A Flexible Non-Normal Random Coefficient Multinomial Probit Model: Application to Investigating Commuter's Mode Choice Behavior in a Developing Economy Context

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

Aupal Mondal
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Tel: 1-512-471-4535; Email: aupal.mondal@utexas.edu

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
There is growing interest in employing non-normal parameter distributions on covariates to account for random taste heterogeneity in multinomial choice models. In this study, we propose a flexible, computationally tractable, structurally simple, and parsimonious-in-specification random coefficients multinomial probit (MNP) model that can accommodate non-normality in the random coefficients. Our proposed methodology subsumes the normally distributed random coefficient MNP model as a special case, thus eliminating the need to a priori decide on the distributional assumption for each coefficient. The approach employs an implicit Gaussian copula to combine the univariate coefficient distributions into a multivariate distribution with a flexible dependence structure. Using our new flexible MNP framework, we investigate the commute mode choice behavior for workers in the city of Bengaluru, a metropolitan city in southern India. Results from our analysis indicate that sociodemographic variables, commute characteristics, and mode-related attributes significantly impact the commute mode choice decision. Importantly, our results indicate the presence of unobserved taste heterogeneity in the sensitivities to the travel time and travel cost variables; moreover, the distribution of the travel time coefficient is found to be significantly non-normal. In terms of data fit, our proposed model statistically outperforms the traditional MNP model as well as an MNP model that imposes normality on the travel time coefficient. The pitfalls of ignoring non-normality in the distribution of parameters are also discussed, as are several policies to promote a shift from private modes of transportation to more sustainable public transportation/walk modes.

Keywords: Non-normal distribution; Unobserved heterogeneity; Gaussian copula; Mode choice; Developing economy.

1 INTRODUCTION

Among the choice models used in the literature, the multinomial logit (MNL) model continues to be the "workhorse" and has been the most widely used structure for modeling travel mode choices. But, over the past several decades, many extensions of the MNL model have been proposed in the literature, which relax the independence and identically distributed (IID) kernel error terms and homogenous response assumptions of the MNL model. Bhat (2020) provides a detailed history of the developments and progress of the field over multiple decades. Among all the choice models, two general econometric model structures have stood out as being the new standards for flexibility in capturing random taste heterogeneity. One is the mixed logit model that has seen several applications in the context of travel mode choice (see for example, Guo et al., 2020, Patil et al., 2020, Parmar et al., 2021, Dias et al., 2022). The second is the multinomial probit (MNP) model (Daganzo, 2014). The MNP model can accommodate random taste heterogeneity in a way similar to the mixed logit, by specifying IID (independent and identically distributed) kernel normal error terms over which a mixing distribution is specified. In this sense, the mixed IID MNP and the mixed logit models have essentially the same structure; the mixed logit has received more attention because, conditional on random taste heterogeneity, it collapses to the familiar closed-form MNL, while the mixed IID probit, again conditional on random taste heterogeneity, requires a one-dimensional integration (see Bhat, 2003). However, as discussed at length in Bhat (2018), when a multivariate normal random taste heterogeneity (mixing) distribution is used, the latter can be much more computationally efficient when the number of random coefficients (say K) is higher than the number of alternatives (say I), as is not uncommon in travel choice modeling. This is because of the conjugate nature of the multivariate normal distribution under additivity, which results in the evaluation of only an (I–1) dimensional truncated (from above; that is, the multivariate normal cumulative distribution (MVNCD) function) in the MNP case relative to a K-dimensional untruncated integral in the mixed logit case (see Figure 1). Besides, in social/spatial interaction contexts or when modeling unordered choices with other types of outcome variables (such as ordered choice or counts), the use of a normal error kernel has substantially more appeal than the extreme value error kernel.

More recently, there has been renewed interest in the use of the normal error kernel in discrete choice modeling, thanks also to the development of new analytic methods to estimate the multivariate normal cumulative distribution function (Bhat, 2018; Wang et al., 2023). Patil et al. (2017) show that these analytic approaches can be much faster than traditional maximum simulated likelihood (MSL) approaches used for the mixed logit models. Also, in the last two decades, there has been growing interest in considering flexible non-normal (mixing) distributions for coefficients on covariates to account for random taste heterogeneity, because the typical normal (mixing) distributional assumptions for model parameters may not be appropriate (see, for example, Bhat, 2000, Train and Sonnier, 2005, Greene et al., 2006, Vij and Krueger, 2017, Bhat and Sidharthan, 2012, and Bhat and Lavieri, 2018). In the context of a travel mode choice model, the most common example is the sensitivity to travel cost, which is expected to be negative; however, the normal distributional assumption does not ensure this negativity constraint. To address this issue, several researchers have proposed the use of bounded distributions such as the lognormal distribution. Moreover, the heterogeneity in sensitivity to other exogenous variables, even if can be unbounded over the real line, may follow a distribution which may be skewed to the right or left (and not necessarily be normally distributed); in such cases, assuming a normal distribution for the response coefficients will generally lead to biased estimates (see Balcombe et al., 2009 and Torres et al., 2011). To address this issue, researchers have used non-normal parameter distributions to provide flexibility. The consideration of such bounded/flexible mixing distributions has the added benefit of avoiding potential misspecification consequences such as poorer data fit, inaccurate trade-off computations, and misinformed policy evaluations.[1]

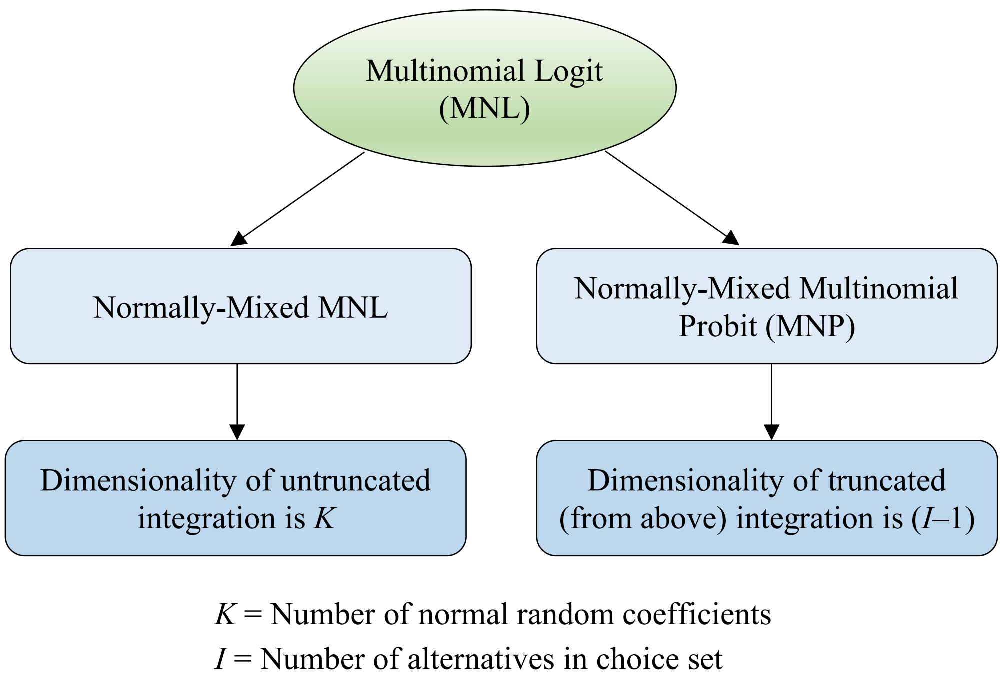

Figure 1. Two typical normally-mixed models in the discrete choice literature

In the discrete choice literature, the mixing distribution for unbounded coefficients, when allowed to be non-normal, is generally either modeled as (a) a discrete-valued random variable vector, or (b) a continuous parametric random variable vector, or (c) a combination of the two (see Figure 2). The method of modeling the mixing distribution as a discrete-valued random variable vector (see left side branch of Figure 2) may itself be achieved in one of two broad ways. The first is to assume a finite number of segments, each segment having a specific fixed coefficient vector in the population. Individuals belong to a specific segment, but this segment membership is not observable, and so a probabilistic model of segment membership is layered over the segment-

[1] The MNP approach, once again has a distinct advantage from a computational standpoint over the mixed logit in situations where there are fewer alternatives than random coefficients, and many random coefficients may be specified to be normal with few left for testing for non-normality. This issue is discussed at length in Bhat and Sidharthan (2012) and Bhat and Lavieri (2018).

specific choice model. This approach (see leftmost twig in Figure 2) corresponds to the latent class of models, as proposed by Bhat (1997) in a travel mode choice context (see also Greene and Hensher, 2003 and Train, 2009). Latent class models are particularly useful in cases when the entire mixing vector (across coefficients) takes a finite and small number of possible value states. On the other hand, if a discrete distribution is considered separately for each individual random coefficient, the result is a non-parametric estimation (see Bastin et al., 2010, Cherchi et al., 2009, Dong and Koppelman, 2014, Krueger et al., 2020, and Bauer et al., 2022). Such discrete (non-parametric) specifications (see right twig under the first branch of Figure 2) allow consistent estimates of the observed variable effects under broad model contexts by making regularity and smoothness (for instance, differentiability) assumptions on an otherwise distribution-free density form. But the flexibility of these methods comes at a high inferential cost because of parameter profligateness arising from series-based or other approximations to the density function (Mu and Zhang, 2018, Denzer, 2019). Also, consistency is achieved only in very large samples, parameter estimates have high variance, and the computational complexity/effort can be substantial (Mittelhammer and Judge, 2011).

The method of modeling the mixing distribution as a continuous multivariate non-normal vector (see center branch of Figure 2) typically takes the form of pre-specifying a multivariate non-normal distribution (such as a multivariate skew-normal distribution to introduce asymmetry and skew, or a multivariate t-distribution to introduce fat tails). An example of this approach is Bhat and Sidharthan (2012), who employ a multivariate skew-normal distribution for taste sensitivity parameters (see leftmost twig of the center branch in Figure 2). The skew normal distribution they use can replicate a variety of non-normal density shapes with heavier left/right tails as well as high/low modal values. The skew distribution also includes the normal distribution as a special case and the log-normal distribution as a limiting case, thus allowing for both boundedness as well as flexibility. Another more general continuous multivariate non-normal approach for the mixing distribution tests a variety of different univariate parametric distributions for each mixing coefficient (the parametric distributions can be different for different coefficients), and then employs a copula to tie the many univariate continuous distributions into a multivariate non-normal distribution (see center twig of the center branch in Figure 2). Bhat and Lavieri (2018) adopt such a general structure for accommodating non-normal distributions in a mixed random coefficients MNP model, combining the different univariate distributions through a Gaussian copula. A Gaussian copula has several advantages, including (a) a flexible dependence structure across coefficients and (b) ease of simulation compared to other copula structures (see Bhat and Eluru, 2009). Bhat and Lavieri (2018) proceed to use a hybrid Maximum Simulated Likelihood (MSL) – Maximum Approximate Composite Marginal Likelihood (MACML) inference approach to estimate their model. Although a flexible approach, the Bhat and Lavieri (2018) approach requires the analyst to assume a continuous distribution for each parameter prior to estimation (or at least test many different combinations prior to imposing a specific distributional assumption on each coefficient). A common distinct advantage of these continuous mixing distributions, relative to the discrete mixing distributions, is parsimony in the number of parameters to be estimated.

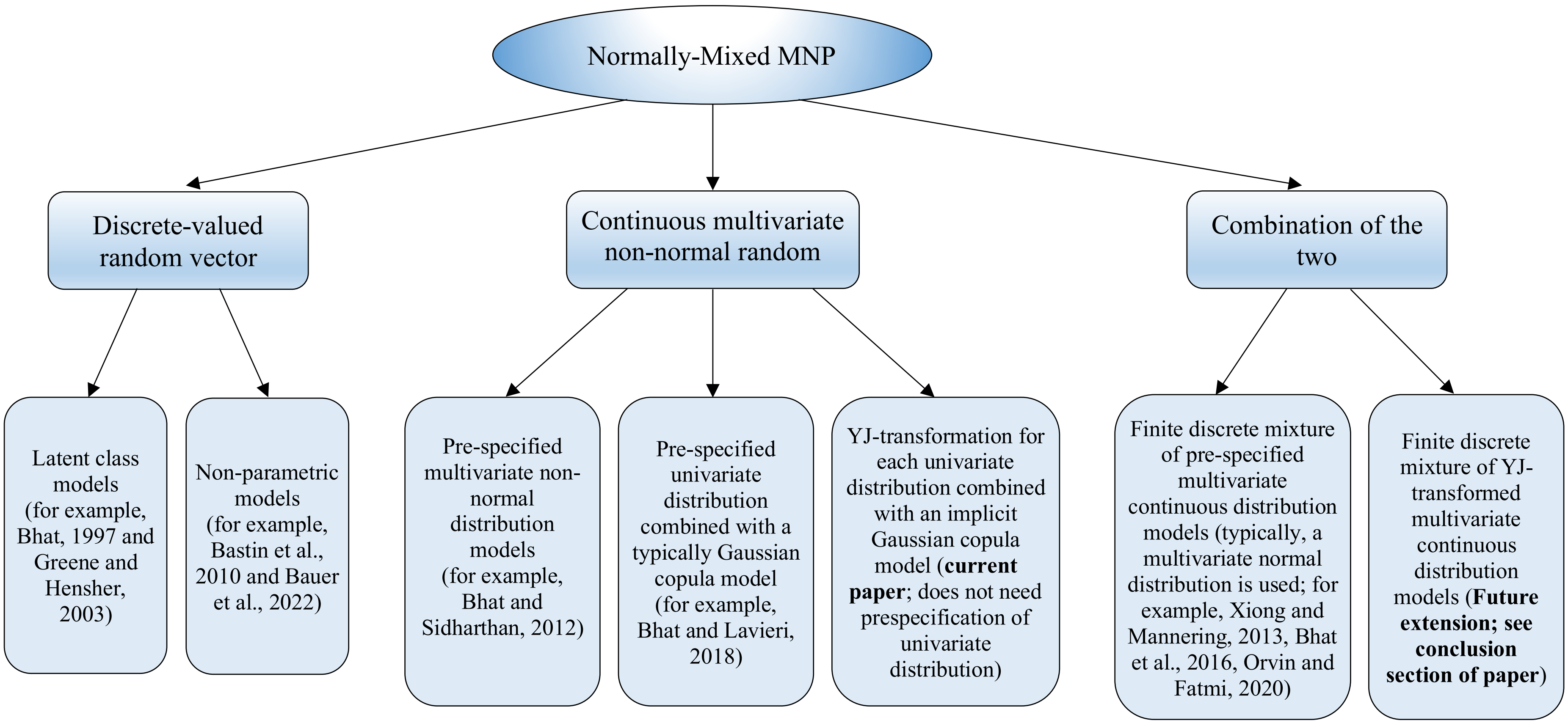

Figure 2. Discrete choice models allowing for non-normal random coefficients

The method of modeling the mixing distribution as a combination of a discrete and continuous distribution vector (see right side branch of Figure 2) typically takes the form of a finite discrete mixture of parametric distributions, including mixture-of-normals, mixture-of-skew-normals, and mixture-of-skew-t distributions. Such approaches may be considered as semiparametric approaches. Of these, the mixture-of-normals has been most commonly used see Xiong and Mannering, 2013, Bhat et al., 2016, Buddhavarapu et al., 2016, Li et al., 2018, and Orvin and Fatmi, 2020). Theoretically speaking, with a large number of components, this mixture distribution can mimic literally any multivariate density function, including the fully nonparametric distribution. Therein lies the challenge. Having a large number of components leads to (a) substantial non-smoothness in the density function and resulting computational challenges, (b) inefficient use of model parameters and resulting estimation challenges (regardless of the inference procedure used) due to high sensitivity to starting parameters/prior distribution assumptions, (c) the danger of the model becoming singular because of identification issues, and (d) potential overfitting tendencies of the inference approaches that can then result in large local maxima and potentially unbounded likelihood functions (see, for example, McLachlan and Rathnayake, 2014, Zhang and Huang, 2015, Hao and Kasahara, 2022, and Cai and Xu, 2023). Due to these reasons, and as stated by Rossi (2014), a statement that remains valid even today, "the really interesting question is not whether the mixture-of-normals can be the basis of a non-parametric density estimation procedure, but, rather, if good approximations can be achieved with relative parsimony." Similarly, in a more recent review of mixture models by McLachlan et al. (2019), they state "Arguably the most obdurate methodological problem associated with mixture distributions is that of identifying the number of components involved in the distribution underlying a set of data". Furthermore, when the target multivariate density has substantial skew, particularly in multiple components, the mixture-of-normals approach can provide distorted and misleading inferences because of overfit problems and weak identification of the many components (due to the need for an unnecessarily high number of components to mimic skewness), which contribute even more to computation/inference problems especially when working with limited sample sizes (Lin et al., 2007, Fruhwirth-Schnatter and Pyne, 2010, Everitt, 2013, Lin et al., 2016, Gallaugher et al., 2020, Smith et al., 2020, and Dong et al., 2023). Of course, one can move to more general component distributions, such as mixtures of skew-normal, skew-t, and other asymmetric and fat-tailed distributions from families of parametric distributions such as the generalized hyperbolic (GH) class; these have the benefit of more appropriately representing the target distribution with parsimony, but bring additional inference and computational challenges, while also leaving unclear about the type of skewness shapes capable of being handled by the different mixing distributions (see Fruhwirth-Schnatter and Pyne, 2010 and Lee and McLachlan, 2022). Besides, as with non-parametric approaches, this general class of semiparametric approaches attain favorable statistical asymptotic properties only when implemented using a large sample size (see Dong and Lewbel, 2015, Mu and Zhang, 2018). In addition, issues of interchangeability of component labels (known as the label-switching problem) need to be carefully addressed through appropriate constraints. As discussed in McLachlan et al. (2019), this is particularly so in a Bayesian framework because of the use of posterior simulations to make inferences.

Overall, non-parametric (discrete) series-based or similar approximations to the density function offer the most flexibility, but are saddled with profligateness and computational/inference challenges. Somewhere closer to the non-parametric distribution with its flexibility but also computational/stability challenges lies the finite discrete mixture-of-parametric distribution class (including mixture-of-normals, mixture-of-skew-normals, and mixture-of-t-distributions, with the mixture-of-normals being much more limited in flexibility than the mixture-of-skew-normals and mixture-of-t-distributions). On the other hand, continuous multivariate non-normal distributions, especially based on the copula approach of Bhat and Lavieri (2018), may offer somewhat less flexibility, but have definitive parsimony, computation, and inference advantages. However, although substantially more flexible than the multivariate normal mixing approach, the Bhat and Lavieri (2018) approach requires the analyst to assume a continuous distribution for each parameter prior to estimation. In this study, we propose an even more flexible, computationally tractable, structurally simpler and parsimonious-in-specification non-normal distribution along each univariate dimension (that is, each random coefficient), which are then tied together across different random coefficients within a Gaussian copula approach (this methodology is listed as the third twig of the center branch labeled "Continuous multivariate non-normal random vector" in Figure 2). Our proposed methodology accounts for the normal distribution as a special case of the non-normal flexible distribution, as well as can mimic a whole range of skewed and fat-tailed univariate distributions, thus eliminating the need to *a priori* decide on the distributional assumption for each coefficient. To do so, we use the Yeo and Johnson (2000) or the YJ transformation for each random coefficient, which extends the well-known Box-Cox transformation of a non-normal random variable/parameter into a normal random variable/parameter using a single parameter. Moreover, this transformation immediately facilitates the use of an implicit Gaussian copula to combine the univariate coefficient distributions into a multivariate distribution with a flexible dependence structure, thereby harnessing the advantages of the Gaussian copula discussed earlier. To our knowledge, our study constitutes the first attempt to introduce this transformation approach for random coefficients in a discrete choice model.

Important to note is that our copula-based method for random coefficients based on the YJ transformation offers much more efficiency than other multivariate nonparametric and semiparametric distributions for accommodating an unknown multivariate distribution with strict unimodality along each univariate dimension. In particular, the YJ transformation appears to do remarkably well in comparison to other approaches to accommodate skewness. As an illustration, Gallaugher et al. (2020) have compared, using Mardia's multivariate skewness and kurtosis metrics (Mardia, 1970) and multiple datasets for cluster analysis, the performance of two types of normal variance mean-mixture based skew distributions with two transformation approaches (including the YJ transformation used in this paper). They conclude that "From the analyses on a variety of datasets…, it appears that no one method consistently outperforms the others and usually the performance is very similar if not identical". Further, they observe that the transformation method has the distinct advantage of being much more parsimonious, while also performing at least as well as the more complicated mixtures of skew distributions. Similarly, Smith et al. (2020) state the following "Our empirical work shows that the Yeo-Johnson transformation is particularly effective and is quickly calibrated using stochastic gradient ascent (SGA); in most cases, faster than calibrating the elliptical or skew-elliptical distributions themselves on the parameter vector." Of course, we must admit that, while parsimonious and offering substantial flexibility to accommodate unimodal skew and fat-tailed distributions, the YJ approach is unable to accommodate distributions with multimodality along one or more dimensions. We leave this for future research, though we suggest one approach in the conclusions section to accommodate multimodality in random coefficients through the fusion of the YJ transformation approach with the finite discrete mixture approach.

Our proposed method is applied to understand commuters' mode choice behavior within the context of a fast-evolving urban mobility landscape of a developing economy, as discussed next.

1.1 Urban Mobility and Mode Choice Models in Developing Economies

The world's urban population, as a percentage of the total population, has seen a steady increase over time, from 47.4% in 2000 to 56.1% in 2020 (World Bank, 2022). Much of this increase may be traced to the increasing urbanization trends in developing countries. For example, according to a World Bank statistic (World Bank, 2022), India alone increased its urban population percentage from 27.6% in 2000 to 34.9% in 2020, representing an urban population percentage increase of 26.4%. Furthermore, not only is the urban population increasing in developing countries, but so is the ownership and use of private vehicles. Again, taking the example of India, the number of registered four-wheelers per 1000 persons increased from 6.6 in 2001 to 28.1 in 2019, representing a percentage increase of about 325% in four-wheeler registration in less than two decades (Ministry of Road Transport and Highways, 2021)! This combination of an increase in urban population and private vehicle ownership has resulted in substantial stress on the transportation infrastructure. The concomitant rise in traffic congestion and mobile-source emissions, along with the limited ability to expand the existing transportation infrastructure, has renewed calls for urban policies to reduce private vehicle usage through urban planning initiatives that promote the use of more sustainable public transportation and non-motorized modes of transportation. The effectiveness of such incentives may be evaluated using travel mode choice models that provide insights into the effects of socio-demographics, built environment characteristics, and transportation-related attributes on individual mode decisions. Such mode choice and travel behavior-related studies abound in the scholarly literature on transportation planning and travel demand. However, compared to the western world (primarily North America and Europe), developing countries have seen less attention in the area of mobility choice analysis (Masoumi, 2019). At the same time, the travel context in developing countries is quite different from that of the western world due to (a) lower levels of private-vehicle ownership, with a high prevalence of "captive" non-private travel mode riders, (b) high levels of two-wheeler usage as a private motorized mode, especially in South-Asian cities, (c) the presence of informal modes of transportation (often termed as "intermediate public transport" or "para-transit"), such as auto-rickshaws in several Indian cities, (d) a rather inadequate walking and bicycling infrastructure, and (e) differences in socio-economic characteristics and attitudes. Thus, there is a clear need for more emphasis on mobility analysis in developing economies.

1.2 Current Paper in Context

Motivated by the discussion above, the overall objective of the current study is to propose a flexible new discrete choice model and utilize the model to understand commuters' mode choice decisions in the context of a developing economy. In doing so, this paper contributes to the literature in at least three ways. *First*, we propose a parsimonious, computationally tractable, and easy-to-specify methodological framework based on a random parameter MNP model that allows non-normality in parameters without the need to prespecify a specific mixing distribution. The framework subsumes the case of normal random parameters as a special case. For the estimation of the model, we use a hybrid simulation approach using a combination of Halton draws for the mixing distribution (see Halton, 1960, Bhat, 2001, and Bhat, 2003), and Bhat's (2018) analytic methods for computing multivariate normal integrals. To our knowledge, this is the first such model proposed in the econometric literature. *Second*, we apply the proposed model to investigate commuters' mode choice behavior for the case of a developing economy. In particular, we analyze workers' mode choice behavior in the city of Bengaluru, a fast-growing technology hub in southern India. Through the empirical analysis, we explore the advantages of non-normal mixing distributions on the coefficients of level-of-service variables and the consequent effect on the resulting valuation of travel time savings (VTTS) measures. *Third*, we go beyond simply presenting the model estimation results to compute pseudo-elasticity measures that provide the direction and magnitude effects of exogenous variables. We then extract information from these elasticity measures to identify potentially promising policies to reduce peak-period traffic congestion and increase the use of public transportation modes.

The rest of the paper is arranged as follows: Section 2 describes the methodological framework proposed in this study. Section 3 contains the entire model application section including the survey sample description, model results, data fit measures, and policy implications. Concluding remarks are provided in Section 4.

2 METHODOLOGY

2.1 The YJ Transformation

The Yeo and Johnson (2000) or the YJ transformation, used to transform non-normal data to normality and symmetry for the marginal distribution, has recently been gaining increasing attention due to its robustness and effectiveness in simplifying computations in econometric analysis (see for example, Smith et al., 2020 and Bhat and Mondal, 2022). The YJ transformation has the distinct advantage of being a single-parameter transformation (so that only one additional parameter needs to be estimated to transform a non-normal distribution to a normal distribution for each margin, which leads to parsimony in estimation). Over a wide variety of skewed/fat-tailed situations, and through both simulation and empirical exercises, it has been shown to be one of the most effective transformations (if not the most effective transformation) relative to other transformations such as the Manly transformation, the Tukey transformation, and the arc-sine transformation (see Osborne, 2010, Jadhav et al., 2023, Watthanacheewakul, 2021, and Marimuthu et al., 2022). It extends the restrictive Box-Cox transformation (which is applicable only to the positive half of the real line; Box and Cox, 1964) to the entire real line, and, by so doing, brings the generality of the BC transformation over the positive real line to the entire real line (note also that, when confined to the positive line, the YJ transformation is the BC transformation, thus subsuming many distributions on the positive line as special cases -- the square root transformation is the BC transformation with a power parameter of 0.5; the corresponding power parameter for the inverse transformation is -1; and the natural logarithm transformation power parameter is 0, which corresponds to the log-normal distribution).

The YJ transformation of a random variable/parameter Z_l to an assumed normal random variable/parameter G_l (with a mean parameter of μ_l and variance of σ_l^2) on the real line is as follows, with an additional parameter $0<\lambda_l<2$:

$$G_l \sim N(\mu_l, \sigma_l^2) = t_{\lambda_l}(Z_l) = \begin{cases} -\dfrac{(-Z_l+1)^{2-\lambda_l}-1}{2-\lambda_l} & \text{if } Z_l < 0 \\ \dfrac{(Z_l+1)^{\lambda_l}-1}{\lambda_l} & \text{if } Z_l > 0 \end{cases} \tag{1}$$

The transformation above is for a non-normal variable/parameter that can take values over the entire real line to the normal distribution. The implied reverse transformation is as follows:

$$Z_l = t_{\lambda_l}^{-1}(G_l) = \begin{cases} 1-\left[1-(2-\lambda_l)G_l\right]^{\left(\frac{1}{2-\lambda_l}\right)} & \text{if } G_l < 0 \\ \left[1+G_l\lambda_l\right]^{\left(\frac{1}{\lambda_l}\right)}-1 & \text{if } G_l > 0 \end{cases} \tag{2}$$

The transformation above allows for an asymmetric distribution for Z_l relative to the traditional normal distribution. To illustrate, Figure 3 plots Z_l for μ_l=0 and σ_l^2=1, and for different values of λ_l ($0<\lambda_l<2$). When $0<\lambda_l<1$, Z_l is skewed to the right with a thicker right tail, while if $1<\lambda_l<2$, Z_l is skewed to the left with a thicker left tail. When λ_l=1, the normal distribution is returned for Z_l. Thus, the YJ transformation allows for skew and fat tails, depending on the estimated value for λ_l. The transformation is quite general, and can represent a variety of unimodal distributions very closely. As an illustration, Figure 4a shows how a standard extreme value distribution may be closely approximated by a YJ transformation with $\mu_l = 0.35$, $\sigma_l = 1.00$, and $\lambda_l = 0.53$. Figure 4b shows how a standard skew normal distribution with a high degree of right skew parameter ($\alpha = 2.065$) may be mimicked by a YJ transformation with $\mu_l = 0.605$, $\sigma_l = 0.35$, and λ_l=0.7. Figure 4c illustrates the same for a standard skew-t distribution with the same high degree of right skew $(\alpha = 2.065)$ and six degrees of freedom. The skew-t distribution shows fatter tails than the skew-normal, which again is quite well reproduced by the YJ transformation with $\mu_l = 0.52$ $\sigma_l = 0.67$, and λ_l=0.8. A variety of other distributions can also be closely approximated by the YJ transformation, demonstrating its flexibility.

Now, consider that there are L such non-normal variables/parameters (l=1,2,…, L). Across the different variables/parameters $Z_1, Z_2, ..., Z_L$, the direction and intensity of skew/tail can vary. For future use, define $\mathbf{Z} = \left(Z_1, Z_2, ..., Z_L\right)'$ $(L \times 1 \text{ vector})$, $\mathbf{G} = \left(G_1, G_2, ..., G_L\right)'$ $(L \times 1 \text{ vector})$, $\mathbf{Z} = \mathbf{t}_{\boldsymbol{\lambda}}^{\mathbf{-1}}(\mathbf{G}) = \left[t_{\lambda_1}^{-1}(G_1), t_{\lambda_2}^{-1}(G_2), ..., t_{\lambda_L}^{-1}(G_L)\right]$ $(L \times 1 \text{ vector})$, and so $\mathbf{G} = \mathbf{t}_{\boldsymbol{\lambda}}(\mathbf{Z})$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_L)$.

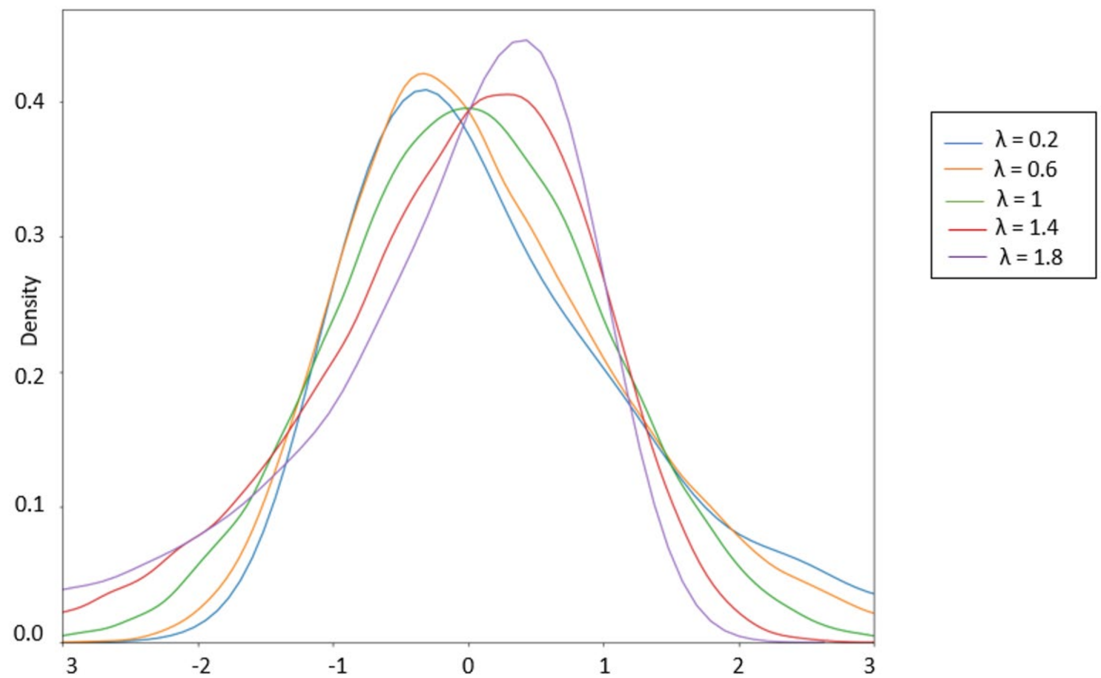

Figure 3. Density of transformed variable for different lambda values

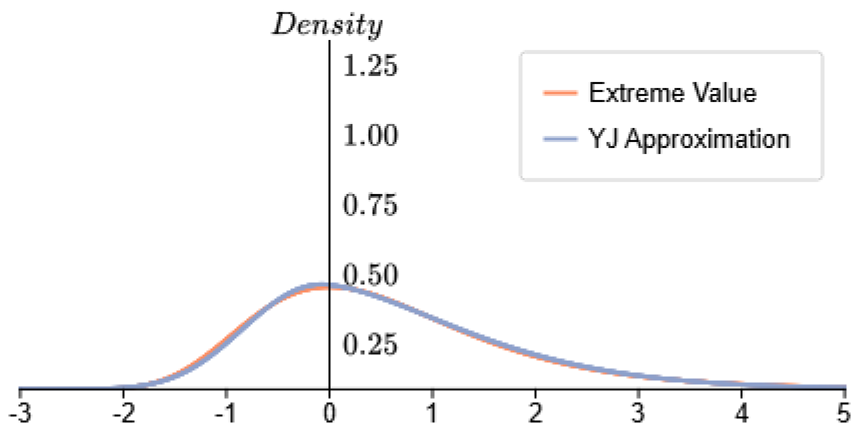

Figure 4a: YJ approximation of standard extreme value distribution

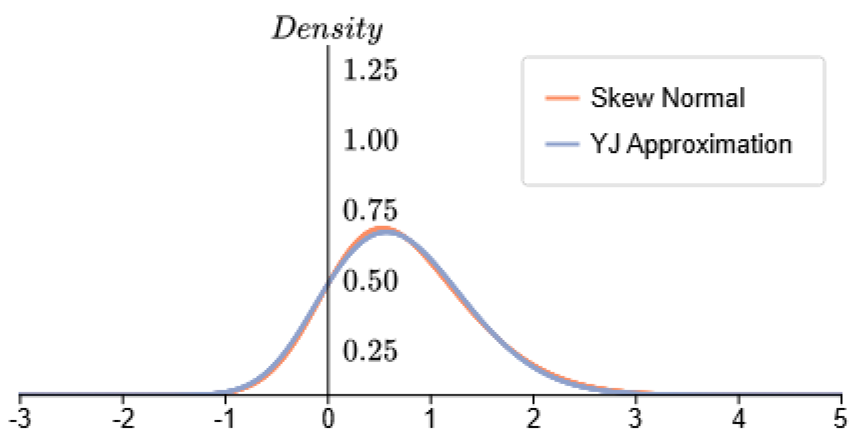

Figure 4b: YJ approximation of standard skew-normal distribution (with slant $\alpha = 2.065$)

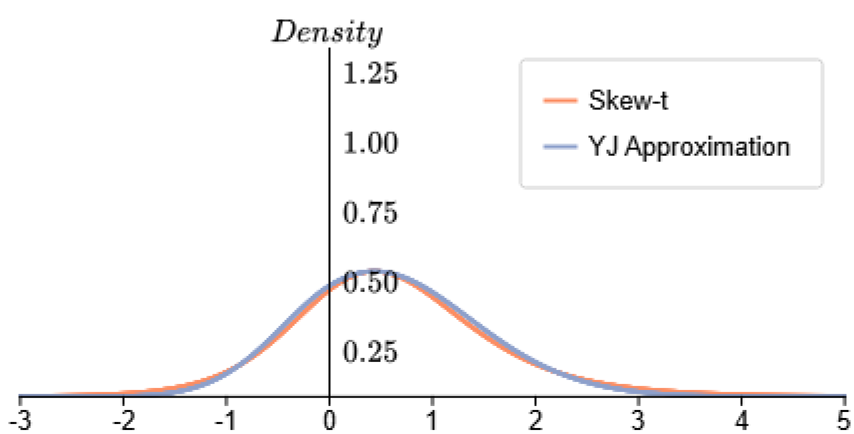

Figure 4c: YJ approximation of standard skew-t distribution (with $\alpha = 2.065$ and 6 dfs)

Figure 4. Ability of YJ transformation to mimic alternative skew/fat-tailed distributions

In consumer behavior studies, it is often also necessary to restrict the distribution of parameters to be consistent with behavioral notions. For instance, in the context of mode choice models, the response coefficient of the travel cost parameter should be strictly negative. A common distributional assumption to ensure such a range restriction is to use a log-normal distribution for the response coefficient. For this, consider the following transformation of bounded random variable/parameter H_v to a normal random variable/parameter S_v with a mean parameter of $\tilde{\mu}_v$ and variance of $\tilde{\sigma}_v^2$) on the real line, such that H_v is restricted to be negative:

$$S_v \sim N(\tilde{\mu}_v, \tilde{\sigma}_v^2) = \tilde{t}(H_v) = \ln(-H_v) \tag{3}$$

The implied reverse transformation is then:

$$H_v = \tilde{t}^{-1}(S_v) = -\exp(S_v) \tag{4}$$

Consider that there are V such log-normal variables. Let $E = L + V$, and let $\mathbf{H} = \left(H_1, H_2, ..., H_V\right)'$ ($V \times 1$ vector), $\mathbf{S} = \left(S_1, S_2, ..., S_V\right)'$ ($V \times 1$ vector), and $\mathbf{H} = \tilde{\mathbf{t}}^{\mathbf{-1}}(\mathbf{S}) = \left[\tilde{t}^{-1}(S_1), \tilde{t}^{-1}(S_2), ..., \tilde{t}^{-1}(S_L)\right]$ ($V \times 1$ vector). Thus, $\mathbf{S} = \tilde{\mathbf{t}}(\mathbf{H})$. Define $\mathbf{K} = (\mathbf{Z}', \mathbf{H}')'$ ($E \times 1$ vector), $\mathbf{R} = (\mathbf{G}', \mathbf{S}')'$ ($E \times 1$ vector), and $\mathbf{K} = \mathbf{c}_{\boldsymbol{\lambda}}^{\mathbf{-1}}(\mathbf{R}) = \left[\left(\mathbf{t}_{\boldsymbol{\lambda}}^{\mathbf{-1}}(\mathbf{G})\right)', \left(\tilde{\mathbf{t}}^{\mathbf{-1}}(\mathbf{S})\right)'\right]'$ ($E \times 1$ vector), so that $\mathbf{R} = \mathbf{c}_{\boldsymbol{\lambda}}(\mathbf{K})$ ($E \times 1$ vector). Now consider a multivariate normal distribution for the transformed **R** variables/parameter vector. We can then write the multivariate cumulative distribution function for the non-normal random vector **K** in terms of the multivariate normal distribution as follows:

$$\begin{aligned}
F_{\mathbf{K}}(\boldsymbol{k}) &= \text{Prob}(\mathbf{K} < \boldsymbol{k}) = \text{Prob}\left[\mathbf{c}_{\boldsymbol{\lambda}}^{\mathbf{-1}}(\mathbf{R}) < \boldsymbol{k}\right],\ \boldsymbol{k} = z_1, z_2, ..., z_L, h_1, h_2, ..., h_V \,\right] \\
&= \text{Prob}\left[\mathbf{R} < \mathbf{c}_{\boldsymbol{\lambda}}(\boldsymbol{k})\right] \\
&= \text{Prob}\left[G_1 < t_{\lambda_1}(z_1), G_2 < t_{\lambda_2}(z_2), ..., G_L < t_{\lambda_L}(z_L), S_1 < \tilde{t}(h_1), S_2 < \tilde{t}(h_2), ..., S_V < \tilde{t}(h_V)\right] \\
&= \text{Prob}(\mathbf{R} < \boldsymbol{r}),\ \boldsymbol{r} = (g_1, g_2, ..., g_L, s_1, s_2, ..., s_V)', \\
&\qquad [g_l = t_{\lambda_l}(z_l),\ l = 1, 2, ..., L,\ s_v = \tilde{t}(h_v),\ v = 1, 2, ..., V\,] \\
&= \tilde{\Phi}_R\left[\boldsymbol{r}; \boldsymbol{\theta}, \boldsymbol{\Omega}\right],\ \boldsymbol{\theta} = (\mu_1, \mu_2, ..., \mu_L, \tilde{\mu}_1, \tilde{\mu}_2, ..., \tilde{\mu}_V)'
\end{aligned} \tag{5}$$

$\tilde{\Phi}_R\left[.; \boldsymbol{\theta}, \boldsymbol{\Omega}\right]$ in the equation above is the multivariate normal distribution function of dimension E with mean **θ** and covariance matrix **Ω**. The implicit copula arises because the mean vector **θ** and the diagonal vector of the **Ω** (corresponding to the vector $(\sigma_1, \sigma_2, ..., \sigma_L, \tilde{\sigma}_1, \tilde{\sigma}_2, ..., \tilde{\sigma}_V)'$) are simply the parameters of the transformed normal variables, so that the only additional parameters to be estimated in the resulting Gaussian copula are the off-diagonal elements of **Ω**. From Equation (5), we also get the following:

$$f_{\mathbf{K}}(\boldsymbol{k}) = \frac{dF_{\mathbf{K}}(\boldsymbol{k})}{d\boldsymbol{k}} = \frac{d\tilde{\Phi}_{\mathbf{R}}\left[\boldsymbol{r}; \boldsymbol{\theta}, \boldsymbol{\Omega}\right]}{d\boldsymbol{r}}\left|\frac{d\boldsymbol{r}}{d\boldsymbol{k}'}\right| = \tilde{\phi}_R(\boldsymbol{r}; \boldsymbol{\theta}, \boldsymbol{\Omega})\left|\frac{d\boldsymbol{r}}{d\boldsymbol{k}'}\right|,\ \left|\frac{d\boldsymbol{r}}{d\boldsymbol{k}'}\right| = \prod_{l=1}^{L}\left(\left|z_l\right| + 1\right)^{\text{sgn}(z_l)(\lambda_l - 1)} \prod_{v=1}^{V}\left(\frac{1}{h_v}\right) \tag{6}$$

where $\tilde{\phi}_R(\boldsymbol{r};\boldsymbol{\theta},\boldsymbol{\Omega})$ represents the multivariate normal density function and $\text{sgn}(z_l)$ takes the value of 1 if z_l is positive, the value of -1 if z_l is negative, and the value of 0 if z_l is zero.

2.2 Model Structure

Consider a multinomial choice context, with the index *q* for the individual, $(q = 1, 2, ..., Q)$ and index *i* for the alternative $(i = 1, 2, ..., I)$. Consider the random-coefficients formulation in which the utility that an individual *q* associates with alternative *i* is written as:

$$\tilde{U}_{qi} = \boldsymbol{\beta}_q' \boldsymbol{x}_{qi} + \boldsymbol{\gamma}' \tilde{\boldsymbol{x}}_{qi} + \tilde{\varepsilon}_{qi}, \tag{7}$$

where $\boldsymbol{x}_{qi}$ is a $(E \times 1)$-column vector of exogenous attributes (without including constants), $\tilde{\boldsymbol{x}}_{qi}$ is another $(D \times 1)$-column vector of exogenous attributes (including dummy variables for constants, except in one of the *I* alternative utilities, say the first alternative), $\boldsymbol{\beta}_q$ is an individual-specific $(E \times 1)$-column vector of coefficients that varies across individuals based on unobserved individual attributes and with each element allowed to be non-normally distributed (here consider that *L* elements are YJ transformed non-normally distributed, and *V* elements are log-normally distributed, where *E=L+V*)[2]. $\boldsymbol{\gamma}$ is another $(D \times 1)$-column vector of coefficients that do not vary across individuals, i.e., the sensitivities to the $\tilde{\boldsymbol{x}}_{qi}$ parameters are assumed to be fixed across individuals[3]. The correspondence of our notations with the previous section should now be clear, with $\boldsymbol{\beta}_q$ taking the place of **K**. In essence, the elements of $\boldsymbol{\beta}_q$ may be partitioned into two vectors: $\boldsymbol{\beta}_q = (\mathbf{Z}_q', \mathbf{H}_q)' = \left[\left(\mathbf{t}_\lambda^{-1}(\mathbf{G}_q) \right)', \left(\tilde{\mathbf{t}}^{-1}(\mathbf{S}_q) \right)' \right]' = \boldsymbol{c}_\lambda^{-1}(\mathbf{R}_q)$. $\mathbf{Z}_q$ corresponds to the elements that are YJ-transformed with parameter vector $\boldsymbol{\lambda}$ based on the underlying multivariate normally distributed vector $\mathbf{G}_q$, and $\mathbf{H}_q$ corresponds to the elements that are exponentially transformed (for the log-normal parameters) from an underlying multivariate normally distributed vector $\mathbf{S}_q$. Defining $\mathbf{R}_q = (\mathbf{G}_q', \mathbf{S}_q')'$ as a multivariate vector of dimension $(E \times 1)$, and writing the transformation in compact form as in the earlier section yields the expression $\boldsymbol{\beta}_q = \boldsymbol{c}_\lambda^{-1}(\mathbf{R}_q)$ and $f_{\boldsymbol{K}_q}(\boldsymbol{k}_q) = \tilde{\phi}_R(\boldsymbol{r}_q;\boldsymbol{\theta},\boldsymbol{\Omega}) \left| \dfrac{\boldsymbol{dr}_q}{\boldsymbol{dk}_q'} \right|$, where $\boldsymbol{k}_q$ is a particular realization of $\boldsymbol{\beta}_q$, and

[2] As discussed earlier, the case of normal distribution is subsumed within the non-normal distribution when λ=1. This obviates the need for pre-specifying certain parameters to be normal or non-normal *a priori.*

[3] If all parameters are assumed to be random, then the $\tilde{\boldsymbol{x}}_{qi}$ vector would only (and at least) contain the alternate-specific constants and the $\boldsymbol{\gamma}$ vector would contain the alternate specific parameters. This is because the randomness in the constants is already absorbed in the kernel error terms and cannot be separately identified.

$\boldsymbol{k}_q = \left[\left(\boldsymbol{t}_\lambda^{-1}(\boldsymbol{g}_q) \right)', \left(\tilde{\boldsymbol{t}}^{-1}(\boldsymbol{s}_q) \right)' \right]' = \boldsymbol{c}_\lambda^{-1}(\boldsymbol{r}_q)$, $\boldsymbol{r}_q = (\boldsymbol{g}_q, \boldsymbol{s}_q)'$. As in the earlier section, $\boldsymbol{\theta}$ represents the mean vector of $\mathbf{R}_q$, and $\boldsymbol{\Omega}$ represents the covariance matrix of $\mathbf{R}_q$. That is, $\boldsymbol{r}_q$ represents an E-variate realization from the multivariate normal distribution of $\mathbf{R}_q$ with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Omega}$.

The (I×1)-vector of kernel error terms, $\tilde{\boldsymbol{\varepsilon}}_q = (\tilde{\varepsilon}_{q1}, \tilde{\varepsilon}_{q2}, \tilde{\varepsilon}_{q3}, \ldots, \tilde{\varepsilon}_{qI})'$, is assumed to have a general covariance structure subject to identifiability considerations such that $\tilde{\boldsymbol{\varepsilon}}_q \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Theta})$. ($\tilde{\boldsymbol{\varepsilon}}_q$ is assumed independent of $\boldsymbol{\beta}_q$). Since only utility differences matter in discrete choice models, appropriate identification conditions are required to be imposed. A common approach is to take the differences of the error terms with respect to the first alternative; therefore, let $\varepsilon_{qi1} = (\tilde{\varepsilon}_{qi} - \tilde{\varepsilon}_{q1})$, and let $\boldsymbol{\varepsilon}_{q1} = (\varepsilon_{q21}, \varepsilon_{q31}, \ldots, \varepsilon_{qI1})$. Next, an appropriate scale normalization is required for identification. To do so, scale the top left diagonal element of this error-differenced covariance matrix to 1. Thus, there are $[(I-1) \times (I/2)] - 1$ free covariance terms in the $(I-1) \times (I-1)$ matrix $\tilde{\boldsymbol{\Theta}}_1$. $\boldsymbol{\Theta}$ is constructed from $\tilde{\boldsymbol{\Theta}}_1$ by adding a top row of zeros and a first column of zeros.

2.3 Model Estimation

With the results and identification considerations from above, we may write Equation (7), conditional on $\boldsymbol{\beta}_q = \boldsymbol{k}_q$ as follows:

$$\tilde{U}_{qi} \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \boldsymbol{k}_q' \boldsymbol{x}_{qi} + \boldsymbol{\gamma}' \tilde{\boldsymbol{x}}_{qi} + \tilde{\boldsymbol{\varepsilon}}_{qi}, \; \boldsymbol{k}_q = (z_{q1}, z_{q2}, \ldots z_{qL}, h_{q1}, h_{q2}, \ldots, h_{qV}), \tag{8}$$

We now set out some additional notation. Define $\tilde{\boldsymbol{U}}_q = (\tilde{U}_{q1}, \tilde{U}_{q2}, \ldots, \tilde{U}_{qI})'$ ($I \times 1$ vector), $\boldsymbol{x}_q = (\boldsymbol{x}_{q1}, \boldsymbol{x}_{q2}, \ldots, \boldsymbol{x}_{qI})'$ ($I \times E$ matrix), $\tilde{\boldsymbol{x}}_q = (\tilde{\boldsymbol{x}}_{q1}, \tilde{\boldsymbol{x}}_{q2}, \ldots, \tilde{\boldsymbol{x}}_{qI})'$ ($I \times D$ matrix). Then, we can write Equation (8) in matrix form as:

$$\tilde{\boldsymbol{U}}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \boldsymbol{x}_q \boldsymbol{k}_q + \tilde{\boldsymbol{x}}_q \boldsymbol{\gamma} + \tilde{\boldsymbol{\varepsilon}}_q \tag{9}$$

It is clear that $\tilde{\boldsymbol{U}}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)$ is multivariate normally distributed:

$\tilde{\boldsymbol{U}}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) \sim MVN_I(\mathbf{V}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q), \boldsymbol{\Theta})$, where $\mathbf{V}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \left[(\boldsymbol{x}_q \boldsymbol{k}_q + \tilde{\boldsymbol{x}}_q \boldsymbol{\gamma}) \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) \right]$.

Let the individual q choose alternative m_q. Define another matrix $\boldsymbol{M}_q$ as an identity matrix of size $(I-1)$ with an extra column of '-1' values added at the m_q^{th} column. Let $\mathbf{B}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \boldsymbol{M}_q \left[\mathbf{V}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) \right]$ and $\tilde{\boldsymbol{\Theta}} = \boldsymbol{M}_q \boldsymbol{\Theta} \boldsymbol{M}_q'$. The parameter vector to be estimated is $\boldsymbol{\delta} = (\boldsymbol{\gamma}', \boldsymbol{\theta}', \boldsymbol{\lambda}', \text{Vech}(\boldsymbol{\Omega}), \text{Vech}(\boldsymbol{\Theta}))'$, where $\text{Vech}(\boldsymbol{\Omega})$ is a column vector obtained by vertically stacking the upper triangle elements of the matrix $\boldsymbol{\Omega}$, and $\text{Vech}(\boldsymbol{\Theta})$ is another column vector obtained by vertically stacking the estimable upper triangular elements of the matrix $\boldsymbol{\Theta}$. The likelihood contribution of individual q conditional on $\boldsymbol{\beta}_q = \boldsymbol{k}_q$ is as below:

$$L_q(\boldsymbol{\delta}) \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \Phi_{(I-1)}\left(\left[-\mathbf{B}_q^* \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)\right], \boldsymbol{\Theta}^*\right), \tag{10}$$

where $\mathbf{B}_q^* \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \boldsymbol{\omega}_\Theta^{-1}\mathbf{B}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)$, $\boldsymbol{\Theta}^* = \boldsymbol{\omega}_{\boldsymbol{\Theta}}^{-1}\boldsymbol{\Theta}\,\boldsymbol{\omega}_{\boldsymbol{\Theta}}^{-1}$, and $\boldsymbol{\omega}_{\boldsymbol{\Theta}}$ is the diagonal matrix of standard deviations of $\boldsymbol{\Theta}$. Finally, the unconditional likelihood contribution of individual q is:

$$\begin{aligned} L_q(\boldsymbol{\delta}) &= \int_{\boldsymbol{k}_q=-\infty}^{\boldsymbol{k}_q=+\infty} \left[\Phi_{(I-1)}\left(\left[-\mathbf{B}_q^* \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)\right], \boldsymbol{\Theta}^*\right) \right] f_{\boldsymbol{\beta}_q}(\boldsymbol{k}_q) \boldsymbol{dk}_q \\ &= \int_{k_q=-\infty}^{k_q=+\infty} \left[\Phi_{(I-1)}\left(\left[-\mathbf{B}_q^* \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)\right], \boldsymbol{\Theta}^*\right) \right] \tilde{\phi}_{R_q}(\boldsymbol{r}_q; \boldsymbol{\theta}, \boldsymbol{\Omega}) \left| \frac{\boldsymbol{dr}_q}{\boldsymbol{dk}_q'} \right| \boldsymbol{dk}_q. \end{aligned} \tag{11}$$

where $\tilde{\phi}_R(\boldsymbol{r}_q; \boldsymbol{\theta}, \boldsymbol{\Omega})$ represents the multivariate normal density function (the second equation above follows from Equation 6). The integration is evaluated by using a simulation technique wherein at each draw $\boldsymbol{r}_q$ from the multivariate normal distribution, the corresponding non-normal vector realization $\boldsymbol{k}_q$ in $\mathbf{V}_q \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q) = \left[(\boldsymbol{x}_q\boldsymbol{k}_q + \tilde{\boldsymbol{x}}_q\boldsymbol{\gamma}) \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)\right]$ (which is itself embedded in $\mathbf{B}_q^* \mid (\boldsymbol{\beta}_q = \boldsymbol{k}_q)$) is computed as $\boldsymbol{k}_q = \boldsymbol{c}_\lambda^{-1}(\boldsymbol{r}_q)$. A total of 500 Halton draws (see Bhat, 2001; Bhat, 2003) of $\boldsymbol{r}_q$ is used for each individual to evaluate the integration. The positive definiteness of the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Theta}$ are ensured by taking the Cholesky decomposition of each of the covariance matrices and estimating the Cholesky elements in the optimization method. Furthermore, since a closed form expression does not exist for the multivariate normal integral, we used the One-variate Univariate Screening technique proposed by Bhat (2018) for evaluating the $\Phi_{(I-1)}$ integral.

3 MODEL APPLICATION

Using our proposed approach, we analyze the intra-city commute mode choice decisions of workers in Bengaluru, a major metropolitan city in the southern region of India. As a major center of economic growth and the capital of the state of Karnataka, the city of Bengaluru has undergone rapid urbanization, especially in the last two to three decades, and continues to witness significant urban sprawl. In particular, the rapid growth of the IT hub and influx of workers in the city has led to excessive network congestion, environmental-related issues, and tremendous pressure on the public transit system (see for example, Harsha et al., 2020). This pattern of urbanization is common in major metropolitan centers of developing economies, typically characterized by overcrowded systems and a supply-demand imbalance of resources; therefore, the present effort provides an important case study for commuters' mode choice in a typical urban region of a developing country.

The data for the analysis comes from a mode choice survey conducted in Bengaluru. The survey collected information on individuals' modal preferences for different trip purposes, with a specific focus on collecting information on the revealed mode choice decisions for their routinely traveled destinations. The objective of the current empirical analysis is to explore the advantages of non-normal mixing distributions on the coefficients of level-of-service variables and the consequent effect on the resulting valuation of travel time savings (VTTS) measures. In the next few sections, we discuss the survey sample, the model results, and important policy implications.

3.1 Sample Data

The data for this analysis is drawn from a survey administered between February and April of 2022 to collect information on the routine travel pattern of the residents of the Bruhat Bengaluru Mahanagara Palike (BBMP) area of Bengaluru. The survey asked respondents about their routine travel destination, along with the purpose and most frequently used mode of their travel to the routine destination. For this study, we used the data of only those who reported commuting as the purpose of travel to their routine destination. Such commuters were asked about their most frequently used primary mode for their commute trips. If a commuter's trip consisted of a combination of two or more modes, the mode with the longest leg (the primary mode) was asked to be reported. Six modal alternatives were considered in our analysis: auto-rickshaw, bus, metro, walk, two-wheelers, and private car[4]. The final sample comprises modal information on commute trips for 914 individuals. Note that, for each individual, we only consider her/his home-to-work leg of the commute trip, i.e., each of the 914 observations corresponds to the home-to-work commute trip of the respective individual. The modal sample shares are reported in Table 1 (see the top row of the table). As evident, the majority of the commute trips are undertaken using two-wheelers, with more than 50% sample share. This is followed by bus and metro modes, with 22.4% and 14.7% shares, respectively, making up more than 37% share for the public transit modes. The mode shares for private cars and auto-rickshaws were found to be 6.4% and 3.3%, respectively. Walking has the least share of all the modes at 2.2%. Note that the modal shares provided in Table 1 are for the entire sample without considering whether each of the six modes was feasible/available to an individual. However, in our analysis, we account for the availability/feasibility of each alternative for each individual. Walking was considered a feasible alternative only when the commute distance was less than 5 kilometers[5]. As a result, walking was a feasible alternative for only 257 individuals. But within the context of these walkable commute trips, the walk mode share is a sizeable 7.7%. Public transit modes (the bus and the metro modes) were considered feasible/available (as the primary mode of travel), only if the total first and last mile access distance between an individual's home and workplace was less than 5 kilometers. Thus, the bus and the metro modes were feasible alternatives for 781 and 253 individuals in the sample respectively; within these individuals, the respective modal shares for the bus and metro modes were 26.2% and 53.1%. Similarly, personal modes (i.e., cars and two-wheelers) were

[4] The final sample collected did not have sufficient users for the ride-hailing mode - only five users reported to have used ride-hailing for commute as their primary mode, and therefore these observations were screened out.

[5] In our sample, the maximum distance corresponding to the chosen walk mode is observed to be about four kilometers. But, owing to the relatively moderate sample size used in this study that may have missed some walkers with a longer commute distance, and the not-too-uncommonly observed one hour of walk commute time prevalent in Indian cities, we decided to be more inclusive in considering the walk mode within an individual's choice set by using a slightly more generous threshold of 5 kilometers (which, at an average walk speed of 5 km per hour, would take an hour).

considered feasible only if the household owned these personal modes. Among the 289 households that had a car available, the car mode share rises to 20.4%, and, among the 755 households who had a two-wheeler available, the two-wheeler mode share rises to 61.7%. Due to the widespread availability of the auto-rickshaw mode throughout the city of Bengaluru, the auto-rickshaw mode was a feasible/available mode in the choice set for all the individuals in the sample (and, for this reason, the auto-rickshaw mode was used as the base alternative in our estimations, even though it has the second lowest share in the market).

In addition to the modal sample shares, the survey also collected information on individual and household-specific socio-demographic variables, such as gender, age, education status, employment type, household income, vehicle ownership, and commute start time. The sample descriptive statistics for these variables are reported in Table 1. The sample has a significantly higher share of men than women, which is not surprising since a similar trend is also reflected in labor force participation rates in India (National Statistical Survey Office, 2021). The sample has a fairly good distribution across age categories. In the context of educational qualification, there is a high proportion of individuals (36.5%) with less than senior school education, despite the sample representing only employed individuals. However, as per the National Statistical Survey Office (2012) data, a large proportion (around 50%) of the employed population in India has studied only until senior school or less. In the context of household attributes, the sample shows a little over 43% of households in the middle-income category (between 20,000 and 100,000 rupees monthly income). Further, more than 83% of households own a two-wheeler, while only 31.6% of households own a car. Finally, in terms of commute characteristics, close to a quarter of respondents have a metro-pass, while over two-thirds start their commute from home (to their regular workplace) during peak hours.

The bottom panel of Table 1 presents information on the travel attributes (travel time in minutes and travel cost in Rupees) for each of the six modes. The average travel time (and the sample standard deviation for time) and the average travel cost (and the sample standard deviation for cost) reported for each of the modes are computed by taking the average (and the sample standard deviation) across all the individuals for whom the respective modes are available/feasible, based on our earlier discussion. For a given commute trip, the travel times (which include the in-vehicle as well as the out-of-vehicle travel times) for different modes were fetched using the Google API (Application Programming Interface), based on the origin (home address), destination (workplace address) and the reported start time of the trip. Expectedly, the average travel times for public transit modes are the highest with over 47 minutes for the bus mode and close to 39 minutes for the metro mode, while that of the private modes and the intermediate public transport (IPT) auto-rickshaw modes are on the lower end (below 28 minutes). Similar to the collection of the travel time data, the travel cost information for metro and bus was fetched using the Google API. For the auto-rickshaw mode, the existing governmental fare structure was used to calculate the travel cost for a given home-to-work trip distance. Travel costs for the private modes (private cars and two-wheelers) were calculated based on the trip distance, prevalent gas price (90 rupees; that is, ₹90 per liter of gas) and by assuming a reasonable vehicle mileage (13 kilometers per liter for private cars and 40 kilometers per liter for two-wheelers). As can be observed from the bottom panel of Table 1, the average travel cost for the bus mode is the least among the motorized modes of transportation, followed by two-wheelers and metro. Quite expectedly, the average travel cost by private cars is at the higher end with a value of over ₹70. The auto-rickshaw mode is reported to have the highest average travel cost with a value of just over ₹125 (this is because the fare structure for such a mode includes the cost of being privately chauffeured in addition to the vehicle operating cost). The reported travel-related attributes' values for commute trips are very reasonable in the context of the city of Bengaluru (see Nayka and Sridhar, 2019, and Sridhar and Nayka, 2022).

Table 1. Details of the sample shares and socio-demographic variables in the data

Sample shares of the alternatives (Sample size (N) = 914)						
	Auto-rickshaw	**Metro**	**Bus**	**Walk**	**Two-wheeler**	**Private car**
Sample shares (%)	3.3%	14.7%	22.4%	2.2%	51.0%	6.4%

Descriptive statistics of the exogenous variables	
Variables	**Sample share**
Individual specific attributes	
Gender	
Male	71.4%
Female	28.6%
Age	
Age 19 - 25 years	19.8%
Age 26 - 35 years	36.9%
Age 36 - 45 years	22.7%
Age greater than 45 years	20.6%
Education	
Less than 12th grade	36.5%
Diploma	13.6%
Undergraduate degree	35.5%
Graduate and above	14.4%
Employment status	
Employed in government sector	8.1%
Employed in private sector	54.5%
Self-employed/Business	37.4%
Household characteristics	
Monthly income	
Less than ₹20,000	27.2%
Between ₹20,000 and ₹100,000	43.4%
More than ₹100,000	29.4%
Household two-wheeler count	
Zero two-wheeler	17.4%
One two-wheeler	51.5%
Two or more two-wheelers	31.1%
Household car ownership	
Zero car	68.4%
One car	27.0%
Two or more cars	4.6%
Commute travel characteristics	
Metro-pass availability	
Available	24.2%
Not available	75.8%
Commute start time from home	
Peak-hours (8:00-11:00, 16:00-22:00)	67.1%
Off-peak hours (Other times)	32.9%

Travel attributes							
		Auto-rickshaw	**Metro**	**Bus**	**Walk**	**Two-wheeler**	**Private car**
Travel Time in minutes	Mean	23.73	38.95	47.51	36.50	25.36	27.30
	Standard deviation	14.50	19.78	24.88	11.54	14.60	15.30
Travel Cost in ₹	Mean	125.70	31.00	19.86	NA	21.51	71.38
	Standard deviation	88.00	10.85	6.29	NA	15.53	52.20

We are unable to examine the representativeness of the sample compared to the population of Bengaluru workers, because almost all census data in India contain information for the population as a whole, and not subpopulation groups such as workers. However, the sample may be compared to the gender distribution of the workforce in the city of Bengaluru for a check on representativeness, for which the census data is available. Our sample, with 28.6% women, is fairly representative of the 25% women in the Bengaluru workforce (Census of India, 2011). Further, the average commute distance in our sample is 10.86 kms, which is very close to the average commute distance for a typical urban dweller in India (Statista, 2019).

3.2 Model Results

In our estimation process, we explored several functional forms for the exogenous variables including a linear form, dummy variable categorization, count variable forms, as well as several interactions of explanatory variables. The dummy variable categorization provided the best fit for most variables except for the travel time and travel cost variables (which came out to be best in continuous form) and two count variables (number of two-wheelers/private cars). The final model specification was obtained after a systematic process of testing alternative combinations of explanatory variables based on statistical fit and parsimony considerations. Importantly, a log-normal distribution for the travel cost coefficient is assumed in our study to restrict the response coefficient to be negative across all individuals (and thus, avoiding the breakdown of VTTS computation due to the singularity problem), while an unrestricted non-normal distribution (as proposed in this paper) is assumed for the travel time response coefficient. The model results are presented in Table 2. These results provide the effect of exogenous variables on the utility of the different modes. The auto-rickshaw mode of transportation is considered the base alternative in our multinomial choice context for the introduction of the effects of alternative-invariant exogenous variables. A '--' in the table for a specific cell indicates that the corresponding row variable has no statistically significant impact on the corresponding column mode.

Individual characteristics

Our results from Table 2 indicate that there are significant gender differences in modal preference. Men generally have a higher inclination toward public transportation (bus and metro), as well as two-wheelers and cars, than women. These results are consistent with previous findings (see, for example, Mahadevia and Advani, 2016, in the context of an Indian city). Public transportation modes are often overcrowded in metropolitan cities in India, posing several challenges related to safety and security issues for women. Moreover, in a society with systematic asymmetries in gender roles, men still are the "primary" users of household-owned private vehicles, which is reflected in the lower usage of private cars among working women (Ram and Dhawan, 2018).

As found in several earlier studies, age is also a key determinant of modal preferences. Older individuals (aged above 45 years) are disinclined toward the use of the metro alternative, presumably because of the high crowding levels that this mass transit system experiences (Panambur and Sushma, 2019). Our results also suggest that middle-aged individuals (age 36-45 years) have the highest preference for the use of bus and walk modes compared to older as well as younger-aged workers. The latter result, in particular, may be attributed to the younger generation's desire for a fast-paced lifestyle that makes slower modes of transportation, such as buses and walking, less attractive. Middle-aged individuals also appear to associate a higher utility to the two-wheeler mode of travel relative to their younger and, especially older, peers, while middle-aged and older workers have a higher propensity for the use of private cars compared to their younger peers in the 19-25 year age group. This is consistent with the notion that young adults at the prime of their physical shape are less troubled by travel discomfort and inconvenience, and may also be more environmentally conscious, and so are less drawn toward the use of private cars (Masoumi, 2019). On the other hand, the results indicate that individuals with higher formal education levels (undergraduate degree or higher) are drawn more toward the use of private cars and less inclined (for those with a graduate degree or higher) toward the walk mode, presumably because of higher time pressure experienced by these individuals compared to individuals with a lower formal level of education (Batur et al., 2019). Finally, within the group of individual characteristics, our results suggest that workers who are self-employed or engaged in business have a lower preference for public transportation (metro and bus) and walk mode compared to individuals employed in government or private jobs.

Household demographics

Not surprisingly, workers from households with higher levels of monthly income (income greater than ₹20,000) have a higher preference for private modes of transportation (two-wheelers and private cars) compared to those from lower-income households, while those in the highest income category (monthly income of more than ₹100,000) have the lowest inclination to use the public transportation (metro) mode. Within the context of the two-wheeler mode, there is an increased preference toward this mode for middle-income households (₹20,000- ₹100,000) compared to other income groups, an observation which is common in typical Indian cities (see for example, Shirgaokar, 2014). Also, as one would expect, workers in households with a higher number of two-wheelers reveal a greater preference for two-wheeler use, and those in households with a higher number of private cars are more inclined toward private vehicle use.

Commute-related characteristics

Among commute characteristics, those holding a metro-pass, quite naturally, are predisposed toward the use of the metro system, though we acknowledge that this is more an association than a causal effect in that those who are predisposed toward the use of the metro system are the ones more likely to purchase the metro-pass in the first place. Interestingly, such pass holders are also more likely to use two-wheelers for their work commute, presumably because of the comparable "convenience in navigation" between these two modes – the metro mode is generally unaffected by road traffic congestion, and the two-wheeler mode is relatively easier to navigate in congested cities like Bengaluru (compared to other motorized modes). Thus, in scenarios when the metro mode may be overcrowded or infrequent in service, these individuals (who hold a metro-pass and also own a two-wheeler) may tend to use their two-wheelers that provide a comparable level of travel time reliability to that of the metro mode. Finally, peak-period commuters have a generic predisposition toward the use of two-wheelers and private cars, if such vehicles are available in their household, perhaps as a way of retaining some sense of control and comfort during the slow traffic movement periods of the day (Ramakrishnan et al., 2020).

Table 2. Mode choice model results

Variables	Auto-rickshaw	Metro		Bus		Walk		Two-wheeler		Private Car	
	(Base)	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Individual characteristics											
Gender (base: Female)											
Male		0.512	1.88	0.201	1.94	--		0.751	3.18	0.628	2.72
Age (base: the omitted categories in each mode)											
Age 19-25		--		--		--		0.194	2.29	--	
Age 26-35		--		--		--		0.299	3.01	--	
Age 36-45				0.329	1.99	0.418	1.81	0.641	3.04	0.306	2.13
Age over 45 years		-0.383	-1.89	--		--		--		0.102	1.99
Education (base: Less than Undergraduate degree)											
Undergraduate degree		--		--		--		--		0.217	1.93
Graduate or more		--		--		-0.223	-1.78	--		0.217	1.93
Employment type (base: Govt. or private)											
Self-employed/Business		-0.580	-3.01	-0.448	-3.32	-0.504	-2.09	--		--	
Household characteristics											
Monthly Income (base: Less than ₹ 20,000)											
₹ 20,000 to ₹ 100,000		--		--		--		0.188	1.89	0.267	1.90
More than ₹ 100,000		-0.431	-1.90	--		--		0.122	1.93	0.267	1.90
Number of two-wheelers		--		--		--		0.129	2.13	--	
Number of four-wheelers		--		--		--		--		0.394	2.61
Commute travel characteristics											
Metro-pass availability (base: Not available)											
Metro-pass available		1.104	4.44	--		--		0.273	2.25	--	
Commute start time (base: Off-peak)											
Peak-hours		--		--		--		0.089	1.62	0.282	1.95
Travel attributes (non-normally distributed)											
Travel time (TT in 100 minutes): [unrestricted non-normal distribution]											
Mean estimate											-0.690 (-3.56)
Std dev.											0.305 (2.22)
Transformation parameter λ_{TT}											1.424 (2.56)
Travel cost (TC in ₹ 100): [log-normally distributed]											
Mean estimate											-0.152 (-4.44)
Std dev.											0.215 (1.89)
Copula Correlation - TT and TC coef.											0.075 (0.96)
Constant		0.601	1.719	0.532	2.09	0.396	1.53	-0.171	-0.52	-0.354	-1.24

NOTE: "--" indicates that the corresponding variable is insignificant.

Travel attributes
As expected, the mean of the travel time and travel cost coefficients are negative (see the bottom row panel of Table 2), highlighting the general disutility associated with travel times and cost[6]. Our results also indicate the presence of significant unobserved heterogeneity (randomness) in the coefficients associated with travel time and travel cost. As proposed in this study, we observe significant non-normality in the unrestricted distribution of the travel time coefficient. In particular, our results suggest that the random coefficient associated with the travel time variable is significantly left-skewed, reflecting a fatter left tail in the coefficient distribution. The estimated mean of the corresponding normally-transformed variable for this coefficient is -0.690, the estimated standard deviation is 0.305, and the estimated transformation parameter (λ_{TT}) is 1.424. The mean and standard deviation are both statistically different from zero, while the t-statistic in the table of 2.56 is with respect to a λ_{TT} value of one (indicating that the estimated distribution of the travel time coefficient is different from that of a normal distribution). As mentioned earlier, a log-normal distribution was assumed for the travel cost coefficient to avoid the breakdown of the VTTS computation; the estimated mean and standard deviation for this distribution are -0.152 and 0.215, respectively. The estimated Gaussian copula correlation between the travel time and travel cost coefficients was very low and not statistically significant in the current empirical context (we show the estimated copula correlation in Table 2 just for completeness, though the model estimates and log-likelihood values hardly changed when the correlation was constrained to zero).

The significant difference of the estimated λ_{TT} parameter relative to the value of 1 is already indicative of a non-normality of the travel time parameter, as discussed above. But, to formally compare our model with a model that restricts the time parameter to normal (while maintaining a log-normal distribution for the travel cost parameter), we estimated a separate random MNP (which we call the N-MNP) model (results not shown to conserve on space). The travel time coefficient mean and standard deviation estimates for the N-MNP model turn out to be -0.739 and 0.463, respectively, and the mean and standard deviation estimates for the travel cost coefficient (in the negative log-normal form) are -0.131 and 0.192, respectively.[7] We plot the distribution of the travel time coefficient obtained from our proposed non-normal model as well as that obtained from the N-MNP model in Figure 5. The non-normality in the coefficient distribution of the travel time variable should be evident from this figure, which essentially indicates a lower sensitivity to travel time in the commuter population as estimated by the N-MNP model compared to that estimated by our proposed model. For completeness, a similar comparative plot for the travel cost distribution is provided in Figure 6 (where the log-normal distribution in

[6] The travel time variable in our analysis includes the sum of in-vehicle and out-of-vehicle travel times. During our model building process, we also estimated separate coefficients for in-vehicle and out-of-vehicle travel times; however, they did not turn out to be statistically different. Therefore, in our final model, we specified a single coefficient for the travel time variable.

[7] Even in this N-MNP model, the copula correlation between the travel time and cost coefficients was almost zero in our empirical context, and not statistically significant even at the 65% confidence level; also, as in the F-MNP model, there was essentially no difference in the model estimates or the log-likelihood value at convergence in the N-MNP model when the copula correlation was dropped.

our proposed specification is compared with the log-normal distribution in the N-MNP model). Several key insights may be taken away from these plots. First, the travel time coefficient appears to have a higher spread (more heterogeneity) compared to the travel cost coefficient, suggesting a broader range of sensitivities for travel time. This may be attributed to the growing use of ICT and technology-related platforms during travel, especially for those who are not active in the act of driving/riding. That is, based on one's technology-savviness, there is likely to be a wide heterogeneity in travel time sensitivity, especially in the context of short-term daily travel (a reason why we did not explicitly restrict the distribution of the travel time coefficient to be exclusively negative). Second, for the travel time coefficient distribution, the results from our proposed model suggest that only 1.2% of the population has a positive valuation of commute travel time (represented by the area enclosed by the density function on the right of the zero abscissa), while the results from the N-MNP model suggest a percentage of 5.5% (see Figure 5, where the area enclosed by the density function of the travel time from the normal MNP to the right of the zero abscissa is larger than the area enclosed by the non-normal density function from our proposed model). Overall, our non-normal model bounds the travel time distribution to a more "expected" parameter space compared to the normal MNP, although we do not explicitly bound the travel time coefficient distribution. Third, while it may seem that the travel time coefficient distribution plots from our model and the more restrictive N-MNP model are not very different, this is also an artifact of the Y-axis scaling. With a fine resolution for the Y-axis scale, the difference can visually seem more (or less) profound. That is the reason for the use of the λ_{TT} t-statistic test (to see if this parameter is statistically different from the value of 1, which it is) and the data fit tests in Section 3.3. To support this, we have used the same Y-axis and X-axis scales for Figures 4 and 5, which should make clear how different the travel time distributions are in Figure 5. Fourth, as expected, we achieve a similar shape for the two log-normally distributed travel cost coefficients (although the travel cost distribution from the N-MNP model is marginally shifted to the left compared to that obtained from our proposed model, indicating a higher sensitivity to travel cost in the commuter population as estimated by the N-MNP model relative to our proposed model; see Figure 6). The differences in the travel time and travel cost coefficients obtained from our proposed model and the N-MNP model have implications for the VTTS estimate, as discussed next.

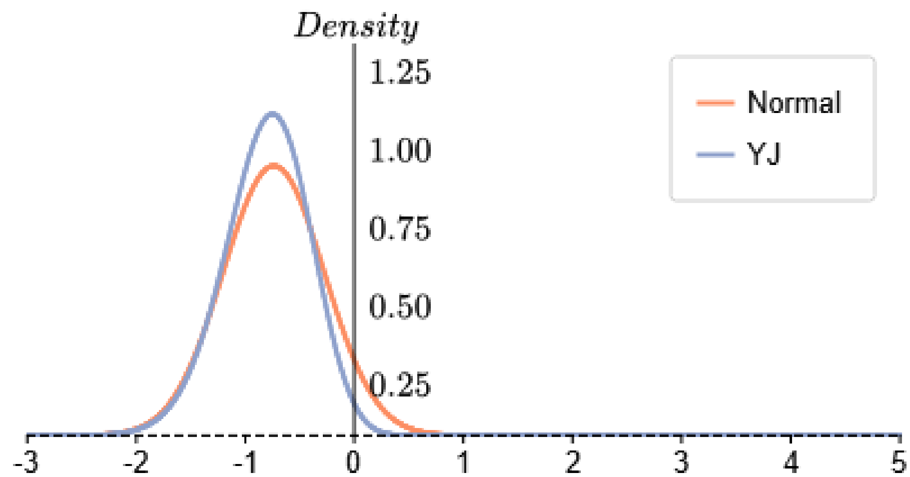

Figure 5. Travel time coefficient distribution plots for normal distribution in N-MNP model and proposed unrestricted non-normal distribution

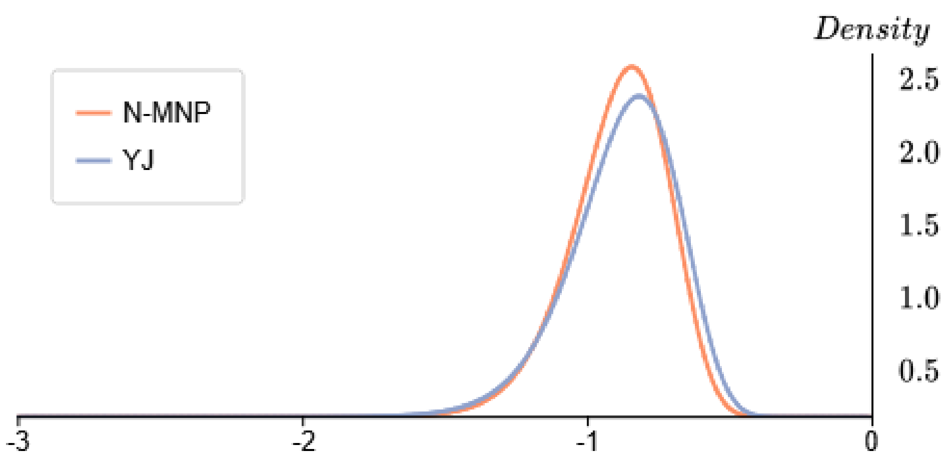

Figure 6. Travel cost coefficient distribution plots for N-MNP model and proposed (YJ) model

<u>*VTTS estimate*</u>
For the VTTS estimate, we compute the median VTTS (rather than the mean) from the travel cost and travel time distributions since it is a better measure of central tendency in cases where there is a possibility of straddling extreme values (for example, when the travel cost coefficient, which appears in the denominator, takes a value very close to zero)[8]. The VTTS median estimate is computed by drawing 10,000 realizations from the bivariate copula distribution of the time and

[8] Moreover, a finite mean does not always exist for complex distributions, especially complex ratio distributions, while the median besides being more robust toward extreme values, always exists, and therefore provides a better measure of central tendency (see Miller, 2015).

cost coefficients, computing the implied VTTS for each bivariate realization by taking the ratio of the time to cost coefficient draws, and then computing the median value across the 10,000 realizations. The median VTTS estimate for our proposed flexible MNP (or F-MNP) model turned out to be ₹55.6/hour, which is a very reasonable estimate for commute trips for an Indian metropolitan city (see Athira et al., 2016). The median VTTS estimate from the N-MNP model which ignores the non-normality in travel time distribution turned out to be ₹51.2/hour, which is about 8% lower (underestimation) than that computed from our proposed model. This VTTS underestimation from the N-MNP model is consistent with the general lower sensitivity to travel time and higher sensitivity to travel cost from the N-MNP model, as discussed in the previous section. In Figure 7, we plot the VTTS distributions (obtained from the random draws) for our proposed F-MNP and N-MNP models. For both the models, the percentage of individuals predicted to have a negative VTTS, of course, reflects the percentages of individuals predicted to have a positive valuation to travel time sensitivities obtained earlier (1.2% and 5.5% respectively).

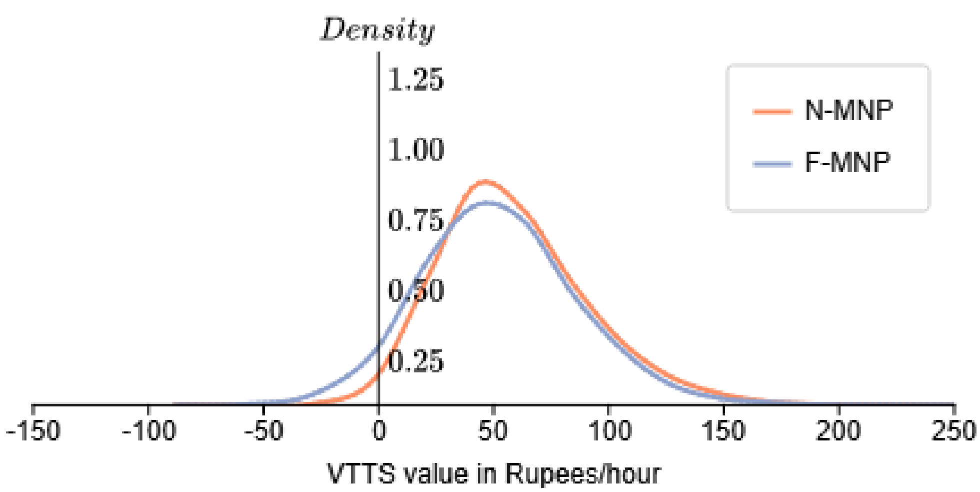

Figure 7. VTTS distributions for N-MNP model and proposed flexible MNP (F-MNP) model

Constants

The alternate specific constants at the bottom of Table 2 do not have any meaningful interpretation and simply provide adjustments to the utility values after accommodating the exogenous variables.

Covariance matrix

As indicated in Section 2, only the covariance matrix of the error differences is estimable (with the scale for one of the error differences normalized to one); in our empirical context, we normalized the first diagonal element of the error-differenced covariance matrix to one, the difference being taken with respect to the auto-rickshaw mode. The differenced error covariance matrix is not interpretable (unless a structural assumption is imposed on the covariance matrix), because multiple (undifferenced) error covariance matrices can originate from the same differenced covariance matrix. In the differenced error covariance matrix, a total of six elements turned out to be significant at an 85% confidence interval – the four non-normalized variance terms in the diagonal and two off-diagonal elements corresponding to the error differenced covariance term between bus and metro, and the error differenced covariance term between metro and two-wheeler (all differences taken with respect to the base auto-rickshaw mode). To conserve on space, and because of interpretation challenges, we do not present the differenced covariance matrix here, but it is available at https://www.caee.utexas.edu/prof/bhat/ABSTRACTS/MNPYJ/OnlineSupp.pdf.

3.3 Data Fit Measures

The flexible multinomial probit model (F-MNP) provides important insights regarding the effects of several sociodemographic, household, and travel-related attributes. However, it is also important to consider the data fit provided by such a model relative to (a) the N-MNP model that ignores the non-normal distribution of the random parameters as well as (b) a non-random MNP (we call this MNP) model that entirely ignores the randomness in the parameters. The three models can be compared using simple pair-wise nested likelihood ratio tests because the MNP model is a nested (restricted) version of the N-MNP model and the N-MNP model, in turn, is a nested (restricted) version of the F-MNP model. We also evaluate the data fit of the three models intuitively and informally at both the disaggregate and aggregate levels. At the disaggregate level, we compute the likelihood-based measures (including the adjusted rho-bar squared value and the Bayesian Information Criterion or BIC) as well as an average (across individuals) probability of correct prediction of the observed choice.[9] At the aggregate level, we compute the predicted shares of each mode from the three models and then compute the weighted average percentage error (WAPE) value (the weighting here is based on the actual observed share of each mode). The results of the data fit evaluations are provided in Table 3.

[9] The adjusted rho-bar squared value is computed as $\bar{\rho}^2 = 1 - \frac{L(\hat{\boldsymbol{\delta}}) - M}{L(\text{c})}$, where $L(\hat{\boldsymbol{\delta}})$ is the log-likelihood value at convergence, L(c) represents the log-likelihood function at constants-only, and M is the number of parameters estimated in the model (excluding the constants). The Bayesian Information Criterion (BIC) is computed as $\left[-L(\hat{\boldsymbol{\delta}}) + 0.5 \times \left(\#\text{ of model parameters including constants}\right) \times \log\left(\text{sample size}\right)\right]$.

Table 3. Data fit measures

Metric		Proposed F-MNP	N-MNP	MNP
Disaggregate fit measures				
Log-likelihood at convergence		-566.26	-569.44	-576.82
Number of non-constant parameters[10]		48	48	46
Log-likelihood at constants-only		-743.53	-743.53	-743.53
Adjusted Rho-squared value		0.174	0.170	0.162
Bayesian Information Criterion or BIC		746.93	750.11	750.68
Average probability of correct prediction		0.641	0.638	0.631
Likelihood ratio (LR) test: N-MNP vs MNP		-	LR = 14.76> $\chi^2_{(2,0.05)}$ =5.991	
Likelihood ratio test: Proposed F-MNP vs N-MNP		LR = 6.36> $\chi^2_{(1,0.05)}$ =3.841		-
Aggregate fit measures				
Modes	Observed share (%)	Predicted share (%)	Predicted share (%)	Predicted share (%)
Auto-rickshaw	3.3%	3.5%	3.6%	3.4%
Metro	14.7%	15.0%	15.2%	15.2%
Bus	22.4%	27.5%	28.0%	28.7%
Walk	2.2%	1.3%	1.2%	1.2%
Two-wheeler	51.0%	47.3%	46.7%	46.2%
Four-wheeler	6.4%	5.4%	5.3%	5.3%
Weighted Absolute Percentage Error		**11.2%**	**12.8%**	**13.8%**

The likelihood-based fit measures and the average probability of correct prediction from the proposed F-MNP model indicate a better fit relative to both the N-MNP and the MNP models (see top row panel of Table 3). The pairwise likelihood tests also confirm that our proposed flexible model is statistically superior to the relatively restrictive models. In terms of aggregate data fit too (see bottom row of the panel), the proposed F-MNP model outperforms the N-MNP and the MNP models. Across all the choice alternatives, the weighted average (weighted on the observed shares) of the absolute percentage error is 11.2% for the proposed model which is superior to the WAPE of 12.8% and 13.8% for the other two models respectively. These differences are moderate because of the high skewness of the modal share distribution in our empirical context (the predictive capability of the two-wheelers substantially impacts the WAPE values, while the prediction accuracy of those modes with very little market shares is hardly consequential). However, such an observation is context-specific, and the results from the disaggregate and aggregate measures overall indicate a superior fit for our proposed model.

To ensure that the data fit of the proposed F-MNP model is not simply an artifact of overfitting on the overall estimation sample, we evaluate the performance of the proposed F-MNP model and the model that ignores unobserved heterogeneity – the MNP model, on various market segments of the estimation sample (Ben-Akiva and Lerman,1985, page 208, refer to such

[10] For the proposed F-MNP model and the N-MNP model, the number of non-constant parameters is set to 48, ignoring the copula correlation coefficient that was almost zero in both models. As discussed in Section 3.2, the log-likelihood values at convergence for both models were essentially the same with or without this copula correlation.

predictive tests as market segment prediction tests). These tests examine the performance of the models for different market segments. In instances when an out-of-sample predictive evaluation test is unavailable, such market-segmentation tests offer a robust means for assessing data fit. At a disaggregate level, we compute the implied predictive log-likelihood and compare the models using an informal chi-squared predictive log-likelihood ratio test. At an aggregate level, we compute the predicted and actual (observed) shares for each market segment in the same manner as for the full estimation sample, and then evaluate the performance of the two models using the WAPE measure. To conserve space, Table 4 presents these data fit statistics for nine market segments based on selected exogenous variables (gender, age, education, employment status, monthly income, household two-wheeler count, household car ownership, metro-pass availability, and commute start time from home). For each selected variable, the data fit for the market segment with the greatest number of observations is presented (for example, for the gender variable, Table 4 provides the data fit for men, because this segment represents 71.4% of the total sample). The results show that the informal predictive log-likelihood ratio tests (see the third numeric row of Table 4) reject the MNP model in preference for the F-MNP model for each market segment (at 0.10 significance level for one segment and at 0.05 significance level for all other segments), and also indicate that the predicted shares from the F-MNP mode are closer to the true shares than the predicted shares from the MNP model for each market segment (see the final numeric row of Table 4 for each segment). These observations provide additional support and validation that the proposed F-MNP model offers a robust data fit and that the superior predictions provided in Table 3 are not simply an artifact of overfitting.

Table 4. Aggregate and disaggregate measures of fit on various market segments of the estimation sample

Market Segment	Gender: Male		Age: Over 26-35 years		Education: Less than Undergraduate	
Measures of Fit	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**
Number of observations	653		337		458	
Mean log-likelihood	-402.91	-409.13	-207.64	-211.10	-282.74	-287.65
Informal predictive likelihood ratio test	12.45 > $\chi^2_{(3,0.05)}$=7.815		7.09 > $\chi^2_{(3,0.10)}$=6.251		9.83 > $\chi^2_{(3,0.05)}$=7.815	
WAPE	10.9%	13.3%	10.7%	12.7%	12.1%	14.3%
Market Segment	**Employment status: Private/Govt. sector**		**Monthly Income: ₹20000 -₹100000**		**Two-wheeler count: One**	
Measures of Fit	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**
Number of observations	572		396		471	
Mean log-likelihood	-353.12	-358.68	-244.21	-248.97	-290.44	-295.24
Informal predictive likelihood ratio test	11.11 > $\chi^2_{(3,0.05)}$=7.815		9.52 > $\chi^2_{(3,0.05)}$=7.815		9.60 > $\chi^2_{(3,0.05)}$=7.815	
WAPE	11.6%	13.9%	11.1%	13.8%	10.8%	13.3%
Market Segment	**Car ownership: Zero**		**Metro-pass: Not available**		**Commute start time: Peak-hours**	
Measures of Fit	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**	**Proposed F-MNP model**	**MNP Model**
Number of observations	625		693		613	
Mean log-likelihood	-385.39	-392.11	-427.46	-433.86	-378.75	-384.34
Informal predictive likelihood ratio test	13.45 > $\chi^2_{(3,0.05)}$=7.815		12.81 > $\chi^2_{(3,0.05)}$=7.815		11.18 > $\chi^2_{(3,0.05)}$=7.815	
WAPE	10.9%	13.3%	11.9%	14.1%	10.0%	12.2%

3.4 Elasticity Effects and Policy Implications

The coefficients in Table 2 provide the exogenous variable effects on the utilities of the choice alternatives; however, they do not directly provide a sense of the direction/magnitude effects of each variable on the discrete outcomes in terms of their impact on the overall shares. Therefore, we compute aggregate-level "pseudo-elasticity effects" of the exogenous variables to characterize the impact of each variable. In the current analysis, we have three types of exogenous variables: categorical variables (they are gender, age category, educational level, employment type, monthly income category, metro pass holder status, commute time category), count variables (they are two-wheeler count and four-wheeler count), and continuous variables (these are travel time and travel cost variables). For each of the binary category variables (gender, employment type, metro pass holder status, commute time category), we first predict the average share of each mode in the sample for the "base" level (which is typically the "0" for the binary variable), and then predict the average shares for the "treatment" level (which is typically the "1" for the binary variable) for the entire sample. The average "pseudo" elasticity effect is then reported as the difference between the "treatment" and the "base" shares as a percentage of the "base" share. For the multi-category (more than two) variables (such as age category, educational level, and monthly income category), we use a similar procedure except that we consider the lower extreme category as the "base" level and the higher extreme category as the "treatment" level, to keep our presentation simple. For the count variable of two-wheeler and four-wheeler numbers, we use the count of "1" as the "base" category and the count of "2" as the "treatment" category. Finally, for the continuous variables, we increase (or decrease) the value of the variable by 25% and express the percentage change with respect to the original shares (i.e., keeping the original values of the continuous variables as the "base"). Since our model specification includes two non-normal random parameters, the calculation of the shares is based on 500 random datasets (corresponding to the generation of the non-normal parameters), which are averaged at the end to obtain the "base" and "treatment" shares in each case discussed above.

Table 5 provides the pseudo-elasticity effects for our proposed F-MNP model. The numbers in the table may be interpreted as the percentage change in the shares of each mode due to a change in the exogenous variable. For example, the first numeric entry of -32.74% in the table indicates that the share of men choosing the auto-rickshaw mode is 32.74% less than the share of women choosing the auto-rickshaw mode. Other numerical entries in Table 5 may be interpreted in a similar manner. Our analysis can help inform several policy decisions, which we discuss next.

3.4.1 Policy implications

The results from Table 2 and the "pseudo" elasticity metrics from Table 5 can be used to identify traffic congestion alleviation strategies by encouraging a shift from private modes of transportation to more sustainable public transportation/walk modes. First, the elasticity effects indicate that older workers (over 45 years of age) are substantially more inclined toward the use of private cars compared to their younger peers (about 81.6% more likely to choose a private car). This large percentage increase is likely because no other modes provide as much convenience and comfort to older workers as does a private car. A possible way to make public transportation more attractive to older workers is to provide free feeder services to metro rail stations, lower-priced or free rail/bus passes for seniors, reserved seats or even reserved compartments for older citizens to ease the impact of overcrowding, and dedicated access and egress walkways inside landmark stations. Similarly, individuals with high formal education degrees, and who are self-employed, are also observed to have a predisposition toward private cars. An approach to draw these individuals away from private cars would be to offer public transportation services that are fast and provide a smooth ride during the journey to open up the possibility of using travel time productively (reading/working), even if at the expense of an elevated cost. Dedicated bus lanes that are not prevalent in many developing economies could also be another approach to achieve fast public transportation services. Second, in addition to making non-private vehicle travel more attractive, a complementary approach would be to discourage private vehicle travel through the use of tolls, congestion pricing and taxes. According to Table 5, an increase of 25% in private car travel costs would lead to a decrease in private car share by about 20% from the current car share. This suggests that imposing congestion pricing in highly congested central business districts (CBDs) or work zones could be effective in discouraging the use of private cars. Table 5 also indicates a substantial jump (by 184.9%) in the share of the private car mode when household vehicle ownership increases from one car to two private cars. Progressively higher taxes on a car purchase as the number of cars owned by a household increases can be one way to reduce private vehicle ownership.

Table 5. "Pseudo" elasticity effects

Variable	Base	Treatment	Auto-rickshaw	Metro	Bus	Walk	Two-wheeler	Private Car
Individual and household characteristics								
Gender	Female	Male	-32.74%	1.01%	-13.45%	-18.23%	21.56%	41.58%
Age	Less than =25	Greater than 45	-1.18%	-17.26%	12.43%	58.39%	-12.29%	81.61%
Education	Below Undergraduate	Graduate or more	0.24%	-1.78%	-5.20%	-80.37%	-0.65%	97.53%
Employment type	Govt./Private job	Self-employed/Business	68.06%	-30.15%	-48.42%	-96.13%	43.63%	80.27%
Monthly income	< ₹20,000	> ₹100,000	-25.30%	-29.94%	-14.61%	-20.80%	24.60%	46.45%
Two-wheeler count	1	2	-10.29%	-4.94%	-11.70%	-28.16%	13.46%	-15.14%
Four-wheeler count	1	2	-47.02%	0.12%	-18.77%	-27.61%	-1.85%	184.90%
Commute travel-related characteristics								
Metro pass holder	No	Yes	2.77%	125.22%	-14.22%	-67.60%	-6.86%	-47.59%
Peak hour commute	No	Yes	-28.39%	-4.72%	-14.41%	-25.49%	5.55%	50.42%
Trip Level Attributes								
Auto-rickshaw travel time	25% decrease		5.81%	-0.16%	-0.41%	-1.94%	-0.51%	-0.55%
Metro travel time	25% decrease		-0.26%	5.01%	-0.90%	0.61%	-0.83%	-1.30%
Bus travel time	25% decrease		-3.11%	-1.16%	5.73%	1.26%	-1.95%	-4.11%
Two-wheeler travel time	25% decrease		-2.51%	-2.01%	-3.04%	-1.99%	3.51%	-4.61%
Private car travel time	25% decrease		-1.94%	-0.24%	-0.61%	-0.44%	-0.99%	8.79%
Auto-rickshaw travel cost	25% increase		-24.36%	0.14%	1.31%	7.93%	0.61%	1.66%
Metro travel cost	25% increase		-0.59%	-2.85%	-0.39%	0.26%	1.11%	0.79%
Bus travel cost	25% increase		1.69%	0.69%	-3.28%	1.78%	1.34%	2.39%
Two-wheeler travel cost	25% increase		1.29%	1.46%	3.81%	2.21%	-3.33%	3.33%
Private car travel cost	25% increase		3.49%	0.48%	1.61%	0.61%	0.89%	-20.06%

Third, congestion reduction measures can also be based on work timing policies. In terms of elasticity effects, Table 5 indicates that the share of private car use is 50% higher during the peak period relative to the off-peak period, with a concomitant reduction in public transportation and walk mode usage during the peak period. This is presumably because the public modes are slow and crowded during peak periods, while the walk mode experiences packed crowds during peak periods and poses traffic safety risks. In this context, staggered work hours can help induce some of the commute traffic to off-peak periods (see, for example, Yildirimoglu et al., 2021). This not only can reduce overall commute travel, but also, based on our results, increase public transportation usage and decrease private mode usage for the commute trips shifted to the off-peak. Such a flexible work policy may be particularly palatable in an environment where there is more acceptance of flexible work arrangements in terms of location and timing. Finally, a generic consideration in many developing cities is the accessibility to public transit systems, including the city of Bengaluru. While the bus mode is available to most of the workers in the dataset (about 85%) based on their home and work locations, the metro mode is available to only about one-third of the workers (see the discussion on availability/feasibility of modes provided in the first paragraph of Section 3.1). Therefore, increasing accessibility to public transit systems is important for sustainability. In a policy-specific analysis, we decreased the out-of-vehicle travel time (as part of the total travel time) for the bus and the metro modes by 50% and compared the modal shares to that of the "base" case. This analysis was specific to a subset of the sample in which at least one of the public transit modes (bus and metro) was available to the user. Based on the availability criteria specified earlier, this out-of-vehicle travel time reduction resulted in an 8% increase in the combined share of both modes, and about a 10% decrease in the private car mode share. Thus, enhancing the accessibility of the public transit network by improving supply-side facilities, such as implementing dedicated feeder routes, increasing the number of stops or stations located within a reasonable walking distance of high-density residential areas, and expanding the transit network, particularly the metro mode, can potentially result in a significant shift in commute mode shares from private transportation to sustainable modes of travel.

3.4.2 Consequences of ignoring non-normality in parameter distribution

In this section, we briefly discuss the pitfalls of ignoring non-normality in the parameter distribution of travel attributes. To highlight this issue, we compute the "pseudo" elasticity effects for the N-MNP that ignores the non-normality in the distribution of the travel time coefficient (the corresponding table is suppressed to conserve on space) and compare these with the results from the F-MNP model. Although there are a few fairly significant differences in the "pseudo" elasticity effects of the sociodemographic variables between the two models, we limit our discussion to the differences in the elasticity effects of the travel attributes. Our analysis from Table 5 indicates that reducing the travel times of buses and metro by 25% leads to an increase of about 8.7% (=5.01+5.73-1.16-0.9) share in public transportation usage. A similar decrease in travel time, as estimated by the N-MNP model, suggests an increase of 6.3% share in public transportation use. Although the underestimation from the N-MNP model seems moderate, it can be enough to discourage investments in the design of exclusive bus lanes or in decisions related to improving the supply side of metro services. Similarly, comparing the "pseudo" elasticities for the travel cost increase treatment, our proposed model results suggest that private car use share would reduce, on average, by about 20% when travel costs for cars are increased by 25%, while the N-MNP model indicates a decrease of about 17% in the private car use. Once again, though the difference seems small, it can be enough to misinform the evaluation and economic analysis of critical policies such as those associated with congestion pricing and purchase taxation.

Another result of ignoring non-normality in travel time and travel cost coefficients, as evident from the discussion in Section 3.2, is the underestimation of the VTTS, which can have significant economic and travel demand-related consequences. The VTTS metric is often used in cost-benefit analysis in transportation infrastructure investments. Moreover, toll prices, exclusive-lane usage, and time-of-day based congestion pricing strategies are all generally based on VTTS estimates (see, for example, Börjesson et al., 2023). But applying an adjustment factor off an already erroneously computed VTTS value will bias results and compound errors in future scenario planning.

4 CONCLUSION

There is a critical need to address issues associated with traffic congestion and environmental considerations in developing economies, including ways to shift travelers from private non-sustainable modes to sustainable public and non-motorized transportation modes. In this context, mode choice behavior studies can provide important insights. In our current study, we propose a new flexible random parameter multinomial probit (F-MNP) model that allows for non-normality in parameter sensitivity. Using the YJ transformation technique, we construct an implicit Gaussian copula to form a multivariate distribution of the random coefficients with a flexible dependence structure. The proposed model is estimated using a hybrid approach of simulation-based likelihood estimation and analytic approximation of multivariate normal integrals.

Using our model, we investigate the commute mode choice behavior of workers in the Indian city of Bengaluru. The data is drawn from a survey administered from February to April of 2022 to collect information on the routine travel pattern of the residents of Bengaluru. The survey asked respondents about their routine travel destination, the purpose and most frequently used mode of travel to the routine destination. For this study, we used the data of only those who reported commuting as their routine travel. The survey also elicited information on individual characteristics, household demographics, and travel-related attributes. Results from our analysis indicate that sociodemographic variables (including gender, age, education level, employment type, household income, and the number of privately owned vehicles), commute characteristics (metro pass availability, commute start time), and mode-related attributes (travel time and travel cost) significantly impact the commute mode choice decision. Importantly, our results indicate the presence of unobserved taste heterogeneity in the sensitivities to the travel time and travel cost variables; moreover, the distribution of the travel time coefficient is found to be significantly non-normal with a fat left tail and thin right tail. In terms of data fit, our proposed model statistically outperforms the traditional MNP model as well as the N-MNP model that imposes normality on the travel time coefficient. By undertaking market segmentation tests, we also show that the improved data fit of our proposed model is not due to overfitting. The model estimates are subsequently translated to VTTS and pseudo-elasticity measures, and policy implications are identified. The consequences of ignoring non-normality are also discussed.

One important limitation of our empirical study is that it did not have a sufficient number of ride-hailing mode users (only five users reported to have used ride-hailing for commute, and therefore these observations were removed from the dataset). Moreover, the sample size was relatively small, although adequate to estimate the proposed model. But the limited sample size precluded us from setting aside a separate out-of-sample set as we opted to utilize the entire sample for model estimation purposes. Future efforts should focus on using a larger sample size to include consideration of all possible modes and the interaction effects of different exogenous variables, as well as on setting aside an exclusive out-of-sample validity set for prediction evaluation.

An important limitation of our methodological approach is that the YJ transformation is not able to accommodate multimodality of random coefficients. A possible approach to handle such multimodality is to transform the target multivariate random coefficient distribution to a mixture-of-normals distribution rather than to a normal distribution (this appears as the rightmost twig in Figure 2). Equation (1) would then take the following extended YJ transformational form:

$$G_l \sim \sum_{i=1}^{I} \pi_i N(\mu_{li}, \sigma_{li}^2) = t_{\lambda_l}(Z_l) = \begin{cases} -\dfrac{(-Z_l+1)^{2-\lambda_l}-1}{2-\lambda_l} & \text{if } Z_l < 0 \\ \dfrac{(Z_l+1)^{\lambda_l}-1}{\lambda_l} & \text{if } Z_l > 0 \end{cases} \tag{12}$$

Such an approach would be even more general than the mixture-of-normals approach because of the transformation, and should have the benefit of more appropriately capturing skew and fat tails with fewer mixture components. Of course, the issues discussed earlier related to the challenges in the estimation of mixture models within the context of econometric methods will need to be borne in mind. Cai and Xu (2023) have recently considered such an extended YJ approach for the statistical clustering of a univariate continuous process. But, as they acknowledge, their Bayesian approach does not address the case of an econometric model framework for a continuous outcome, leave alone within an econometric framework for other limited-dependent outcomes. Besides, even within their simple univariate continuous outcome clustering exercise, they observe that their approach cannot handle many mixture components. Thus, accommodating a fusion of a YJ approach and a mixture approach within a discrete choice model for random coefficients, and developing approaches to estimate such a model, is a substantial challenge that we leave for future research. More generally, the issue of how to introduce non-normality of random coefficients in discrete choice models remains a wide area for additional research. For instance, even within just the finite discrete mixture of parametric continuous distributions class of approaches, there are so many different possibilities for the component distributions, as discussed in a recent overview by Lee and McLachlan, 2022. Most of these approaches have been examined for the simple clustering analysis of univariate continuous processes, with relatively little work into examining the different mixture approaches for the econometric modeling of continuous or limited-dependent variable models. Thus, just within the transportation discrete choice field, we are not aware of any study that has used a discrete mixture-of-skew normal distributions or a discrete mixture-of-skew-t distributions for random coefficients, and only a handful of studies that have used a discrete mixture-of-normals for random coefficients (as discussed in the introduction section). Given the situation discussed above, we leave a comparison of our YJ-based approach with the vast landscape of other approaches, including the mixture-of-normals approach, to future studies, especially because the main emphasis of this paper is to introduce a new way to consider non-normality for random coefficients. Besides, the basis of the effectiveness of an approach should not rest on application to a single empirical context. As Lee and McLachlan (2022) state, there are pros and cons of the many different ways to introduce non-normality, and "appropriate choices of models will depend on the application at hand ….and factors and features ….such as computational affordability, ease of interpretation, and facilitation of downstream analysis". From a methodological perspective, we have proposed, and demonstrated the value of, the YJ transformation approach for random coefficients relative to the more traditional normal mixing approach, and positioned our flexible formulation as a valuable addition to the toolbox of discrete choice modelers, which also has clear theoretical and conceptual advantages of parsimony in parameters and ease in estimation relative to most other approaches to incorporating non-normality. Of note also is that the YJ transformation can be used both in a frequentist context as well as within a variational inference approach in a scalable approximate Bayesian inference context to approximate high-dimensional target distributions with skews and heavy tails (see Bean et al., 2016, Smith et al., 2020, and Loaiza-Maya et al., 2022, Cai and Xu, 2023). Substantively, the results from our investigation provide useful insights regarding commuters' mode choice behavior in the city of Bengaluru. To conclude, we hope our study will spur many other investigations of travel choices using flexible discrete choice model structures, especially in the context of developing economies.

ACKNOWLEDGMENTS
This research was partially supported by the U.S. Department of Transportation (USDOT) under Grant 69A3552047138 for the Center for Automated Vehicles Research with Multimodal AssurEd Navigation (CARMEN) University Transportation Center. The empirical data used in this paper was collected as part of a project entitled "*Data-driven models and decision support tools for improved transit reliability in Indian cities*", funded by the Ministry of Electronics and Information Technology of the Government of India. The authors acknowledge the inputs of Karthik Srinivasan and Ganesh Ambi Ramakrishnan in survey design and the efforts of Mehek Biswas, Sandhya Sourirajan, Ruchika Mattoo, Deepa L, Najeebul Feroz Malik, Shobhit Saxena, and several survey enumerators in survey administration and data preparation. The authors are grateful to Lisa Macias for her help in formatting this document, and also thank Ganesh Bharate for help with the plots in

Figures 1 and 2, and Dale Robbennolt for help with the plots in Figures 3, 4, and 5. We also appreciate the valuable anonymous review input obtained on an earlier version of this paper.

REFERENCES

Athira, I. C., Muneera, C. P., Krishnamurthy, K., and Anjaneyulu, M. V. L. R. (2016). Estimation of value of travel time for work trips. *Transportation Research Procedia*, *17*, 116-123.

Balcombe, K., Chalak, A., and Fraser, I. (2009). Model selection for the mixed logit with Bayesian estimation. *Journal of Environmental Economics and Management*, *57*(2), 226-237.

Bastin, F., Cirillo, C., and Toint, P.L. (2010). Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. *Transportation Science*, *44*(4), 537-549.

Batur, I., Sharda, S., Kim, T., Khoeini, S., Pendyala, R. M., and Bhat, C. R. (2019). Mobility, time poverty, and well-being: How are they connected and how much does mobility matter? Technical paper, Arizona State University.

Bauer, D., Büscher, S., and Batram, M. (2022). Non-parametric estimation of mixed discrete choice models. *arXiv preprint arXiv:2212.04855*.

Bean, A., Xu, X., and MacEachern, S. (2016). Transformations and Bayesian density estimation. *Electronic Journal of Statistics*, *10*, 3355-3373.

Ben-Akiva, M. E., and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand* (Vol. 9). MIT press.

Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, *31*(1), 34-48.

Bhat, C. R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A*, *32*(7), 495-507.

Bhat, C.R. (2000). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science*, *34*(2), 228-238.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, *35*(7), 677-693.

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, *37*(9), 837-855.

Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B*, *109*, 238-256.

Bhat, C.R. (2020). Consumer choice modeling: The promises and the cautions. *Mapping the Travel Behavior Genome*, Chapter 5, 63-80, edited by K.G. Goulias and A.W. Davis, Elsevier.

Bhat, C. R., and Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, *43*(7), 749-765.

Bhat, C. R., and 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., and Mondal, A. (2022). A new flexible generalized heterogeneous data model (GHDM) with an application to examine the effect of high density neighborhood living on bicycling frequency. *Transportation Research Part B*, *164*, 244-266.

Bhat, C. R., and Sidharthan, R. (2012). A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B*, *46*(7), 817-833.

Bhat, C. R., Astroza, S., and Bhat, A. C. (2016). On allowing a general form for unobserved heterogeneity in the multiple discrete–continuous probit model: Formulation and application to tourism travel. *Transportation Research Part B*, *86*, 223-249.

Börjesson, M., Asplund, D., and Hamilton, C. (2023). Optimal kilometre tax for electric vehicles. *Transport Policy*, *134*, 52-64.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211-243.

Buddhavarapu, P., Scott, J. G., and Prozzi, J. A. (2016). Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. *Transportation Research Part B*, *91*, 492-510.

Cai, J., and Xu, X. (2023). Bayesian analysis of mixture models with Yeo-Johnson transformation. *Communications in Statistics-Theory and Methods*, *53*(18), 6600-6613.

Census of India (2011). https://censusindia.gov.in/census.website/

Cherchi, E., Cirillo, C., and Polak, J. (2009). User benefit assessment in presence of random taste heterogeneity: Comparison between parametric and nonparametric models. *Transportation Research Record*, *2132*, 78-86

Daganzo, C. (2014). *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Elsevier.

Denzer, M. (2019). Estimating causal effects in binary response models with binary endogenous explanatory variables: A comparison of possible estimators. Discussion paper number 1916, Gutenberg School of Management and Economics, Johannes Gutenberg University Mainz, Germany.

Dias, C., Abdullah, M., Lovreglio, R., Sachchithanantham, S., Rekatheeban, M., and Sathyaprasad, I. M. S. (2022). Exploring home-to-school trip mode choices in Kandy, Sri Lanka. *Journal of Transport Geography*, *99*, 103279.

Dong, A., Melnykov, V., Wang, Y., and Zhu, X. (2023). Conditional mixture modelling for heavy-tailed and skewed data. *Stat*, *12*(1), e608. doi.org/10.1002/sta4.608.

Dong, X., and Koppelman, F.S. (2014). Comparison of continuous and discrete representations of unobserved heterogeneity in logit models. *Journal of Marketing Analytics*, *2*, 43-58.

Dong, Y., and Lewbel, A. (2015). A simple estimator for binary choice models with endogenous regressors. *Econometric Reviews*, *34*(1-2), 82-105.

Everitt, B. (2013). *Finite Mixture Distributions*. Springer Science and Business Media.

Fruhwirth-Schnatter, S., and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-*t* distributions. *Biostatistics*, *11*(2), 317-336.

Gallaugher, M.P.B., McNicholas, P.D., Melnykov, V., and Zhu, X. (2020). Skewed distributions or transformations? Modelling skewness for a cluster analysis. https://doi.org/10.48550/arXiv.2011.09152, accessed January 7, 2024

Greene, W. H., Hensher, D. A., and Rose, J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B*, *40*(1), 75-92.

Greene, W.H., and Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, *37*(8), 681-698.

Guo, J., Feng, T., and Timmermans, H. J. (2020). Co-dependent workplace, residence and commuting mode choice: Results of a multi-dimensional mixed logit model with panel effects. *Cities*, *96*, 102448.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*, 84-90.

Hao, Y., and Kasahara, H. (2022). Testing the number of components in finite mixture normal regression model with panel data. *arXiv preprint arXiv:2210.02824*.

Harsha, V., Karmarkar, O., and Verma, A. (2020). Sustainable urban transport policies to improve public transportation system: A case study of Bengaluru, India. *Transportation Research Procedia*, *48*, 3545-3561.

Jadhav, A., Dhaulakhandi, D., Kumar, S., Malviya, L., and Mewada, A. (2023). Data transformation: A preprocessing stage in machine learning regression problems. *Artificial Intelligence Techniques in Power Systems Operations and Analysis*, Singh, N., Tamrakar, S., Mewada, A., and Gupta, S.K. (Eds.). Auerbach Publications, https://doi.org/10.1201/9781003301820

Krueger, R., Rashidi, T. H., and Vij, A. (2020). A Dirichlet process mixture model of discrete choice: Comparisons and a case study on preferences for shared automated vehicles. *Journal of Choice Modelling*, *36*, 100229.

Lee, S. X., and McLachlan, G. L. (2022). An overview of skew distributions in model-based clustering. *Journal of Multivariate Analysis*, *188*, 104853.

Li, Z., Chen, C., Wu, Q., Zhang, G., Liu, C., Prevedouros, P. D., and Ma, D. T. (2018). Exploring driver injury severity patterns and causes in low visibility related single-vehicle crashes using a finite mixture random parameters model. *Analytic Methods in Accident Research*, *20*, 1-14.

Lin, T. I., Lee, J. C., and Yen, S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, *17*, 909-927.

Lin, T. I., McLachlan, G. J., and Lee, S. X. (2016). Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *143*, 398-413.

Loaiza-Maya, R., Smith, M. S., Nott, D. J. and Danaher, P. J. (2022), Fast and accurate variational inference for models with many latent variables. *Journal of Econometrics*, *230*(2), 339-362.

Mahadevia, D., and Advani, D. (2016). Gender differentials in travel pattern–the case of a mid-sized city, Rajkot, India. *Transportation Research Part D*, *44*, 292-302.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519-530.

Marimuthu, S., Mani, T., Sudarsanam, T. D., George, S., Jeyaseelan, L. (2022). Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research. *Clinical Epidemiology and Global Health*, *15*, 101043.

Masoumi, H. E. (2019). A discrete choice analysis of transport mode choice causality and perceived barriers of sustainable mobility in the MENA region. *Transport Policy*, *79*, 37-53.

McLachlan, G. J., and Rathnayake, S. (2014). On the number of components in a Gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(5), 341-355.

McLachlan, G.J., Lee, S.X., and Rathnayake, S.I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, *6*, 355-78.

Miller, S.J. (2015). Probability Lifesaver: Order Statistics and the Median Theorem. https://web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/supplementalchap_mediantheoremandorderstatistics.pdf

Ministry of Road Transport and Highways. (2021). Road Transport Year Book (2017-18 and 2018-19). https://morth.nic.in/sites/default/files/RTYB-2017-18-2018-19.pdf

Mittelhammer, R.C., and Judge, G. (2011). A family of empirical likelihood functions and estimators for the binary response model. *Journal of Econometrics*, *164*(2), 207-217.

Mu, B., and Zhang, Z. (2018). Identification and estimation of heteroscedastic binary choice models with endogenous dummy regressors. *Econometrics Journal*, *21*(2), 218-246.

National Statistical Survey Office. (2012). Employment Unemployment Survey. Ministry of Statistical Programme and Implementation, GoI.

National Statistical Survey Office. (2021). Periodic Labor Force Survey, Ministry of Statistical Programme and Implementation, GoI.

Nayka, S., and Sridhar, K. S. (2019). Determinants of intra urban mobility: A study of Bengaluru. Bengaluru, India: Institute for Social and Economic Change.

Orvin, M. and Fatmi, M.R. (2020). Modeling destination choice behavior of the dockless bike sharing service users. *Transportation Research Record*, *2674* (11), 875-887.

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, *15*, Article 12.

Panambur, V. R., and Sushma, V. (2019). Study of challenges faced by visually impaired in accessing Bangalore metro services. In *Proceedings of the 10th Indian Conference on Human-Computer Interaction* (1-10).

Parmar, J., Saiyed, G., and Dave, S. (2021). Analysis of taste heterogeneity in commuters travel decisions using joint parking and mode choice model: A case from urban India. *arXiv preprint arXiv:2109.01045*.

Patil, G. R., Basu, R., and Rashidi, T. H. (2020). Mode choice modeling using adaptive data collection for different trip purposes in Mumbai metropolitan region. *Transportation in Developing Economies*, *6*, Article 9.

Patil, P.N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., and Bhat, C. R. (2017). Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of Choice Modelling*, *23*, 9-20.

Ram, S.G., and Dhawan, H. (2018). No ban but plenty of bias: Why so few women drive in India. *The Times of India*. https://timesofindia.indiatimes.com/home/sunday-times/no-ban-but-plenty-of-bias-why-so-few-women-drive-in-india/articleshow/64809217.cms

Ramakrishnan, A., Kalkuhl, M., Ahmad, S., and Creutzig, F. (2020). Keeping up with the Patels: Conspicuous consumption drives the adoption of cars and appliances in India. *Energy Research and Social Science*, *70*, 101742.

Rossi, P. (2014). *Bayesian non-and semi-parametric methods and applications*. Princeton University Press. https://doi.org/10.1515/9781400850303

Shirgaokar, M. (2014). Employment centers and travel behavior: Exploring the work commute of Mumbai's rapidly motorizing middle class. *Journal of Transport Geography*, *41*, 249-258.

Smith, M. S., Loaiza-Maya, R., and Nott, D. J. (2020). High-dimensional copula variational approximation through transformation. *Journal of Computational and Graphical Statistics*, *29*(4), 729-743.

Sridhar, K. S., and Nayka, S. (2022). Determinants of commute time in an Indian city. *Margin: The Journal of Applied Economic Research*, *16*(1), 49-75.

Statista. (2019). Length of daily commute by urban dwellers across India in 2019. https://www.statista.com/statistics/1103900/india-urban-dwellers-daily-commute-distance/.

Torres, C., Hanley, N., and Riera, A. (2011). How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management*, *62*(1), 111-121.

Train, K. (2009). *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, Cambridge.

Train, K., and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In *Applications of Simulation Methods in Environmental and Resource Economics*, 117-134. Springer, Dordrecht.

Vij, A., and Krueger, R. (2017). Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions. *Transportation Research Part B*, *106*, 76-101.

Wang, K., Bhat, C. R., and Ye, X. (2023). A multinomial probit analysis of shanghai commute mode choice. *Transportation*, *50*(4), 1471-1495.

Watthanacheewakul, L. (2021). Transformations for left skewed data. *Proceedings of the World Congress on Engineering 2021 (WCE 2021)*, July 7-9, 2021, London, U.K.

World Bank. (2022). Urban Population (% of the total population). United Nations Population Division. Available at: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS.

Xiong, Y., and Mannering, F. (2013). The heteroscedastic effects of guardian supervision on adolescent driver-injury severities: A finite mixture-random parameters approach. *Transportation Research Part B*, *49*, 39-54.

Yeo, I. K., and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954-959.

Yildirimoglu, M., Ramezani, M., and Amirgholy, M. (2021). Staggered work schedules for congestion mitigation: A morning commute problem. *Transportation Research Part C*, *132*, 103391.

Zhang, H., and Huang, Y. (2015). Finite mixture models and their applications: A review. *Austin Biometrics and Biostatistics*, *2*(1), 1013.