data). At the same time, a logical consideration in such model systems can be to restrict correlations across certain specific dimensions. For instance, in a multiple discrete-continuous model system, a behaviorally consistent consideration would be to restrict correlations between the discrete preference of a good i and the continuous preference of a good j ($j \neq i$) to zero.

On the estimation side, separating out the scale from correlations is helpful in multivariate mixed outcome models (these models have a mix of different types of outcomes, such as continuous, ordinal, grouped, count, and unordered-response outcomes). In such models, the scale of the latent variables underlying the non-continuous and non-grouped outcomes will be fixed to one for identification purposes, while those for the continuous and grouped outcomes will be left free for estimation. Directly estimating a covariance matrix to adhere to these constraints (of some diagonal elements constrained to one and others left free), while also using decomposition techniques to preserve positive-definiteness, leads to a complex and unwieldy situation. In such situations, it is much easier (if not the only way) to use the partitioning strategy into a separate scale matrix and a separate correlation matrix with uniform entries of one on the matrix diagonal.

This issue becomes particularly obvious when coding software for the estimation of such models. Besides, de-scaling before estimating interdependencies generally leads to much faster and more stable convergence (see Kohli et al., 2019), and, in a Bayesian inference context, leads to a simple computational strategy for obtaining the posterior distributions of the scale and the correlation matrices (Barnard et al., 2000).

The importance of the separation strategy, while invoked as a potentially effective strategy for Bayesian estimation in Barnard et al. (2000) in many model applications, has particularly started seeing much application in the estimation context of mixed multivariate outcome modeling (see, for example, Jiryaie and Khodadadi, 2019), skew-normal copula models (which are one form of mixed outcome modeling, because the skew can be generated through a latent variable censoring mechanism with the scale of the latent variable set to one; see, for example, Sidharthan and Bhat, 2012), and multinomial probit modeling (see, for example, Bhat and Lavieri, 2018). The precise unconstrained parameterization adopted to ensure that the correlation matrix is positive definite in such applications (as well as in other applications where a correlation matrix is of interest, such as in multivariate ordinal response systems or latent construct-based factorization models; see Bhat, 2015) has varied from simple one-level Cholesky decomposition schemes that write the diagonal Cholesky elements in a form conforming to unit diagonal in the correlation matrix (see, for example, Srinivasan and Bhat, 2005 and Bhat and Lavieri, 2018) to specific multi-level Cholesky parameterization schemes. The problem with the first one-level decomposition scheme is that the estimation can break down unless the code imposes a steep penalty for the diagonal elements of the Cholesky if any of these elements turn out to be zero or negative during the search process. While a reasonable strategy, such estimations can require a good bit of handholding during estimation. In the second set of multi-level decomposition schemes, three methods are available, all of which have a common second-level parameterization for the Cholesky elements but differ in the third-level of parameterization (see Bhat and Mondal, 2021 for a detailed discussion): (1) the partial correlations method (Joe, 2006), (2) the spherical parameterization method (Pinheiro and Bates, 1996 and Rebonato and Jäckel, 2000), and (3) the radial parametrization method (van Oest, 2021). Of these, the first partial correlations method tends to be relatively complicated in application, and the latter two parameterizations have been shown to be essentially equivalent but for a scaling difference (Bhat and Mondal, 2021).

In this paper, we first examine the value of the separation or "divide and conquer" strategy as it relates to convergence during estimation and the recovery of "true" parameters, and compare this approach with the more traditional Cholesky decomposition approach for covariance matrices. In addition, we also exploit the separation-based strategy to explore the possibility of imposing restrictions on the covariance matrix. Specifically, we explore the possibility of restricting some correlation elements to zero.

Surprisingly, there has been very little exploration of these issues, with most applications using the traditional Cholesky decomposition approach. Of course, as already discussed, the traditional Cholesky decomposition approach is not appropriate for mixed models with some diagonal entries of the covariance matrix normalized to one for the identification or restricting some correlation values to zero. Thus, in the simulation experiments used in the comparison of the two alternative covariance decomposition methods to maintain semi-positive definiteness, we consider a mixed model with unrestricted diagonal entries, except for a top diagonal restriction that can be easily accommodated in both the decomposition approaches. However, we still restrict some correlation values to zero to highlight the efficacy of the proposed strategy. Specifically, we consider the case of a multivariate mixed model with one multinomial unordered-response outcome (with four alternatives), a grouped outcome, and a continuous outcome (leading to a five-dimensional covariance matrix in estimation). Issues of convergence, computation time, as well as accuracy and precision in recovering parameters are examined using both the separation method and the traditional Cholesky method. Further, in the separation method, Bhat and Mondal (2021) indicate that the scale embedded in the logistic function of the spherical (or equivalently, radial)[4] parameterization (for the correlation matrix) can impact convergence rates and computational times. Thus, we explore such scale impacts too within the context of the separation approach. Second, we consider a motivating example for the case of a multivariate dependent variable model where some diagonal entries of the covariance matrix are normalized to one for identification. In this case, unless external boundary constraints are placed through a restricted maximum likelihood approach (which, as discussed earlier, is to be avoided because it can lead to very substantial estimation difficulties), one has to employ the separation approach. In this context, we demonstrate the application of the separation strategy through an empirical analysis. In doing so, we also present and demonstrate the use of an approach to estimate model specifications with restricted correlation matrices (that is, some correlation elements are restricted to zero). This is an issue that has received surprisingly no attention (as far as we are aware) in the literature but is important because restrictions on the correlation matrix do not immediately translate to convenient restrictions that can be imposed on the Cholesky elements of the correlation matrix. We develop an algorithm to impose such restrictions, along with an algorithm for the gradient of the correlation elements with respect to active Cholesky elements. The algorithm is implemented in Gauss software and is made available for its use.

The rest of the paper is structured as follows: Section 2 discusses different parameterization approaches to facilitate the estimation of the covariance matrix. Next, building on the separation-based strategy with spherical parametrization, we discuss (and derive) conditions to restrict some correlation parameters to zero. In Section 3, we present simulation experiments to exemplify the benefits of the proposed strategy over the traditional Cholesky parametrization approach. In Section 4, we illustrate the use of the proposed separation-based strategy as well as the procedure developed for restricting specific correlations to zero for an empirical application to analyze the usage and extent of use of non-private modes in Bengaluru, India. Finally, Section 5 concludes the paper with a quick summary.

[4] In rest of the paper, we use the phrase "spherical parameterization" to represent both spherical and radial parameterization strategies, since both strategies are essentially equivalent (Bhat and Mondal, 2021).

2 ALTERNATIVE DECOMPOSITION PROCEDURES FOR A COVARIANCE MATRIX

2.1 The Traditional Cholesky Decomposition

The traditional econometric approach to ensure positive semi-definiteness of the covariance matrix, without requiring constraints on the parameters, is to reparametrize the covariance matrix $\mathbf{\Omega}$ in terms of its Cholesky decomposition. The basic idea is to write the covariance matrix in terms of its "square root" matrix so that $\mathbf{\Omega} = \mathbf{G}'\mathbf{G}$, where $\mathbf{G}$ is an upper triangular Cholesky matrix. This is a very effective approach, though it can run into convergence problems unless the diagonal elements are specified to be positive. Specifically, the estimated Cholesky parameters are not necessarily unique, since multiplying a subset of rows of $\mathbf{G}$ by –1 results in the same covariance matrix $\mathbf{\Omega}$ (see Pinheiro and Bates, 1996). However, this issue can be quite easily addressed by further parameterizing the diagonal Cholesky elements in logarithmic form, so that the actual diagonal elements are strictly positive. This variant of the Cholesky decomposition is the one used in the current paper. However, even in this case, there can be convergence problems if the variances of specific elements of the matrix $\mathbf{\Omega}$ are of quite different magnitudes. The Cholesky decomposition has been widely discussed in standard econometric texts (see, for example, Train, 2009) and so is not presented in detail in this paper.

2.2 The Separation-Based Spherical Parameterization

In the separation-based approach, the covariance matrix is first separated out into a diagonal standard deviation matrix $\boldsymbol{\omega}$ and a correlation matrix **R**, so that $\mathbf{\Omega} = \boldsymbol{\omega}\mathbf{R}\boldsymbol{\omega}$. Next a Cholesky decomposition is applied to the correlation matrix **R**, such that $\mathbf{R} = \mathbf{L}'\mathbf{L}$ (note that the Cholesky is with respect to the correlation matrix here, so the matrix **L** is different from the matrix $\mathbf{G}$ earlier, and $\mathbf{\Omega} = \boldsymbol{\omega}\mathbf{L}'\mathbf{L}\boldsymbol{\omega}$). Thus, consider an $M \times M$ correlation matrix **R** as follows, with $r_{i,j} = r_{j,i}$ and $-1 < r_{i,j} < 1\ \forall i \neq j.$:

$$R = \begin{pmatrix} 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,M-1} & r_{1,M} \\ r_{2,1} & 1 & r_{2,3} & \cdots & r_{2,M-1} & r_{2,M} \\ r_{3,1} & r_{3,2} & 1 & \cdots & r_{3,M-1} & r_{3,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{M-1,1} & r_{M-1,2} & r_{M-1,3} & \cdots & 1 & r_{M-1,M} \\ r_{M,1} & r_{M,2} & r_{M,2} & \cdots & r_{M,M-1} & 1 \end{pmatrix} \tag{1}$$

The Cholesky decomposition of the above matrix **L** such that $\mathbf{R} = \mathbf{L}'\mathbf{L}$ is given by:

$$
L=\begin{pmatrix}
1 & l_{1,2} & l_{1,3} & \cdots & l_{1,M-1} & l_{1,M} \\
0 & \sqrt{1-l_{1,2}^2} & l_{2,3} & \cdots & l_{2,M-1} & l_{2,M} \\
 & & \sqrt{1-l_{1,3}^2-l_{2,3}^2} & \cdots & l_{3,M-1} & l_{3,M} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \sqrt{1-\sum_{k=1}^{M-2} l_{k,M-1}^2} & l_{M-1,M} \\
0 & 0 & 0 & \cdots & 0 & \sqrt{1-\sum_{k=1}^{M-1} l_{k,M}^2}
\end{pmatrix} \tag{2}
$$

Estimating the elements above for **L** directly may work, but an issue with the above Cholesky decomposition for the correlation matrix **R** is that the diagonal will frequently get into imaginary space when the quantity within the square root goes to a value of zero or less. Thus, it is common place to further parameterize the Cholesky elements $l_{i,j}$ in additional levels to adhere to a positive value for the diagonal quantities throughout the iterative estimation process. Multiple parameterizations are possible, as discussed in Bhat and Mondal (2021), including (1) the partial correlations method (Joe, 2006), (2) the spherical parameterization method (Pinheiro and Bates, 1996 and Rebonato and Jäckel, 2000), and (3) the radial parametrization method (van Oest, 2021). Of these three methods, the first partial correlation method tends to be relatively complicated in application. And Bhat and Mondal (2021) show that the spherical and radial parameterization essentially are the same in the context of estimation, based on appropriate scaling of the embedded logistic function embedded in both these parameterizations. Further, the spherical parameterization is typically faster in large problems. So, here we will use the spherical parameterization, but will test for the scale that provides the best combination of speed and convergence. The spherical parameterization method has also seen the most used because of its ease and interpretability of the parameters (see Madar, 2015, Pourahmadi and Wang, 2015, and Tsay and Pourahmadi, 2017).

For the spherical parameterization, first consider the parameterization of each element $l_{i,j}=h_{i,j}\sqrt{\prod_{k=1}^{i-1}\left(1-h_{k,j}^2\right)}$, $-1<h_{i,j}<1\ (i\neq j)$, $h_{i,i}=1\,\forall i$, and $l_{1,j}=h_{1,j}\ \forall j$ (because the Cholesky refers to the upper diagonal matrix, the entries correspond to $l_{i,j}$ such that $i\leq j$, a point we will not belabor over in all following notations). The values for each row *i* for the $h_{i,j}$ values are built up successively from the corresponding column values for the previous *i*–1 columns. Thus, the parameterization looks like what follows below:

$$L=\begin{pmatrix} 1 & h_{1,2} & h_{1,3} & \cdots & h_{1,M-1} & h_{1,M} \\ & \sqrt{1-h_{1,2}^2} & h_{2,3}\sqrt{1-h_{1,3}^2} & \cdots & h_{2,M-1}\sqrt{1-h_{1,M-1}^2} & h_{2,M}\sqrt{1-h_{1,M}^2} \\ & & \sqrt{\left(1-h_{1,3}^2\right)\left(1-h_{2,3}^2\right)} & \cdots & h_{3,M-1}\sqrt{\left(1-h_{1,M-1}^2\right)\left(1-h_{2,M-1}^2\right)} & h_{3,M}\sqrt{\left(1-h_{1,M}^2\right)\left(1-h_{2,M}^2\right)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & \cdots & \sqrt{\prod_{k=1}^{M-2}\left(1-h_{k,M-1}^2\right)} & h_{M-1,M}\sqrt{\prod_{k=1}^{M-2}\left(1-h_{k,M}^2\right)} \\ & & & \cdots & & \sqrt{\prod_{k=1}^{M-1}\left(1-h_{k,M}^2\right)} \end{pmatrix} \quad (3)$$

It is easy to see that the parameterization above satisfies the condition in Equation (2) that the diagonal terms must be $l_{i,i}=\sqrt{1-\sum_{k=1}^{i-1}l_{k,i}^2}$. For example, $l_{3,3}=\sqrt{(1-h_{1,3}^2)(1-h_{2,3}^2)}=\sqrt{1-h_{1,3}^2-h_{2,3}^2(1-h_{1,3}^2)}=\sqrt{1-l_{1,3}^2-l_{2,3}^2}$, which is as required. The additional condition that all diagonal terms $l_{i,i}>0$ is also satisfied as long as the condition $-1<h_{i,j}<1\ (i\neq j)$ holds. For this, in a third-level spherical parameterization, we write $h_{i,j}=\cos\left[\pi F\left(\theta_{i,j}\right)\right],\ i\neq j$, where $F(\theta_{i,j})=\dfrac{1}{1+\exp\left(-\dfrac{\theta_{i,j}}{\lambda}\right)}$ is a logistic function with λ as the scale parameter. Note that the suitability of different levels of the scale parameter (λ = 0.8, 1.0, and 1.2) is explored in this study.

2.3 Restricted Covariance Structures

The elements of the correlation matrix **R** corresponding to the covariance matrix $\mathbf{\Omega}$ may be written as a function of the Cholesky matrix elements of **R**:

$$R_{ij}=\left(\sum_{s=1}^{i-1}L_{si}L_{sj}\right)+L_{ij}\sqrt{1-\sum_{s=1}^{i-1}L_{si}^2}\ . \quad (4)$$

With the convention that $\prod_{k=1}^{M-1}\left(1-h_{k,j}^2\right)=1$, and after some tedious but straightforward algebraic manipulations exploiting the structure of the $l_{ij}\rightarrow h_{ij}$ parameterization, the elements above can be written as a function of the elements h_{ij} as:

$$R_{ij}=\left(\sum_{s=1}^{i-1}\left[h_{si}h_{sj}\sqrt{\prod_{k=1}^{s-1}\left(1-h_{ki}^2\right)\left(1-h_{kj}^2\right)}\right]\right)+h_{ij}\sqrt{\prod_{k=1}^{i-1}\left(1-h_{ki}^2\right)\left(1-h_{kj}^2\right)} \quad (5)$$

Thus, if any element R_{ij} of the correlation matrix is zero, it must be true that:

$$h_{ij} = -\frac{\left(\sum_{s=1}^{i-1}\left[h_{si}h_{sj}\sqrt{\prod_{k=1}^{s-1}\left(1-h_{ki}^2\right)\left(1-h_{kj}^2\right)}\right]\right)}{\sqrt{\prod_{k=1}^{i-1}\left(1-h_{ki}^2\right)\left(1-h_{kj}^2\right)}} \tag{6}$$

The important point to note is that any restriction in any column j of the i^{th} row only requires that the corresponding h_{ij} element be developed as a function of elements in the $(i–1)^{\text{th}}$ row and earlier. Thus, based on the specification provided for the restrictions to be imposed, it is possible to construct the appropriate restricted h_{ij} values row by row until the last row.

In our estimations, for efficiency in implementation, we develop values for a given row all at once, if there is a restriction in a correlation element in that row (there is no need to compute restricted values for rows of the correlation matrix without any constraints). That is, if a particular element of the row is specified to have zero correlation, the corresponding element h_{ij} should be as given in Equation (6). Once this is computed, the implied $\theta_{i,j}$ value is calculated using the following inverse transformation: $\theta_{i,j} = \ln\left(\frac{\arccos(h_{i,j})}{\pi - \arccos(h_{i,j})}\right),\ i \neq j,$ and used for the computation of the spherically parameterized Cholesky elements in subsequent rows.

The coding of the restricted estimation procedure is tricky, which is probably why we could not find any reference in the literature to econometric model estimations with a restricted correlation matrix. The basic issue is to develop an appropriate spherically-parameterized Cholesky decomposition of the correlation matrix **R** that adheres to the zero-correlation constraints specified in the matrix **R**. Additionally, for use in estimation, the gradient of all elements of the correlation matrix needs to be written as a function of the active correlation parameters, after imposing restrictions on the correlation elements based on Equation (6). An important note here is that some zero-restricted values may not be consistent with a positive definite correlation matrix, because the restrictions may be such that the implied value for one or more h_{ij} values do not adhere to the –1 to +1 range (this issue, of course, does not arise in an unrestricted estimation, because each correlation element is appropriately parameterized to ensure that each and every h_{ij} element is bounded by –1 and +1). In such instances, the code adjusts to produce a positive-definite correlation matrix, though it will then not strictly adhere to the constraints of zero correlations for specific elements. The code also identifies which specific element had to be released to ensure a positive-definite matrix. This information can be used by the analyst to inform the restricted correlation specification.

3 SIMULATION STUDY

We undertook simulations to investigate the performance of the proposed algorithm that allows restricting specific correlation values to zero. In the simulation experiments, we consider a multivariate mixed model with one multinomial unordered-response outcome (with four alternatives), a grouped outcome, and a continuous outcome (leading to a five-dimensional covariance matrix in estimation). Further, to showcase the effectiveness of the proposed strategy, we restricted two correlation parameters to zero (more on this in the subsequent discussion). The model structure for the above setup is discussed next.

3.1 Model Structure

Consider the nominal (unordered-response) outcome for an individual, and let i be the corresponding index for alternatives ($i = 1, 2, 3, ..., I$; $I = 4$ in the current simulation setting). Let the individual under consideration choose the alternative m. Also, assume the usual random utility structure for each alternative i.

$$U_i = \boldsymbol{\beta}_i' \boldsymbol{x} + \varepsilon_i, \tag{7}$$

where $\boldsymbol{x}$ is an (H×1) fixed column vector of exogenous variables, $\boldsymbol{\beta}_i$ is another (H×1) column vector of corresponding coefficients, and ε_i is a normal random error term. To move forward, let $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_I)'$ ($I \times 1$ vector), and $\boldsymbol{\varepsilon} \sim MVN_I(\mathbf{0}_I, \boldsymbol{\Lambda})$, where $MVN_I(\mathbf{0}_I, \boldsymbol{\Lambda})$ stands for a multivariate normal distribution of dimension I with a mean $I \times 1$ vector of zeros and a covariance matrix $\boldsymbol{\Lambda}$. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}$ of the error differences, $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\varepsilon}_2, \tilde{\varepsilon}_3 ..., \tilde{\varepsilon}_I)'$ (where $\tilde{\varepsilon}_i = \varepsilon_i - \varepsilon_1$, $i \neq 1$), is estimable. In addition, the usual identification restriction is imposed such that one of the alternatives serves as the base when introducing alternative-specific constants and variables that do not vary across alternatives (that is, whenever an element of $\boldsymbol{x}$ is individual-specific and not alternative-specific, the corresponding element of $\boldsymbol{\beta}_i$ is set to zero for at least one alternative i). Next, define $\mathbf{U} = (U_1, U_2, ..., U_I)'$ ($I \times 1$ vector), and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_I)'$ ($I \times H$ matrix). Then, in matrix form, we may write:

$$\mathbf{U} = \boldsymbol{\beta}\mathbf{x} + \boldsymbol{\varepsilon}. \tag{8}$$

Next, for later use, note that, under the utility maximization paradigm, $u_{im} = U_i - U_m$ must be less than zero for all $i \neq m$, since the individual chose alternative m. Stack the latent utility differentials into a vector $\boldsymbol{u} = \left[\left(u_{1m}, u_{2m}, ..., u_{Im} \right)'; i \neq m \right]$ and the corresponding error differentials (taken with respect to the alternative m) as , $\vec{\boldsymbol{\varepsilon}} = (\vec{\varepsilon}_2, \vec{\varepsilon}_3, ..., \vec{\varepsilon}_I)'$ (where $\vec{\varepsilon}_i = \varepsilon_i - \varepsilon_m$, $i \neq m$).

Further, consider the structure for a grouped model with say J groups. Let y_n^* be the underlying latent variable whose horizontal partitioning leads to the observed outcome for the grouped variable. An example of a grouped variable may be annual household income, which is typically collected in grouped windows such as <50,000, 50,000-100,000, etc. or annual vehicle miles of travel, which may be in groups such as <5,000, 5,000-7,500, 7,500-10,000, etc.; the difference between an ordinal variable and a grouped variable is that the thresholds demarcating different groups are known in advance in a grouped model, while these thresholds have to be estimated in the ordinal model. But because the thresholds are observed in the grouped case, the scale of the underlying variable is estimable. So, assume that the individual under consideration chooses the a^{th} grouped category. Then, in the usual latent variable representation for the grouped variable, we may write the following for this individual:

$$y^* = \boldsymbol{\gamma}'\mathbf{z} + \xi, \text{ and } \psi_{a-1} < y^* < \psi_a, \tag{9}$$

where $\mathbf{z}$ is an $A\times 1$ vector of exogenous variables (including a constant), $\boldsymbol{\gamma}$ is a corresponding vector of coefficients to be estimated, the ψ terms represent the observed thresholds corresponding to the grouping in which the individual is observed, and ξ is a normal random error $\left(\xi \sim N(0,\sigma_4)\right)$.

For the continuous variable, let $y = \boldsymbol{\delta}'\mathbf{w} + \zeta$ in the usual linear regression fashion, where $\mathbf{w}$ is an $(C\times 1)$ vector of exogenous variables (including a constant). $\boldsymbol{\delta}$ is a corresponding compatible coefficient vector, and $\left(\zeta \sim N(0,\sigma_5)\right)$.

To consider jointness, we assume the following upper diagonal error structure for the five-error vector $\boldsymbol{\eta} = (\tilde{\boldsymbol{\varepsilon}}, \boldsymbol{\xi}, \zeta)'$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ & & & \sigma_{44} & \sigma_{45} \\ & & & & \sigma_{55} \end{bmatrix} \tag{10}$$

The first three rows and three columns of the matrix above correspond to the utility differences (taken with respect to the first alternative's error term) for the unordered-response variable, the third row/column corresponds to the binary choice outcome, the fourth row/column to the ordered-response outcome, and the last column/row to the continuous outcome. In the scale matrix, the scales associated with $\tilde{\varepsilon}_2 = \varepsilon_2 - \varepsilon_1$ and $\tilde{\varepsilon}_3 = \varepsilon_3 - \varepsilon_1$ are identified (as discussed earlier), as are the scales for the grouped and continuous outcome errors. The correlation matrix is, of course, symmetric and must be positive-definite.

3.2 Model Estimation

For model estimation, the first step is to obtain the implied covariance matrix corresponding to $\boldsymbol{\mu} = (\boldsymbol{\varepsilon}, \boldsymbol{\xi}, \zeta)'$ from that of $\boldsymbol{\eta} = (\tilde{\boldsymbol{\varepsilon}}, \boldsymbol{\xi}, \zeta)'$. To do so, define a (6×5) matrix $\boldsymbol{D}$ as follows:

$$D=\begin{bmatrix}0&0&0&0&0\\1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix} \tag{11}$$

Then, the covariance matrix of $\boldsymbol{\mu}$ may be developed as $\boldsymbol{\Omega}=\mathbf{D\Sigma D}'$. All parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on utility differences. At the same time, it provides a consistent means to obtain the covariance matrix $\boldsymbol{\Xi}$ of $\boldsymbol{\tau}=(\ddot{\boldsymbol{\varepsilon}},\boldsymbol{\xi},\zeta)'$, where $\ddot{\boldsymbol{\varepsilon}}$ is the vector of error differences with respect to the chosen alternative. To write this utility differential-based vector compactly in terms of the original utilities, define a matrix **M** of size 5×6 with all zero entries. Insert an identity matrix of size 3 after supplementing it with a column of '–1' values in the column corresponding to the chosen alternative (the identity matrix would then become a 3×4 matrix), and insert an identity matrix of size 2 into the last two rows and two columns of the matrix **M**. For our simulation setting, assuming the individual under consideration selects the second alternative, the matrix **M** takes the following form:

$$\mathbf{M}=\begin{bmatrix}1&-1&0&0&0&0\\0&-1&1&0&0&0\\0&-1&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix} \tag{12}$$

With the matrix **M** as defined, the covariance matrix $\boldsymbol{\Xi}$ is given by $\boldsymbol{\Xi}=\mathbf{M\Omega M}'=\mathbf{MD\Sigma D'M}'$. To proceed, define $\tilde{\mathbf{B}}=[(\boldsymbol{\beta}\mathbf{x})',(\boldsymbol{\gamma}'\mathbf{z})',(\boldsymbol{\delta}'\mathbf{w})')]'$ $[6\times1]$ vector, $\mathbf{B}=\mathbf{M}\tilde{\mathbf{B}}$ $[5\times1]$ vector, and $\mathbf{g}=\left[\mathbf{u}',y^*,y\right]'$. Then, $\mathbf{g}\sim MVN_5(\mathbf{B},\boldsymbol{\Xi})$. Then, partition $\mathbf{g}$ as $\mathbf{g}=(\tilde{\mathbf{u}},y)$, where $\tilde{\mathbf{u}}=\left[\mathbf{u}',y^*\right]'$ and, correspondingly partition **B** and $\boldsymbol{\Xi}$ as $\mathbf{B}=(\mathbf{B}_{\tilde{\mathbf{u}}},B_y)'$ and $\boldsymbol{\Xi}=\begin{bmatrix}\boldsymbol{\Xi}_{\tilde{\mathbf{u}}} & \boldsymbol{\Xi}_{\tilde{\mathbf{u}}y}\\ \boldsymbol{\Xi}'_{\tilde{\mathbf{u}}y} & \Xi_y\end{bmatrix}$. By the conditioning property of the multivariate normal distribution, we then have $\tilde{\mathbf{u}}\,|\,y\sim MVN_4(\vec{\mathbf{B}}_{\tilde{\mathbf{u}}},\vec{\boldsymbol{\Xi}}_{\tilde{\mathbf{u}}})$, where $\vec{\boldsymbol{B}}_{\tilde{\mathbf{u}}}=\boldsymbol{B}_{\tilde{\mathbf{u}}}+\boldsymbol{\Xi}'_{\tilde{\mathbf{u}}y}\ \Xi_y^{-1}\left(y-B_y\right)$, $\vec{\boldsymbol{\Xi}}_{\tilde{\mathbf{u}}}=\boldsymbol{\Xi}_{\tilde{\mathbf{u}}}-\boldsymbol{\Xi}'_{\tilde{\mathbf{u}}y}\ \Xi_y^{-1}\boldsymbol{\Xi}_{\tilde{\mathbf{u}}y}$. Next, define observed threshold vectors for the individual as follows: $\vec{\boldsymbol{\psi}}_{low}=\left[\left(-\boldsymbol{\infty}_{I-1}\right)',\psi_{a\text{-}1}\right]'$ ($[I\times1]$ vector) and $\vec{\boldsymbol{\psi}}_{up}=\left[\left(\mathbf{0}_{I-1}\right)',\psi_a\right]'$ ($[I\times1]$ vector), where $-\boldsymbol{\infty}_{I-1}$ is a $(I-1)\times1$-column vector of negative infinities and $\mathbf{0}_{I-1}$ is another $(I-1)\times1$-column vector of zeros (I=4 in our simulations). Also, let $\vec{\boldsymbol{\omega}}_{\tilde{\mathbf{u}}}$ be a diagonal matrix containing the square root of the variance elements of $\ddot{\boldsymbol{\Xi}}_{\tilde{\mathbf{u}}}$, and let $\boldsymbol{\tau}_{low} = \left(\ddot{\boldsymbol{\omega}}_{\tilde{\mathbf{u}}}\right)^{-1}(\ddot{\boldsymbol{\psi}}_{low} - \ddot{\mathbf{B}}_{\tilde{\mathbf{u}}})$, $\boldsymbol{\tau}_{up} = \left(\ddot{\boldsymbol{\omega}}_{\tilde{\mathbf{u}}}\right)^{-1}(\ddot{\boldsymbol{\psi}}_{up} - \ddot{\mathbf{B}}_{\tilde{\mathbf{u}}})$, and $\ddot{\boldsymbol{\Xi}}^{*}_{\tilde{\mathbf{u}}} = \left(\ddot{\boldsymbol{\omega}}_{\tilde{\mathbf{u}}}\right)^{-1} \ddot{\boldsymbol{\Xi}}_{\tilde{\mathbf{u}}} \left(\ddot{\boldsymbol{\omega}}_{\tilde{\mathbf{u}}}\right)^{-1}$.

Let ϑ be the collection of parameters to be estimated: $\vartheta = [\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\delta}', \text{Vech}(\boldsymbol{\Sigma})]$, where the operator "Vech(.)" vectorizes all the non-zero elements of the matrix/vector on which it operates. Then the likelihood function for the individual may be written as:

$$L(\vartheta) = f(y \mid B_y, \Xi_y) \times \Pr\left[\ddot{\boldsymbol{\psi}}_{low} < \tilde{\boldsymbol{u}} < \ddot{\boldsymbol{\psi}}_{up}\right].$$

The expression above can be re-written as follows:

$$L(\vartheta) = \frac{1}{\sqrt{\Xi_y}} \times \phi\left(\frac{y - B_y}{\sqrt{\Xi_y}}\right) \times \left(\Phi_4(\boldsymbol{\tau}_{up}; \ddot{\boldsymbol{\Xi}}^{*}_{\tilde{\mathbf{u}}}) - \Phi_4(\boldsymbol{\tau}_{low}; \ddot{\boldsymbol{\Xi}}^{*}_{\tilde{\mathbf{u}}})\right),$$

where $\Phi_4(.,.)$ is the standard four-variate normal cumulative distribution (MVNCD) function.

3.3 Experimental Design

To compare and evaluate the performance of the traditional Cholesky decomposition and the separation-based spherical parameterization approaches, as well as compare the performance of the latter approach with different values for the scale parameter (λ) in the embedded logistic function, we undertake a simulation exercise for the mixed system described in the previous sections. To do so, we first generate 500 independent data sets, each with 3000 observations. A pre-specified value for the parameter vector ϑ is used to generate the samples, as discussed below.

In the setup, to focus on the covariance matrix estimation, we use a simple specification for the exogenous variable vectors $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{w}$. Specifically, we use a constant and three continuous exogenous variables in the $\mathbf{x}$ vector for the nominal variable. The values for the continuous variables are drawn from univariate normal distributions as follows: $x_2 \sim N(0, 0.75)$, $x_3 \sim N(0.25, 1.0)$, and $x_4 \sim N(0.5, 1.5)$. Thus, $\mathbf{x} = (1, x_2, x_3, x_4)$, and the systematic utility $\boldsymbol{\beta}\mathbf{x}$ for the nominal variable is specified as:

$$\boldsymbol{\beta}\mathbf{x} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{2c} & \beta_5 & 0 & 0 \\ \beta_{3c} & 0 & \beta_5 & 0 \\ \beta_{4c} & 0 & 0 & \beta_5 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.25 & 1.00 & 0 & 0 \\ -0.50 & 0 & 1.00 & 0 \\ +0.50 & 0 & 0 & 1.00 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{13}$$

The exogenous variables in the $\mathbf{z}$ and $\mathbf{w}$ vectors are specified to include a constant and a dummy variable. For the dummy variable d_1 in the $\mathbf{z}$ vector (for the grouped dependent variable), for each of the 500 datasets, we draw 3000 independent values from the standard uniform distribution. If the value drawn is less than 0.5, the value of '0' is assigned for the dummy variable. Otherwise, the value of '1' is assigned. For the dummy variable d_2 in the $\mathbf{w}$ vector (for the continuous dependent variable), for each of the 500 datasets, we again draw 3000 independent values from the standard uniform distribution. If the value drawn is less than 0.7, the value of '0' is assigned for the dummy variable (to create an asymmetry with more values of the dummy variable toward the value of zero than the value of one). The coefficient on the constant in the **z** vector is specified to be +0.5, while the coefficient on the dummy variable is specified to be +0.75. The constant in the **w** vector is specified to be +1.0, while the coefficient on the dummy variable is specified to be –1.5. Thus, we have the following:

$$\boldsymbol{\gamma}'\mathbf{z} = [\gamma_1 \quad \gamma_2]\begin{bmatrix}1\\ d_1\end{bmatrix} = [0.5 \quad 0.75]\begin{bmatrix}1\\ d_1\end{bmatrix} \text{ and } \boldsymbol{\delta}'\mathbf{w} = [\delta_1 \quad \delta_2]\begin{bmatrix}1\\ d_2\end{bmatrix} = [1.0 \quad -1.5]\begin{bmatrix}1\\ d_2\end{bmatrix} \tag{14}$$

Once generated, the exogenous variables are held fixed for the rest of the simulation exercise (that is, across all the 500 data samples, the same exogenous variable values are used).

Next, we generate, for each of the 3000 observations in each of the 500 data samples, a five-variate realization of the error term vector $\boldsymbol{\eta} = (\tilde{\boldsymbol{\varepsilon}}, \boldsymbol{\xi}, \zeta)'$ with a predefined positive-definite covariance structure ($\boldsymbol{\Sigma}$) such that both positive and negative covariances are represented. Importantly, restrictions are imposed on the covariance matrix, with elements σ_{13} and σ_{24} fixed to zero. In addition, we use a design covariance matrix that embeds both variance heterogeneity (across the diagonal elements) as well as relatively high correlations (because it is well known that estimation of mixed models becomes more challenging with high correlations). The design matrix is as below:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ & & & \sigma_{44} & \sigma_{45} \\ & & & & \sigma_{55} \end{bmatrix} = \begin{bmatrix} 1 & 0.90 & 0 & -1.00 & -1.00 \\ & 2.25 & 1.35 & 0 & -1.50 \\ & & 2.25 & 1.80 & 0.60 \\ & & & 4.00 & 2.40 \\ & & & & 4.00 \end{bmatrix} \tag{15}$$

Separating the above matrix into a correlation matrix and a diagonal scale matrix as:

$\boldsymbol{\Sigma} = \boldsymbol{\Omega} = \boldsymbol{\omega}\mathbf{R}\boldsymbol{\omega}$

$$\text{where, } \boldsymbol{\omega} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ & 1.5 & 0 & 0 & 0 \\ & & 1.5 & 0 & 0 \\ & & & 2.0 & 0 \\ & & & & 2.0 \end{bmatrix}, \text{ and } \mathbf{R} = \begin{bmatrix} 1 & 0.60 & 0 & -0.50 & -0.50 \\ & 1 & 0.60 & 0 & -0.50 \\ & & 1 & 0.60 & 0.20 \\ & & & 1 & 0.60 \\ & & & & 1 \end{bmatrix} \tag{16}$$

Note that the restrictions imposed on the covariance parameters translate to the correlation parameters, where elements $r_{1,3}$ and $r_{2,4}$ are restricted to be zero. With these restrictions in place, a total of 8 parameters shall be estimated in the correlation matrix, along with the 4 scale parameters, thus resulting in a total of 12 parameters to be estimated for the covariance matrix.

To simulate the multivariate outcome data, the five-variate error term realization for each observation and each variable is added to the appropriate systematic components (with no error term added to the utility of the first alternative in the nominal variable). For the nominal variable, the alternative with the highest resulting utility is declared as the chosen alternative for each of the 3000 observations. For the grouped variable, we use a set of four threshold values (resulting in five grouped categories) as follows: $\psi_1 = 0.25$, $\psi_2 = 0.5$, $\psi_3 = 0.75$, and $\psi_4 = 1.0$. Based on the value of y^* corresponding to the error realization, relative to the thresholds, the grouped category for each observation is determined. As discussed previously, 500 such datasets are generated. The estimations are undertaken both with the proposed approach that ensures the imposed restrictions (for three different λ values), and the traditional Cholesky decomposition approach, thus resulting in 2000 multivariate joint model estimations.

3.4 Performance Evaluation

Using the generated data, we evaluate the performance of the separation-based spherical parameterization that ensures the imposed restrictions on the covariance matrix as discussed above. In addition, we also evaluate the performance of the separation-based strategy of estimating the covariance matrix with different levels of scales (λ) in the spherical parameterization. Specifically, we consider three scale levels: 0.8, 1.0, and 1.2. Thus, for each of the 500 data sets generated, we estimate a total of 3 models with different parameterizations/scales. The performance of the models is evaluated using multiple metrics for recovering model parameters as well as for the actual predictions. The procedure is as follows:

(1) Estimate the parameters for each of the 500 datasets, for each of the traditional Cholesky parameterization and the separation-based parameterization with three different scale values. This will result in four estimations for each dataset. Estimate the standard errors. For each of the four sets of estimations, do the following:

(2) Compute the percentage of non-convergence estimations among the 500 datasets (for each of the traditional Cholesky parameterization and the separation-based parameterization). Next, undertake the following:

(3) Compute the mean estimate for each model parameter across each of the 500 datasets. Compute the **absolute percentage (finite sample) bias (APB)** as:
$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$$

(4) Compute the standard deviation of each parameter estimate across the 500 datasets, and label this as the **finite sample standard deviation or FSSD** (essentially, this is the empirical standard error).

(5) Compute the mean of standard errors for each model parameter across the 500 datasets, and label this as the **asymptotic standard error or ASE** (essentially, this is the standard error

of the distribution of the estimator as the sample size gets large, and is a theoretical approximation to the FSSD).

(6) Next, to evaluate the accuracy of the asymptotic standard error formula for the finite sample size used, compute the **relative efficiency (RE)** for each parameter relative to the corresponding finite sample standard deviation as:

$$RE = \frac{\text{ASE}}{\text{FSSD}} \tag{17}$$

(7) Examine the data fit at a disaggregate level by comparing the log-likelihood values at the convergence of the models. The model with the higher log-likelihood value is to be preferred because all the models have the same number of estimated parameters. Based on the log-likelihood values for each of the 500 runs (corresponding to the 500 datasets), compute a mean log-likelihood value.

3.5 Simulation Results

The overall summary of the simulation results is presented in Table 1. The three set of rows in Table 1 correspond to the sample statistics discussed above for the three scaling factors of 0.8, 1.0 and 1.2, respectively. Further, while we undertook estimations with the traditional Cholesky factorization approach, those results are not presented here (to conserve space). However, some of the discernible differences and advantages of the separation strategy that allows the imposition of restrictions on the correlation matrix are discussed in the subsequent sections. Also, the convergence rates across the 500 datasets are not reported simply because we did not encounter any situation where the model estimation did not converge (across the 2000 model estimation runs).

3.6 Summary of the Findings

The experimental setup involved the estimation of 10 mean parameters (i.e., $[\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\delta}']$ vector) and 12 covariance parameters. Thus, a total of 22 parameters were estimated. The estimated parameters are reported for the three scale levels in Table 1. Note that columns 2 through 11 contain the details of estimates of the mean parameters, whereas columns 12 through 21 correspond to the 10 covariance parameters (out of which two correlations were fixed to zero). Finally, the last four columns report the estimated scales (i.e., the square root of the diagonal elements of the covariance matrix).

As is evident from Table 1, the parameters are retrieved with reasonable accuracy for all three scale levels, with average APB values being 4.75%, 4.21%, and 2.78% (not shown in the table) for the scale levels of 0.8, 1, and 1.2, respectively. Across the 10 mean parameters, the APB values lie in the range of 0.03% to 19.84%, with the highest APB value of 19.84% corresponding to the constant for the third nominal alternative (that is, β_{3c}) for the scale level of 0.8. However, with higher scale levels, the bias in this parameter reduces to less than 1.1% (for the scale level of 1.2).

Further, the constant corresponding to the second nominal alternative is associated with a slightly higher percentage bias (in the range of 16.23% to 17.46% for the three scale levels), which could be due to the relatively small magnitude of the parameter value (parameter's true value is –0.25 and the corresponding estimated value is –0.29). The correlation matrix and the scale parameters are also retrieved with good accuracy, despite them entering in a complex, non-linear fashion in the likelihood function. In contrast, parameter estimates from the traditional Cholesky factorization approach are associated with significantly higher bias, especially in the estimated Cholesky factors, with APB values as high as 27.6% (and an average APB of 7.7% as compared to APB values in the range of 2.7% to 4.3% for the separation based spherical parameterization approach), thus highlighting the usefulness of the proposed approach. Besides, the traditional Cholesky factorization approach is not even applicable when specific correlation values are to be restricted to zero, which highlights the importance of the proposed separation-based strategy in the first place. However, similar to the separation-based spherical parameterization approach, we did not come across any convergence-related issues with the Cholesky parameterization approach.

From the standpoint of precision in the estimated parameters, the ASE and the FSSD values of the estimated parameters are not unreasonably large, thus indicating that the parameter estimates are precise. The ASE values indicate the efficiency of the estimates in large samples, whereas the FSSD indicates the level of precision across different samples. When the number of samples and the sample sizes are sufficiently large, the ASE values tend to the FSSD values since ASE values are essentially the approximations of the FSSD values. As expected, the ASE and FSSD values reported in Table 1 are reasonably close to each other, suggesting that the ASE is a good estimator of FSSD across the three scale levels of the spherical parameterization.

Table 1. Simulation Results

Number of datasets: 500; Sample size: 3000																								
Parameters	β_{2c}	β_{3c}	β_{4c}	β_5	β_6	β_7	γ_1	γ_2	δ_1	δ_2	r_{12}	r_{13}	r_{14}	r_{15}	r_{23}	r_{24}	r_{25}	r_{34}	r_{35}	r_{45}	σ_{22}	σ_{33}	σ_{44}	σ_{55}
True value	-0.25	-0.50	0.50	1.00	1.00	1.00	0.50	0.75	1.00	-1.50	0.6	0	-0.50	-0.50	0.60	0	-0.50	0.60	0.20	0.60	1.50	1.50	2.00	2.00
	Scale for the spherical parametrization = 0.8																							
Estimated value	-0.29	-0.40	0.46	0.97	0.88	0.91	0.49	0.74	1.00	-1.51	0.58	0 *(fixed)*	-0.49	-0.49	0.59	0 *(fixed)*	-0.51	0.61	0.20	0.60	1.30	1.35	2.01	2.00
APB	17.46	19.84	8.82	2.52	11.95	9.32	2.17	1.51	0.03	0.80	3.45	--	2.20	2.20	2.38	--	2.71	0.92	1.45	0.07	13.55	10.31	0.27	0.03
ASE	0.06	0.10	0.06	0.06	0.10	0.11	0.06	0.08	0.04	0.06	0.05	--	0.03	0.03	0.03	--	0.02	0.02	0.03	0.01	0.16	0.16	0.09	0.03
FSSD	0.05	0.11	0.05	0.06	0.10	0.07	0.05	0.07	0.04	0.06	0.04	--	0.03	0.03	0.04	--	0.02	0.02	0.03	0.01	0.16	0.12	0.09	0.03
RE	1.21	0.92	1.05	1.05	1.07	1.42	1.02	1.04	1.00	1.03	1.38	--	1.18	1.35	0.93	--	1.05	1.01	1.00	1.02	0.99	1.40	1.03	0.99
Avg. LL.	-11220.636																							
	Scale for the spherical parametrization = 1.0																							
Estimated value	-0.29	-0.44	0.46	0.99	0.91	0.94	0.48	0.74	1.00	-1.52	0.57	0 *(fixed)*	-0.49	-0.49	0.58	0 *(fixed)*	-0.51	0.61	0.20	0.60	1.33	1.38	1.98	2.00
APB	17.35	12.83	8.83	0.53	8.59	6.48	4.63	1.04	0.06	1.07	5.43	--	2.81	2.81	2.73	--	2.41	1.11	1.89	0.27	11.09	7.76	1.24	0.05
ASE	0.06	0.11	0.06	0.06	0.11	0.11	0.06	0.08	0.04	0.06	0.06	--	0.03	0.04	0.03	--	0.02	0.02	0.03	0.01	0.16	0.17	0.09	0.03
FSSD	0.05	0.07	0.05	0.05	0.12	0.12	0.06	0.08	0.04	0.06	0.04	--	0.03	0.03	0.03	--	0.02	0.02	0.03	0.01	0.18	0.19	0.09	0.03
RE	1.30	1.47	1.20	1.17	0.87	0.94	0.98	1.00	0.98	1.02	1.56	--	1.22	1.40	1.09	--	1.14	1.12	1.08	1.04	0.88	0.91	1.00	0.99
Avg. LL	-11220.068																							
	Scale for the spherical parametrization = 1.2																							
Estimated value	-0.29	-0.49	0.47	1.02	0.98	0.98	0.49	0.73	1.00	-1.50	0.56	0 *(fixed)*	-0.49	-0.49	0.58	0 *(fixed)*	-0.52	0.61	0.20	0.60	1.41	1.44	1.97	2.00
APB	16.23	1.09	5.90	2.20	2.49	1.95	1.79	2.13	0.16	0.04	7.23	--	2.59	2.59	2.71	--	3.44	0.93	1.45	0.41	6.24	3.67	1.53	0.05
ASE	0.06	0.12	0.06	0.06	0.12	0.12	0.05	0.08	0.04	0.06	0.06	--	0.03	0.04	0.03	--	0.02	0.02	0.03	0.01	0.17	0.18	0.09	0.03
FSSD	0.05	0.08	0.05	0.06	0.13	0.16	0.06	0.08	0.04	0.06	0.04	--	0.02	0.02	0.03	--	0.02	0.02	0.03	0.01	0.19	0.14	0.09	0.03
RE	1.23	1.45	1.16	1.11	0.88	0.73	0.96	0.99	0.95	1.02	1.57	--	1.25	1.43	1.09	--	1.10	1.09	1.07	1.06	0.88	1.24	0.99	1.00
Avg. LL	-11219.812																							

--: Not applicable. APB: Absolute percentage bias. ASE: Asymptotic standard error. FSSD: Finite sample standard deviation.
RE: Relative efficiency. Avg. LL: Average log-likelihood across the 500 datasets.

Lastly, the goodness-of-fit values were not significantly different across the three scale levels in the case of the spherical parameterization approach, with the likelihood value being the highest for the case with the scale level of 1.2. However, despite higher number of parameters in the traditional Cholesky factorization approach, the average log-likelihood value was lower than that in the spherical parametrization approach (an average log-likelihood value of –11222.59 with 24 parameters for the Cholesky decomposition as compared to an average log-likelihood of –11219.81 with 22 parameters for the separation-based spherical parameterization approach)[5].

Overall, the simulation results highlight the effectiveness of the separation-based strategy with spherical parameterizations of the Cholesky matrix in estimating the correlation matrix, especially when restrictions are imposed on specific correlation values. Even in situations when no restrictions are imposed on the covariance matrix, the proposed strategy can serve as an alternative to the traditional Cholesky factorization-based approach for estimating covariance matrices in multivariate models.

4 EMPIRICAL ANALYSIS

4.1 The Empirical Context: Analyzing the Usage (and the Extent of Use) of Non-Private Modes of Transport in Bengaluru, India

In this section, we illustrate the use of the separation strategy as well as the procedure developed for restricting specific correlation values to zero for an empirical application to analyzing usage (i.e., usage and the extent of use) of different non-private[6] modes of transportation in Bengaluru, a metropolis in the southern region of India.

Bengaluru, like most major cities around the globe, has a variety of non-private transport modes. These include public transport options such as buses and metro trains, intermediate public transit (IPT) options such as auto-rickshaws, ride-hailing services such as Ola and Uber, and shared ride modes such as shared taxis and other informal carpool options. This section presents a multivariate analysis of the socio-demographic determinants of individuals' choice and extent of usage of these non-private modes of transportation. The primary purpose of this analysis is to demonstrate the benefits of the separation strategy vis-à-vis the traditional Cholesky factorization approach for estimating correlation matrices with a priori structure imposed in such multivariate dependent variable models. We also briefly discuss substantive findings related to the influence of sociodemographic attributes on the usage of various non-private modes in Bengaluru, India.

4.2 Data

The empirical data for the analysis is drawn from the Ease of Moving Index survey conducted by the Ola Mobility Institute in India in 2018. The survey collected information on individuals' usual modes of travel and the frequency of usage of those modes. In addition, the survey collected

[5] Note that one cannot compare the log-likelihood values directly. However, since the number of parameters are higher for the Cholesky based approach, the AIC and BIC values will reflect the same trend as the log-likelihood values.

[6] We use the term "*non-private modes*" throughout this article to represent modes that are not personally owned and used. In the subsequent sections, we use the same term to represent the modes considered in the analysis.

individuals' socio-demographic information as well as their attitudes and perceptions of transit service.

Table 2. Descriptive Statistics of the Sample

Dependent outcomes				
	Transit	Ride-hailing	Shared ride	IPT
Percentage of those who used the mode	23.1%	11.1%	22.2%	12.9%
Extent of usage (for those who used)				
Monthly (1-2 times in a month)	9.1%	25.0%	21.3%	44.8%
Weekly (1-2 times a week)	29.5 %	52.3 %	46.2%	44.5%
Daily (more than 3 times a week)	61.4 %	22.7%	32.5%	10.7%

Descriptive statistics of the exogenous variables	
	Sample shares
Gender	
Males	57.9%
Females	42.1%
Age	
Age less than 20 years	21.7%
Age between 20 and 40 years	47.8%
Age more than 40 years	30.5%
Educational qualification	
Less than 10^{th} grade	1.7%
Between 10^{th} and 12^{th} grade	19.1%
Graduate	41.2%
Post-graduate and above	38.0%
Employment status	
Student	29.9%
Employed	34.0%
Unemployed	23.9%
Homemaker	12.1%
Household two-wheeler ownership	
Zero two-wheeler	36.1%
One two-wheeler	12.8%
Two or more two-wheelers	51.1%
Household car ownership	
Zero car	51.1%
One car	34.7%
Two or more cars	14.1%
Individual monthly income (in INR; across those who are employed)	
Less than 15K	5.8%
Between 15K and 30K	36.4%
Between 30K and 50K	40.3%
More than 50K	17.5%
Sample size (*N*)	2337

For this empirical context, we focused on individuals' usage and the extent of usage of non-private modes, including transit, ride-hailing services (such as Ola, Uber, etc.), shared ride (i.e., carpool, shared taxi, etc.), and intermediate public transport modes (i.e., IPT modes such as auto-rickshaws and e-rickshaws)[7]. The final sample used in the analysis comprises 2337 respondents. The details of the descriptive analysis of the sample are presented in Table 2. As can be observed from the percentages of individuals using different modes in Table 2, transit is the most used non-private mode considered in the analysis (i.e., more than 90% of the individuals who used transit used it at least 1-2 times a week, and more than 60% used transit on a daily basis). This is not surprising since Bengaluru has an extensive bus transit system, with more than 6000 buses plying daily with a daily ridership of more than 3 million (at the time of the survey). Also to be noted is that the percentages of adoption of each mode (first line of Table 2) do not sum to 100% across the four modes, because adoption is represented as a binary choice of whether that mode is part of the "usual" repertoire of modes chosen by individuals for their travel. These are not for a specific trip (such as the commute trip), but for overall usage (and hence the appearance of the frequency dimension for each mode too, conditional on adoption).

Ride-hailing services have a share of around 11%, with 45% of these users opting to use ride-hailing services only once or twice a month. While this appears to be a low usage, the ride-hailing services started in Indian cities around 2013. Some literature suggests that the rapid increase in the mode share of these services is leading to a decline in the public transit mode share (Ngo et al., 2021; Babar and Burtch, 2020).

As can be seen from Table 2, the estimation sample has a larger proportion of men than women, a larger representation of employed individuals and students (than unemployed and homemakers) and a higher proportion of individuals with at least a graduate degree. This could be one reason for a significant proportion of individuals (around 49%) not using these non-private modes of transport and possibly using personal modes such as cars and two-wheelers.

4.3 Analytical Framework Used for the Empirical Analysis

The analysis in this study is carried out using a joint, multivariate binary and ordinal response probit model. That is, we develop binary probit models to represent the discrete decision of whether a mode is used or not and an ordinal response probit model to represent the frequency of usage of each mode – as monthly, weekly, or daily – if the mode is used. As a result, the model framework involves four binary outcomes that correspond to the discrete choice of each of the four modes and four ordinal outcomes for the frequency of usage of each of the four modes. All these outcomes are modelled in a joint multivariate framework. Such a model structure involves the estimation of an eight-by-eight correlation matrix[8] (i.e., four dimensions corresponding to the discrete outcomes and the next four dimensions corresponding to the propensity of usage of each of the four modes in the analysis), with 28 correlation parameters.

There are several reasons why the traditional Cholesky decomposition approach may be difficult to implement while estimating such a correlation matrix. First, estimating a correlation

[7] While Bengaluru currently has an operational metro transit system, at the time of the data collection, the metro lines were still not fully in operation. As a result, the share of metro users in the data was very small (less than 2%), and hence, metro was not considered in our analysis.

[8] The model structure involves a correlation matrix, as opposed to a covariance matrix, as in Section 3. This is because the scale parameters are not identified (hence normalized to 1) for all the binary and ordinal outcomes.

matrix using the traditional Cholesky decomposition is infeasible since restricting the diagonal elements to one is not possible with the traditional Cholesky decomposition, unless the separation-based strategy is implemented. Moreover, as discussed in the introduction section, even the separation-based strategy runs into estimation issues unless spherical parameterization is also implemented. Second, estimating such a large dimensional correlation matrix (with 28 correlation parameters) may not be easy. Moreover, there are no behavioral/logical reasons why certain specific dimensions must be correlated. For instance, the propensity of choosing a specific mode is likely to be correlated with the corresponding propensity that represents its frequency of use. However, there are no substantive reasons why the discrete preference of a mode must be correlated with the propensity representing the frequency of use of another mode. Alternatively, restricting such correlations to zero, while practical both from the standpoint of ease of estimation and behavioral interpretation, is not possible with the traditional Cholesky decomposition approach. In such a situation, the separation-based strategy not only allows estimation of the correlation matrix by fixing the diagonal terms to one but also allows restricting certain specific correlation terms to zero. Such restrictions, in addition to being behaviorally consistent, also ease the estimation procedure by reducing the number of parameters to be estimated. Third, the ability to restrict specific correlation values to zero also allows one to drop statistically insignificant parameter estimates in the correlation matrix.

4.4 Empirical Estimation Results

The parameter estimates of the joint model are presented in Table 3. The final specification was carefully developed, supported by, as is inevitable in specification testing, insights from previous literature, parsimony in specification, and statistical fit/significance considerations. In particular, we began our specification analysis by examining the variables impacting each mode separately within a bivariate modeling system of mode adoption (a binary variable) and frequency of use of the mode (an ordered-response variable). In this mode-specific bivariate specification, we included all explanatory variables, considering alternative functional forms. We also estimated the correlation between the binary adoption and ordered frequency dimensions specific to that mode. Based on this bivariate model estimation, we retained all variable effects that had a t-statistic higher than 1. In our estimation trials, the correlation between the binary adoption of transit and its ordered frequency was insignificant (with t-statistic less than 1), possibly because of the relatively lower magnitude of the estimated correlation value. However, we decided to retain this correlation in the preliminary specification. Next, we took these results into the estimation of the multivariate eight-dimensional model across all the four modes.[9] As discussed earlier, in estimating this multivariate model, we made some correlation restrictions and eliminated variables that had a t-statistic of less

[9] In the experience of the authors, the effects and statistical significance of variables in marginal models for a subset of dependent outcomes will generally be carried over to more comprehensive multivariate models. That is, the introduction of additional correlation effects in unobserved factors, generally, will not change the magnitude and significance of the effect of a variable observed to be statistically insignificant in a marginal model. In fact, econometrically speaking, this result is the basis of the composite marginal likelihood inference approach for model estimation of large multivariate model systems (see Bhat, 2015).

than 1. Notably, the correlation between the binary adoption of transit and the ordered frequency turned out to be marginally significant and therefore, was retained in the final specification. The final multivariate model specification includes some variables that are not highly statistically significant, but which are included because of their intuitive effects and potential to guide future research and survey efforts in the field. This is particularly so because of the multi-dimensional nature of the model system and the rather skewed statistics toward non-use of each of the non-private modes. Also, due to identification considerations, no constants are estimated in the binary and ordered-response equations (because a full set of thresholds mapping the latent variables underlying the categorical outcomes to the outcomes themselves are estimated).

One additional point of note about our empirical specification. In the context of factors influencing mode choice decisions and frequency of use, besides the socio-demographic attributes, individuals' subjective perceptions can also play an important role (Hensher et al., 2003; Carrel and Walker, 2017; Johansson et al., 2006). While it is convenient to use these indicators directly as attributes in the model specification, they are often riddled with measurement errors, which are important to recognize (see Bhat and Dubey, 2014 and Deepa et al., 2022). However, in this study, since the focus is primarily on showcasing the ability of the proposed approach to restrict specific correlations to zero, we use the direct measures of these subjective perceptions (particularly, the perception toward reliability of transit) as attributes without recognizing that these measures could be prone to errors. Of course, recognizing these attitudes and perceptions as latent variables in the model to account for measurement errors, while also incorporating our strategy to restrict specific covariances to zero, is a worthwhile research direction to pursue in the future.

4.4.1 Influence of Exogenous Variables

In the context of the influence of gender, men are more likely to use shared rides and use it more frequently than women, possibly because of the safety and comfort issues that women might have when sharing rides with strangers. Interestingly, however, there were no significant differences between men and women in preferences toward other modes. In the context of age-related variables, while there were no significant differences in the likelihood of usage of transit across different age groups, younger individuals (age < 40 years) are more frequent users of transit than older (age >40) individuals. Also, individuals of age less than 20 years have a higher preference for shared ride modes than those in other age groups.

Individuals with higher education (i.e., at least a graduate degree) are less likely to use transit. Surprisingly, however, such individuals are also less likely to use ride-hailing services, possibly because such individuals own private vehicles and therefore do not prefer to use non-private modes of transportation. In the context of employment status, employed individuals do not show significant differences from unemployed people or homemakers in their usage of transit or shared-ride modes. But employed individuals show a lower preference for ride-hailing services and a higher preference for IPT modes, possibly because IPT modes such as auto-rickshaws were easily accessible and associated with shorter wait times as compared to ride-hailing options at the time of the survey (2018). Students are more likely to rely on non-private modes of transportation, with a higher preference for (and a higher frequency of usage of) transit and shared-ride modes.

Students also show a higher frequency of usage of IPT modes (if they use the IPT modes) than that of unemployed people or homemakers.

Individuals' perception of the reliability of transit also plays an important role, as individuals with positive perceptions of the reliability of transit services are more likely to use it whereas those with negative perceptions are less likely to use transit (Deepa et al., 2022). In the context of vehicle ownership, not owning a personal vehicle (either a two-wheeler or a four-wheeler) increases reliance on the non-private modes. While the influence of not-owning a two-wheeler is significant only on the likelihood of using transit and ride-hailing services (that too, the effects are marginally significant), owning cars significantly reduces the usage (and the extent of use of) of most non-private modes. Further, individuals from households with multiple cars are much less likely to use any of the non-private modes. These results are in line with other findings in the literature that increased car ownership levels reduce the usage of non-private modes of transportation (Thompson et al., 2002).

The information on income was available for only those individuals who were employed. Therefore, the effects of income on preferences of these modes were captured only for employed individuals. In this regard, as expected, individuals with a monthly income of less than ₹30,000 are more likely to use non-private modes than those with a higher monthly income. Further, interactions between gender and work status indicate that female students tend to rely more on non-private modes of transportation, particularly transit and ride-hailing. Also, female students who chose IPT modes are likely to use them more frequently than their male counterparts.

Table 3. Empirical Estimation Results

Explanatory variables	**Transit**		**Ride-hailing**		**Shared ride**		**IPT**	
	Discrete preference	Propensity of usage	Discrete preference	Propensity of usage	Discrete preference	Propensity of usage	Discrete preference	Propensity of usage
Constants	-0.90 *(-4.72)*	--	-1.04 *(-8.12)*	--	-1.06 *(-13.24)*	--	-1.19 *(-25.64)*	--
Socio-demographic variables								
Gender (Base: Female)								
Male	**	**	**	**	0.21 *(3.26)*	0.29 *(1.64)*	**	**
Age (Base: Age more than 60 years)								
Age less than 20 years	**	0.19 *(1.08)*	**	**	0.80 *(2.73)*	**	**	**
Age between 20 and 40 years	**	0.44 *(1.50)*	**	**	**	**	**	**
Edu. qualification (Base case: High school and below)								
Graduate and above	-0.74 *(-8.90)*	**	-0.19 *(-2.00)*	**	**	**	**	**
Employment status (Base: Unemployed and homemakers)								
Student	0.29 *(2.17)*	0.93 *(3.84)*	**	**	0.18 *(3.34)*	0.78 *(3.42)*	**	0.87 *(3.61)*
Employed	**	**	-0.37 *(-2.55)*	**	**	**	0.11 *(1.44)*	**
Vehicle ownership								
Zero two-wheeler ownership	0.09 *(1.12)*	**	0.10 *(1.26)*	**	**	**	**	**
Zero car ownership	0.99 *(11.80)*	0.43 *(1.88)*	0.19 *(1.80)*	0.51 *(2.48)*	0.24 *(2.83)*	**	**	-0.18 *(-1.30)*
Owning 2 or more cars	-0.53 *(-3.10)*	**	-0.52 *(-3.21)*	**	-0.16 *(-1.62)*	**	-0.21 *(-2.05)*	**
Perception of reliability in transit (Base: Transit services are somewhat reliable)								
Transit services are not reliable	-0.29 *(-3.63)*	**	--	--	--	--	--	--
Transit services are very reliable	0.47 *(4.24)*	**	--	--	--	--	--	--
Interaction effects								
Employed × Income less than ₹15K (*Base: Income > ₹30K*)	0.91 *(3.81)*	**	0.85 *(3.14)*	0.36 *(1.13)*	0.21 *(1.00)*	0.85 *(1.60)*	**	**
Employed × Income between ₹15K - ₹30K (*Base: Income > ₹30K*)	0.87 *(7.03)*	0.65 *(2.42)*	0.66 *(4.46)*	**	0.50 *(4.10)*	0.35 *(2.16)*	0.28 *(2.41)*	**
Male × Student	-0.33 *(-2.41)*	-0.51 *(-1.61)*	-0.43 *(-3.91)*	**	**	**	**	-0.69 *(-1.82)*
Male × Employed	**	-0.49 *(-1.70)*		**	**	**	**	**
Employed × Zero car ownership	**	0.77 *(3.50)*	0.32 *(1.50)*	**	-0.71 *(-5.23)*	**	**	**
Thresholds								
Monthly | Weekly	--	-0.09 *(-0.14)*	--	0.08 *(0.15)*	--	0.34 *(0.73)*	--	-0.36 *(-1.83)*
Weekly | Daily	--	1.15 *(1.81)*	--	1.52 *(3.86)*	--	1.59 *(5.12)*	--	1.04 *(2.66)*

**: Statistically insignificant. --: Not applicable.

4.4.2 Correlation Matrix

The parameter estimates of the correlation matrix are reported in Table 4. We should report here that when we implemented the separation-based strategy without spherical parameterization, the diagonal elements went into imaginary space, breaking down the estimation routine. This result is not surprising. since without the additional levels of parameterization (as in Equation (3), with further parameterization of the $h_{i,j}$ elements), the diagonal elements of the decomposed matrix $\mathbf{L}$ can venture into imaginary spaces, thus breaking down the estimation procedure. Therefore, we employed the separation-based strategy with spherical parameterization to estimate the correlation matrix.

Table 4. Estimated Correlation Matrix with Restrictions

		Discrete preference				Propensity of usage			
		Transit	Ride hailing	Shared ride	IPT	Transit	Ride hailing	Shared ride	IPT
Discrete preference	Transit	1	0.34 *(6.54)*	**	0.18 *(3.03)*	0.07 *(1.25)*	*0.00#*	*0.00#*	*0.00#*
	Ride hailing		1	**	0.27 *(4.00)*	*0.00#*	0.20 *(2.17)*	*0.00#*	*0.00#*
	Shared ride			1	0.21 *(4.26)*	*0.00#*	*0.00#*	0.40 *(1.33)*	*0.00#*
	IPT				1	*0.00#*	*0.00#*	*0.00#*	**
Propensity of usage	Transit					1	0.26 *(1.83)*	**	**
	Ride hailing						1	0.15 *(1.14)*	0.13 *(1.03)*
	Shared ride							1	0.53 *(5.30)*
	IPT								1

**: The estimated correlation was statistically insignificant. Hence dropped. #Restricted to be zero

However, estimating such a large correlation matrix with 28 correlation parameters is anyway not easy, particularly with datasets that are not very large. Also, as already discussed, there are no behavioral reasons as to why the discrete preference of a mode *i* should be correlated with the propensity of usage of another mode *j* ($j \neq i$). Therefore, we restricted such correlations to zero, thus reducing the number of parameters from 28 to 16. Furthermore, five of the estimated correlation parameters were estimated to be statistically insignificant and were restricted to zero in the final specification, which would not have been possible with the traditional Cholesky decomposition approach. In sum, we estimated 11 correlation parameters in the final model, thus substantially reducing the estimation burden.

From the estimated correlation matrix in Table 4, several observations can be made. First, the correlations between the discrete preferences of all the modes were positive, except that the correlations between transit and shared ride and ride-hailing and shared ride modes were insignificant. Intuitively, it is reasonable that the preferences of these modes are positively correlated, since several unobserved factors such as perception towards safety, cleanliness, etc. are likely to influence the preference of these modes in a similar manner. Similar trends are also observed in the propensity of usage of these modes, with a positive correlation between all pairs, except between transit and shared ride and transit and IPT, where the correlations were insignificant. Second, while correlations between the discrete preference of a mode i and the propensity of usage of another mode j ($j \neq i$) were restricted to be zero, such correlations between the discrete preference and the propensity of usage for a mode i were freely estimated. As expected, these correlations were estimated to be positive across all modes, except IPT, where the correlations were statistically insignificant. In essence, the correlation parameter estimates are all interpretable.

4.4.3 Goodness-of-Fit Measures

The log-likelihood of the final model was –5188.56 relative to a model with only constants in the binary mode adoption models and thresholds in the mode use frequency models being –5788.632. A likelihood ratio test indicates that our model outperforms the simple naïve constants-only model at any reasonable level of significance, indicating the explanatory value of our model as well as the importance of recognizing correlations due to unobserved factors.

5 CONCLUSIONS

Estimation of most multivariate model systems involves the estimation of covariance matrices. While most model systems rely on the traditional Cholesky decomposition approach to ensure positive definiteness of the covariance matrix, there are situations where in addition to the positive definiteness requirement of the covariance matrix, other restrictions (such as fixing the diagonal elements of the covariance matrix to specific values or restricting specific covariance elements to zero) on the elements of the covariance matrix are also needed. In such model systems, the usual Cholesky decomposition approach may not work. In this regard, a common strategy is to separate the covariance matrix into a diagonal scale matrix and a correlation matrix. However, applying Cholesky decomposition of the correlation matrix while ensuring the unit diagonal requirement can lead to situations where the diagonal elements of the correlation matrix do not always conform to the real line. To this end, a spherical parameterization of the Cholesky matrix is undertaken. The spherical parameterization ensures that the off-diagonal elements of the correlation matrix always lie between –1 to 1, which automatically ensures that the diagonal elements are always positive and lie on the real line. However, strategies to impose restrictions on the correlation elements have not been explored in any of the above parameterization approaches.

In this paper, we build on the above-discussed parameterization strategies to explore the possibilities of restricting elements of the correlation matrix to zero. To do so, we revisit the separation-based strategy with spherical parameterization of the covariance matrix and derive conditions to restrict the specific correlation values to zero. The proposed approach is then validated using extensive simulation runs. The effectiveness of the developed algorithm is highlighted in terms of accurate and precise recovery of parameters while ensuring that the specific correlation values are restricted to zero. In comparison with the traditional Cholesky decomposition approach, the simulation results indicate that the proposed strategy allows a more accurate recovery of the covariance parameters. Therefore, the proposed strategy can serve as an alternative to the traditional Cholesky decomposition approach.

Finally, the developed strategy is implemented in a joint, multivariate binary and ordinal response probit model system to analyze the usage of (and the extent of usage of) non-private modes of transportation in Bengaluru, India. The empirical results indicate that socio-demographic variables, such as gender, age, employment status, income, and ownership of personal vehicles, significantly influence the discrete preference as well as the extent of usage of these modes. Further, the estimation results reveal that the factors that influence the preference for a mode might be different than those that influence its extent of usage. Importantly, the estimation of the resulting correlation matrix is facilitated by the proposed separation-based strategy with spherical parameterization that allows imposing zero-correlation restrictions. The importance of the proposed strategy is highlighted in the ability to restrict specific correlation parameters that do not offer behaviorally explainable interpretations. Besides, restricting the correlation parameters reduced the number of parameters to be estimated, thus substantially reducing the estimation burden. Of course, further investigations into estimation approaches that are able to strictly adhere to the desired correlation restrictions is a fruitful avenue for additional research, as is an investigation of alternative normalization methods for identification.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support provided by the Ministry of Human Resource Development (MHRD) of the Government of India through its Scheme for Promotion of Academic and Research Collaboration in facilitating this research collaboration. The authors are also grateful to the Ola Mobility Institute for sharing the survey data, and to three anonymous reviewers for providing comments that helped improve the presentation in the paper.

REFERENCES

Babar, Y., & Burtch, G. (2020). Examining the heterogeneous impact of ride-hailing services on public transit use. *Information Systems Research*, 31(3), 820–834.

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10, 1281–1311.

Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50–77.

Bhat, C. R., & Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, 67, 68–85.

Bhat, C. R., & Lavieri, P. S. (2018). A new mixed MNP model accommodating a variety of dependent non-normal coefficient distributions. *Theory and Decision*, 84(2), 239–275.

Bhat, C. R., & Mondal, A. (2021). On the almost exact-equivalence of the radial and spherical unconstrained Cholesky based parameterization methods for correlation matrices. Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bhat, C. R., Sener, I. N., & Eluru, N. (2010). A flexible spatially dependent discrete choice model: Formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B*, 44(8–9), 903–921.

Carrel, A., & Walker, J. L. (2017). Understanding future mode choice intentions of transit riders as a function of past experiences with travel quality. *European Journal of Transport and Infrastructure Research*, 17(3).

Deepa, L., Mondal, A., Raman, A., Pinjari, A. R., Bhat, C. R., Srinivasan, K. K., Pendyala, R. M., & Ramadurai, G. (2022). An analysis of individuals' usage of bus transit in Bengaluru, India: Disentangling the influence of unfamiliarity with transit from that of subjective perceptions of service quality. *Travel Behaviour and Society*, 29, 1–11.

Dennis, J. E., & Schnabel, R. B. (1989). A view of unconstrained optimization. In *Handbooks in Operations Research and Management Science*: *Vol. 1*. *Optimization* (pp. 1–72).

Dias, F. F., Lavieri, P. S., Sharda, S., Khoeini, S., Bhat, C. R., Pendyala, R. M., Pinjari, A. R., Ramadurai, G., & Srinivasan, K. K. (2020). A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C*, 114, 643–656.

Higham, N.J. (2009). Cholesky factorization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(2), 251–254.

Hensher, D. A., Stopher, P., & Bullock, P. (2003). Service quality—developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A*, 37(6), 499–517.

Jiryaie, F., & Khodadadi, A. (2019). Simultaneous optimization of multiple responses that involve correlated continuous and ordinal responses according to the Gaussian copula models. *Journal of Statistical Theory and Applications*, 18(3), 212–221.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10), 2177–2189.

Johansson, M. V., Heldt, T., & Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40(6), 507–525.

Kohli, N., Peralta, Y., & Bose, M. (2019). Piecewise random-effects modeling software programs. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1), 156–164.

Leonard, T., & Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4), 1669–1696.

Lindstrom, M. J., & Bates, D. M. (1988). Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404), 1014–1022.

Madar, V. (2015). Direct formulation to Cholesky decomposition of a general nonsingular correlation matrix. *Statistics & Probability Letters*, 103, 142–147.

McNeish, D., & Bauer, D. J. (2022). Reducing incidence of nonpositive definite covariance matrices in mixed effect models. *Multivariate Behavioral Research*, 57(2–3), 318–340.

Müller, D., & Czado, C. (2018). Representing sparse Gaussian DAGs as sparse R-vines allowing for non-Gaussian dependence. *Journal of Computational and Graphical Statistics*, 27(2), 334–344.

Ngo, N. S., Götschi, T., & Clark, B. Y. (2021). The effects of ride-hailing services on bus ridership in a medium-sized urban area using micro-level data: Evidence from the Lane Transit District. *Transport Policy*, 105(February), 44–53.

Ong, V. M.-H., Nott, D. J., & Smith, M. S. (2018). Gaussian variational approximation with a factor covariance structure. *Journal of Computational and Graphical Statistics*, 27(3), 465–478.

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

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, 87(2), 425–435.

Pourahmadi, M., & Wang, X. (2015). Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor. *Statistics & Probability Letters*, 106, 5–12.

Rana, T. A., Sikder, S., & Pinjari, A. R. (2010). Copula-based method for addressing endogeneity in models of severity of traffic crash injuries: Application to two-vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2147(1), 75–87.

Rebonato, R., & Jäckel, P. (2000). The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, 2(2), 17–27.

Schöttle, K., and Werner, R. (2004). Improving the most general methodology to create a valid correlation matrix. *WIT Transactions on Ecology and the Environment*, 77.

Sidharthan, R., & Bhat, C. R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, 44(4), 321–349.

Srinivasan, S., & Bhat, C. R. (2005). Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation*, 32(5), 523–544.

Teixeira-Pinto, A., & Harezlak, J. (2013). Factorization and latent variable models for joint analysis of binary and continuous outcomes. In de Leon, A.R., and K.C. Chough (Eds.) *Analysis of Mixed Data: Methods & Applications* (1st ed.), pp. 81–91, Chapman and Hall/CRC, Boca Raton, FL.

Thompson, B. J., Perone, J. S., & Gabourel, K. (2002). Transit non-user survey: Restful riding rather than stressful driving (No. NCTR-416-08.4). University of South Florida. Center for Urban Transportation Research.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Tsay, R. S., & Pourahmadi, M. (2017). Modelling structured correlation matrices. *Biometrika*, 104(1), 237–242.

van Oest, R. (2021). Unconstrained Cholesky-based parametrization of correlation matrices. *Communications in Statistics - Simulation and Computation*, 50(11), 3607–3613.