A Flexible Multinomial YJ-Based Model with Non-Normal Random Coefficients and Non-Normal Kernel Error Terms

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

Dale Robbennolt
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Email: dar4836@utexas.edu

Abstract
In this study, we propose a flexible, computationally tractable, structurally simple, and parsimonious-in-specification random coefficients multinomial Yeo-Johnson (MNYJ) model to accommodate both non-normal random coefficients as well as non-normal kernel errors, while also accommodating dependence across the random coefficients and the kernel error terms. This is accomplished using an implicit Gaussian copula to tie the many random elements together. The net result is that our proposed random coefficients MNYJ model (or the MNYJRC) model, while providing substantially more flexibility than the traditional multinomial probit (MNP) model, also enables intuitive interpretations of scaling on, and dependencies across, the structural kernel error terms, as opposed to a lack of such interpretations in the MNP model (where the analyst can only estimate the scale of, and correlations across, the differenced kernel errors). This additional interpretation ease is because of the asymmetric distribution specification, along with varying skew effects, on the kernel error terms, so that the differences in kernel errors are not symmetric. We validate the ability of our model, and the associated estimation procedure, to recover underlying parameters through simulation experiments, as well as demonstrate the pitfalls of using traditional models that a priori impose symmetric distributions on the kernel error differences. Using our new flexible MNYJ framework, we also investigate future vehicle ownership intentions (the choice among purchasing a regular vehicle, purchasing an autonomous vehicle, or relying on autonomous ridehailing services), highlighting the policy implications of ignoring kernel error asymmetry.

Keywords: Asymmetric and flexible kernel error distributions; Unobserved heterogeneity; Gaussian copula; Choice models; Inverse YJ transformation

1 INTRODUCTION

Discrete choice models are widely used in a variety of disciplinary fields to understand the factors influencing choice from a set of available discrete alternatives, such as the choice of a specific brand of yoghurt, or the choice of where to live, or the choice of which mode to travel by for a specific trip or a series of trips. The most common discrete choice models are based on utility-maximizing theory, with different models arising from different assumptions made regarding the kernel error terms in the latent utility of each alternative. The two most common distributions used for the kernel error terms are the Type 1 extreme value (an asymmetric distribution) and the normal distribution (a symmetric distribution). Combined with the rather restrictive (and typically difficult to justify) identically and independently distributed (IID) assumption across the errors of the alternatives, the Type 1 extreme value kernel results in the familiar multinomial logit (MNL) model. The closed form structure of this model derives from the property that the difference of two IID Type 1 extreme value errors is the symmetric logistic distribution. But the IID structure also saddles the MNL with the independence of irrelevant alternatives (IIA) property at the individual level. On the other hand, the normal distribution assumption for the kernel errors allows a more flexible covariance structure among the utilities, and leads to the multinomial probit (MNP) model. Again, the effective distribution that matters is the stochastic component of utility differences, which, like in the MNL model, takes a symmetric form because of the closure property of the normal distribution under subtraction.

Of course, over the past several decades, many extensions of the MNL and MNP models have been proposed, based on mixing approaches over the MNL and MNP kernels (see Bhat, 2020 for a detailed review of random utility-based discrete choice literature; Train, 2009 is a textbook resource). In this mixing context, the arguments and positioning by McFadden and Train (2000) has sometimes been interpreted, somewhat inappropriately, to imply that any random utility model (RUM) can be approximated by a mixed MNL or a mixed MNP model. However, an important caveat here is that this is only true if the random distributions used in the mixing process (whether for random coefficients associated with observed variables or for error components introduced to engender correlations/heteroscedasticity across alternatives) are appropriately specified. In this regard, it is important to note that most mixed models focus on a pure random coefficients specification (with IID utility conditional on the random coefficients) to accommodate potential taste variations across individuals in the sensitivity to observed exogenous variables (see, for example, Guo et al., 2020, Patil et al., 2020, Dias et al., 2022, Vichiensan et al., 2025, Wu et al., 2026). Within such random coefficients-based mixing models, recent studies have explored a variety of non-normal asymmetric parametric distributions or semi- and non-parametric flexible asymmetric distributions for the random coefficients themselves (see Bhat et al., 2025 for a detailed discussion). However, even though such specifications may induce asymmetric overall utility distributions, the kernel error terms themselves continue to be specified in a way that lead to symmetric differenced error distributions. Consequently, conditional on the random coefficients, the residual stochastic component of each utility difference remains symmetric around zero. Thus, such models may allow asymmetric heterogeneity in observed-variable sensitivities across individuals, but they retain symmetry in the conditional kernel error differences.

Similarly, while fewer mixed models with a logit kernel consider an error-components approach to induce heteroscedasticity and covariance effects across alternatives, such error-components specifications also almost universally rely on normally distributed error components layered on top of the IID Type 1 extreme value kernel. As a result, the implied differenced kernel

error structure remains symmetric. Further, when an MNP kernel is used (rendering additional error-components structures unnecessary because of the already flexible covariance structure of the MNP kernel), the differenced utility kernels are immediately symmetric because of the closure property of the normal distribution under subtraction.

In summary, in mixing models, whether of the random-coefficients variety or the error-components variety, it is almost universally the case that the kernel error term distributions are such that the differences in the kernel error terms are symmetric (in contrast to the substantial attention on asymmetric taste variations on specific exogenous variables in random-coefficients specifications). However, kernel error asymmetry in utilities (and, in particular, in utility differences) may arise from several fundamental behavioral mechanisms, including asymmetric intrinsic (and unobserved to the analyst) preferences toward alternatives and/or loss-gain asymmetry where individuals perceive losses more acutely than gains (see Kahneman and Tversky, 1979). An example of asymmetric preferences in a mode choice context between driving and public transport is that many individuals may exhibit only a mild preference for driving, while a smaller subset places exceptionally high value on the flexibility and control associated with driving. This would generate a rightward skew in the utility distribution associated with driving. Conversely, many individuals may exhibit only a modest aversion toward public transport, while a smaller subset may be strongly averse to schedule constraints and the perceived lack of control associated with transit use, leading to a leftward skew in the utility distribution associated with public transport. An example of loss-gain asymmetry in decisions regarding autonomous vehicle (AV) adoption is that many individuals may perceive the potential benefits of AV ownership, such as reduced driving burden or increased mobility, as relatively modest, while a smaller subset may place substantial weight on concerns regarding technological reliability, safety, liability, cybersecurity, or the loss of driving enjoyment. Such a pattern would induce a leftward skew in the kernel error distribution associated with AV ownership utility, with many individuals being only mildly favorable toward AV ownership but a smaller subset being strongly opposed to it because of intrinsic concerns not observed by the analyst. Conversely, AV ridehailing may represent a lower-commitment pathway for experiencing AV technology. Many individuals may exhibit only a modest aversion toward AV ridehailing because of concerns regarding privacy, safety, or loss of personal control. However, a smaller subset may view AV ridehailing exceptionally favorably because it allows them to obtain many of the anticipated benefits of AV technology without the substantial financial commitment and perceived risks associated with vehicle ownership. This would generate a rightward skew in the utility distribution associated with AV ridehailing.

Overall, the consideration of asymmetric kernel error terms (specifically, allowing for asymmetric distributions in differenced utility kernels) is not simply a statistical esoteric pursuit, but is important from a behavioral standpoint. In particular, ignoring kernel error asymmetry (including differing extents of asymmetry across alternatives) invites the pitfalls of biased and misleading estimates of the effects of exogenous variables, as the kernel asymmetry and thicker tails may contaminate the parameters characterizing the distribution effects of exogenous variables. More generally, ignoring kernel error asymmetry when it is present constitutes a likelihood misspecification, and leads maximum likelihood estimators to converge to *pseudo-true* values that minimize Kullback–Leibler divergence, rather than the underlying structural behavioral "truth" (White, 1982). This is also borne out from empirical studies showing that imposing symmetric links when the data-generating process is asymmetric produces appreciable bias in

slope parameters and preference distributions (Stukel, 1988; Bhat et al., 2025), reinforcing the importance of allowing flexible and asymmetric kernel error structures in discrete choice models.

1.1. Related Studies

Some early studies in econometrics considered asymmetric kernel error distributions in choice models, though there has been a long hiatus in the literature on this topic. These early models considered a binary choice case, embedding asymmetry through asymmetric link functions such as the complementary log–log link (Prentice and Gloeckler, 1978), the scobit model (Nagler, 1994), or generalized logistic families (Stukel, 1988). These studies impose asymmetry directly on the differenced utilities between the two alternatives. While useful, the econometrics changes considerably between binary choice and multinomial choice when asymmetric kernel error terms are used. In particular, while in binary choice models, the direction in which utility differences are taken does not affect the model results (except for a change in asymmetry from one direction to another), imposing asymmetry directly on utility differences in a multinomial choice context is not invariant to which alternative is used as the base.

Some other studies have specified relatively restrictive asymmetric distributions directly for the kernel errors in multinomial choice models (see Paleti, 2019 for an expansive review). To develop a model that does not exhibit the IIA property, Daganzo (1979) used independent negative exponential distributions with different variances for the random-error components to propose a closed-form discrete choice model (the variance differences cause skew in the differenced utilities). However, the negative exponential distribution is inherently asymmetric because it is defined only over the positive real line with support $[0,\infty)$, implying the rather unrealistic situation that errors can only increase latent utility. So, the Daganzo model has not seen much application. The Weibit model of Castillo et al., 2008 and Fosgerau and Bierlaire, 2009 may be viewed as an extension of the Daganzo model with a Weibull error term, with the additional twist that the typical additive kernel error term in each alternative's utility is replaced with a multiplicative error term. As with the negative exponential, the error term is foundationally asymmetric, since the Weibull distribution is one-sided and skewed. This restrictive range of the Weibull distribution implies that the systematic utility components must satisfy specific sign restrictions (e.g., being strictly positive or negative depending on the parameterization) for the choice probabilities to remain valid. Besides, the shape of the asymmetry is tightly tied to the Weibull distribution's form, and may not reflect the full spectrum of skew and tail behaviors that empirical data may demand. Further, the multiplicative specification of shocks can be more difficult to justify than the additive specification. Together, these considerations have limited the use of the Weibit model for general-purpose discrete choice modeling. A third model that allows the kernel error terms to span the entire real line is Bhat's (1995) heteroscedastic extreme value (HEV) model. Using the usual additive error structure, Bhat's model adopts Type 1 extreme value error terms but with different scales for each alternative, which leads to asymmetric kernel error term differences and again relaxes the strict IID assumption of the MNL. Again, though, this model is restrictive in that the skew and tail behaviors are controlled by only two parameters that characterize the Type 1 extreme value distribution's form. Also, while analytically convenient and allowing the kernel error distribution to have full range, the HEV, like the negative exponential and Weibit models, generates asymmetry in the differenced kernel errors through scale variations across alternatives (with the scale of one alternative being normalized to one and other scales being estimated relative to the normalized scale). However, all of these models maintain independence across the kernel

error terms of the alternatives and so cannot generate flexible cross-alternative dependencies through the kernel terms.

A class of multinomial choice models, although not focused on asymmetry as in the current paper, but that allows for cross-alternative correlations with also allowance for symmetric heavy-tailed distributions at both extremes, are those reviewed recently by Kreuger et al. (2023). The most flexible of these are the multinomial robit (MNR) models of Dubey et al. (2020), Peyhardi (2020), and Kreuger et al. (2023). All of these models assume that the utility differences follow a symmetric and heavy-tailed multivariate t-distribution. Dubey et al. (2020) and Peyhardi (2020) begin from the structural model of utilities, with the kernel error terms across alternatives having a single generic degree of freedom (DOF) parameter that determines tail thickness. Krueger et al. (2023) claim that this is a flexibility restriction, though there is a substantial advantage to adopting a single DOF MNR formulation. Specifically, doing so ensures that the differenced utilities are also multivariate-t, because of the closure property of the elliptically contoured multivariate-t distribution under subtraction. In addition, the identification considerations are exactly the same as for the MNP model, where the scale of one of the differenced utility errors is normalized. Importantly, just as in the MNP model, the results are invariant to which one of the scales in the differenced utility space is normalized. On the other hand, Krueger et al. (2023) directly impose a multivariate t-distribution on the differenced errors (with differences taken with respect to a designated last alternative), allowing for different DOF parameters across the differenced errors (that is, allowing for a non-elliptically contoured multivariate t-distribution; they label this as the Gen-MNR model). Unlike the case of a single DOF parameter in the structural utility space, it is unclear in the Gen-MNR formulation what the multivariate distribution of the structural errors is. Technically, it is difficult then to associate results to tail thickness intensities of different alternatives. More importantly, Kreuger et al.'s formulation has an exchangeability problem in that taking differences with respect to a different base alternative (than the last alternative), and allowing different DOF parameters in this new differenced space, will produce different estimation results.

In this paper, we adopt the inverse Yeo-Johnson (YJ) transformation of latent normal error terms to generate asymmetry in the kernel errors. Our model, which we label as the MNYJ model, offers a general, flexible, and elegant formulation by allowing a broad spectrum of asymmetric shapes, with the added advantage of heterogeneity in the skewed distributions across alternatives in scale, skew, and Kurtosis (tailedness). Our formulation is based on specifying asymmetry in the distributions of the kernel error terms of the original (structural) utilities, rendering the model invariant to the way utility differences are taken. Bhat et al. (2024) discuss the inverse YJ transformation in detail, demonstrating that it can mimic a whole variety of other flexible unimodal distributions closely, including the extreme value, the skew-normal, and skew-t distributions. Further, we are able to allow a flexible cross-alternative correlation structure through the use of an implicit Gaussian copula. Compared to other multivariate skew distributions, this transformation-based multivariate inverse YJ distribution is much more parsimonious, while also performing at least as well as even complicated mixtures of traditional Azzalini-type multivariate skew distributions (see Azzalini, 2005, Gallaugher et al., 2020).

In summary, to our knowledge, the current paper constitutes the first formulation and application of a flexible alternative-specific asymmetric multinomial choice model. The probability expressions collapse to the form of a mixing (over a univariate normal density function) of an (I–1) multivariate normal cumulative distribution function (MVNCD), where I is the number of alternatives. Random coefficients do not materially change the structure of the model, enabling

the seamless extension to asymmetric random-coefficients specifications along with an asymmetric kernel error term. Estimation is achieved through the use of an analytic evaluation of the MVNCD function based on Bhat (2018), followed by an integration using Gauss-Hermite quadrature. Identification conditions are quite different from the case of the MNP model, as the nonlinear transforms through the inverse YJ provides for true additional flexibility by way of being able to estimate quite a few more parameters than when the utility difference matrix has a Gaussian covariance structure. The net result is that the formulation generically enables the identification and interpretation of relative scales, correlations, and shapes at the structural kernel error level, unlike the case of the MNP where only the error difference covariance matrix is identified and it is not possible to interpret the covariance matrix at the structural level of the original utilities.

2 METHODOLOGY

2.1 The YJ Transformation

The Yeo and Johnson (2000) or the YJ transformation is a single-parameter transformation from an asymmetric distribution to a more symmetric target distribution. While any symmetric target distribution may be used, it is typical to use the normal distribution because of the nice properties of the normal distribution and the ability to build multivariate normal distributions using the transformed marginal normals. Consider an asymmetric random error ε_i that is YJ transformed into a standard normal random term η_i as follows (λ_i is a transformation parameter: $0<\lambda_i<2$):

$$\eta_i \sim N(0,1) = t_{\lambda_i}(\varepsilon_i) = \begin{cases} -\dfrac{(-\varepsilon_i+1)^{2-\lambda_i}-1}{2-\lambda_i} & \text{if } \varepsilon_i < 0 \\ \dfrac{(\varepsilon_i+1)^{\lambda_i}-1}{\lambda_i} & \text{if } \varepsilon_i > 0 \end{cases} \tag{1}$$

The more relevant transformation for discrete choice modeling is the inverse YJ transformation to characterize the asymmetric distribution of ε_i as (note that because the YJ transformation is strictly monotonic, its inverse transformation exists):

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

In the above inverse transformation, when $0<\lambda_i<1$, ε_i is skewed to the right with a thicker right tail, while if $1<\lambda_i<2$, ε_i is skewed to the left with a thicker left tail. When $\lambda_l=1$, the normal distribution is returned for ε_i. The transformation is quite general, and is able to represent a whole variety of asymmetric unimodal distributions very closely (including the skew normal and skew-t) as illustrated in Bhat et al. (2025). The cumulative distribution function (CDF) and probability density function (PDF) of ε_i may be obtained as:

$H_{\varepsilon_i}(z) = \text{Prob}(t_{\lambda_i}^{-1}(\eta_i) < z) = \text{Prob}(\eta_i < t_{\lambda_i}(z)) = \Phi\left[t_{\lambda_i}(z)\right] = \Phi\left[g\right]$, where $g = t_{\lambda_i}(z)$, and

$$h_{\varepsilon_i}(z) = \frac{\partial H_{\varepsilon_i}(z)}{\partial z} = \frac{\partial \Phi\left[t_{\lambda_i}(z)\right]}{\partial t_{\lambda_i}(z)} \times \left|\frac{\partial t_{\lambda_i}(z)}{\partial z}\right| = \phi\left[g\right] \times \left|\frac{\partial g}{\partial z}\right| = \phi\left[g\right] \times \left(|z|+1\right)^{\text{sgn}(z)(\lambda_i - 1)}, \tag{3}$$

$\Phi(.)$ and $\phi(.)$ above refer to the CDF and PDF of the standard univariate normal distribution. $\text{sgn}(z)$ is a sign-based dummy variable with a value of 1 if z is positive, a value of –1 if z is negative, and a value of 0 if z is zero. The moments of ε_i (including the mean, standard deviation and the skew/kurtosis) are all dependent upon λ_i. While there is no known closed form for these moments as a function of the YJ parameter λ_i, they are easily computed from the density function in Equation (3) as:

$$\mu_i(\lambda_i) = \int_{z=-\infty}^{+\infty} z h_{\varepsilon_i}(z)\, dz = \int_{g=-\infty}^{+\infty} t_{\lambda_i}^{-1}(g)\phi(g)\, dg, \text{ and}$$

$$\sigma_i(\lambda_i) = \sqrt{\int_{z=-\infty}^{+\infty} z^2 h_{\varepsilon_i}(z)\, dz - \left[\mu_i(\lambda_i)\right]^2} = \sqrt{\left(\int_{g=-\infty}^{+\infty} \left[t_{\lambda_i}^{-1}(g)\right]^2 \phi(g)\, dg\right) - \left[\mu_i(\lambda_i)\right]^2}. \tag{4}$$

These quantities may be estimated quickly for given λ_i using Gauss-Hermite quadrature. Then, for future use, and to ensure that the inverse YJ transformation parameter λ_i is estimated solely based on shape (that is, skew and kurtosis), and does not confound location/scale with shape, we define a standardized error term as follows:

$$\varsigma_i = \frac{\varepsilon_i - \mu_i(\lambda_i)}{\sigma_i(\lambda_i)}. \tag{5}$$

To illustrate, Figure 1 provides the density for the standardized error term ς_i, for several values of λ_i alongside the traditional symmetric standard normal error distribution (which is equivalent to $\lambda_i = 1$). The inverse YJ transformed error terms ε_i can be tied together across different error terms (say $\varepsilon_1, \varepsilon_2, ...\varepsilon_i, ...\varepsilon_I$) into a multivariate distribution using an implicit Gaussian copula, given that the standard normal η_i marginals can be brought together in a multivariate normal distribution using a correlation matrix $\mathbf{R}$. The direction and intensity of skew/tail can vary across the random terms. Define $\boldsymbol{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, ..., \varepsilon_I\right)'$, $\boldsymbol{\eta} = \left(\eta_1, \eta_2, ..., \eta_I\right)'$, $\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}} = \left[\lambda_1, \lambda_2, ..., \lambda_I\right]'$, $\boldsymbol{\varepsilon} = \mathbf{t}_{\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}}^{-1}(\boldsymbol{\eta}) = \left[t_{\lambda_1}^{-1}(\eta_1), t_{\lambda_2}^{-1}(\eta_2), ..., t_{\lambda_I}^{-1}(\eta_I)\right]$, and so $\boldsymbol{\eta} = \mathbf{t}_{\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}}(\boldsymbol{\varepsilon})$. Also, define $\boldsymbol{\varsigma} = \left(\varsigma_1, \varsigma_2, ..., \varsigma_I\right)'$, and $\boldsymbol{\mu}(\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}) = \left[\mu_1(\lambda_1), \mu_2(\lambda_2), ..., \mu_I(\lambda_I)\right]'$. All of the above vectors are $(I \times 1)$ column vectors. Also, for later use, define $\boldsymbol{\Sigma}(\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}) = \begin{bmatrix} \sigma_1(\lambda_1) & 0 & 0 & 0 \\ 0 & \sigma_2(\lambda_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_I(\lambda_I) \end{bmatrix}$. Then, $\boldsymbol{\varsigma} = \left[\boldsymbol{\varepsilon} - \boldsymbol{\mu}(\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}})\right]\left(\boldsymbol{\Sigma}(\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}})\right)^{-1}$.

The cumulative distribution of the asymmetric error vector $\boldsymbol{\varepsilon}$ may then be written as:

$$H_{\boldsymbol{\varepsilon}}(\mathbf{z}) \;=\; \text{Prob}(\boldsymbol{\varepsilon} < \mathbf{z}) = \text{Prob}\left[\mathbf{t}_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{\eta}) < \mathbf{z}\right] = \text{Prob}\left[\boldsymbol{\eta} < \mathbf{t}_{\boldsymbol{\lambda}}(\mathbf{z})\right] = \Phi_I\left[\mathbf{g}, \mathbf{R}\right], \text{ with } \mathbf{g} = \mathbf{t}_{\boldsymbol{\lambda}}(\mathbf{z}) \qquad (6)$$

For later use, let $\boldsymbol{\eta}_{-i}$ be the $[(I-1)\times 1]$ column vector of $\boldsymbol{\eta}$ sans the *i*th element η_i, and similarly for other column vectors and let $\Phi_{I-1}(.;.)$ denote the multivariate cumulative normal distribution (MVNCD) function of dimension $(I-1)$. Define an $[(I-1)\times I]$ mask matrix $\mathbf{D}_i$ that holds an $[(I-1)\times(I-1)]$ identity matrix supplemented by an $[(I-1)\times 1]$ column of zeros in the *i*th column and another $(I\times 1)$ matrix (column vector) $\mathbf{G}_i$ with zero elements in each row except an entry of '1' in the *i*th row. Then, using the conditional properties of the multivariate normal distribution, we may write:

$$\boldsymbol{\eta}_{-i} \,\big|\, (\eta_i = g_i) \sim MVN_{I-1}[\mathbf{B}_{-i}(g_i), \mathbf{C}_{-ii}],$$
$$\text{where } \mathbf{B}_{-i}(g_i) = \left[\mathbf{D}_i\mathbf{R}\mathbf{G}_i \times g_i\right] \text{ and } \mathbf{C}_{-ii} = \left[\mathbf{D}_i\mathbf{R}\mathbf{D}_i' - \left(\mathbf{D}_i\mathbf{R}\mathbf{G}_i\right)\left(\mathbf{D}_i\mathbf{R}\mathbf{G}_i\right)'\right] . \qquad (7)$$

From above, and defining $\boldsymbol{\omega}_{-ii}$ as a diagonal $[(I-1)\times(I-1)]$ matrix holding the square root of the diagonal elements of $\mathbf{C}_{-ii}$, we can also write the following conditional probability:

$$\text{Prob}\left[\left(\boldsymbol{\eta}_{-i} < \mathbf{g}_{-i}\right)\right) \big| \left(\eta_i = g_i\right)\right] = \Phi_{I-1}\left[\left[\left(\mathbf{g}_{-i} - \mathbf{B}_{-i}(g_i)\right)\boldsymbol{\omega}_{-ii}^{-1}\right]; \left(\boldsymbol{\omega}_{-ii}^{-1}\mathbf{C}_{-ii}\,\boldsymbol{\omega}_{-ii}^{-1}\right)\right]. \qquad (8)$$

2.2. The MNYJ Model with Fixed Coefficients

Consider a multinomial choice context, with the index *i* for the alternative $(i = 1, 2, ..., I)$ (we will suppress the index *q* for individuals at this point). The utility associated with alternative *i* is written as:

$$U_i = \mathbf{b}'\mathbf{x}_i + s_i\varsigma_i, \qquad (9)$$

where $\mathbf{x}_i$ is a $(E\times 1)$-column vector of exogenous attributes (including dummy variables for constants, except in one of the *I* alternative utilities, say the first alternative), $\mathbf{b}$ is a fixed $(E\times 1)$-column vector of coefficients, s_i is an alternative-specific scale parameter, and ς_i is a standardized asymmetric error form as defined in Equation (5). As in any discrete choice model, only utility differences matter, but now the differences correspond to nonlinear and asymmetric transformations of normal errors, which enables the estimation of alternative-specific scales and shapes for the structural errors, as well as a full correlation matrix for the structural errors. That is, the probability functions in the MNYJ model depend on more than a Gaussian difference covariance, and so we are able to work directly with the structural errors without needing to work in the differenced covariance space as in the MNP (this is an issue that has not been realized or discussed in any earlier literature, as far as we are aware). This has important interpretation benefits in addition to flexibility benefits. However, there is one normalization needed in the MNYJ model in relation to the structural scales because the multiplication of utilities by a common factor does not change the ordering of utilities. Thus, the scale s_i for one alternative needs to be normalized (say to one) and other scales would be estimated as relative scales. The choice of which alternative's utility to scale is innocuous, as the results are invariant to this choice (except that the scales of the alternatives and the $\mathbf{b}$ vector will be rescaled appropriately). Alternatively, the analyst

may also impose an across-alternatives scale constraint such as $\sum_{i=1}^{I} s_i^2 = 1$ to obtain a single set of coefficient estimates and scales. In this paper, we adopt the latter normalization (we will return to this issue later).

The probability of choice of alternative *i* may be written as:

$$\begin{aligned}
\text{Prob}(i) &= \text{Prob}\left[(U_i > U_j)\ \forall\, j \neq i\right] = \text{Prob}\left[(U_j - U_i < 0)\ \forall\, j \neq i\right] \\
&= \text{Prob}\left[\left[s_j\varsigma_j < \left(\mathbf{b}'\left[\mathbf{x}_i - \mathbf{x}_j\right] + s_i\varsigma_i\right)\right]\ \forall\, j \neq i\right] \\
&= \text{Prob}\left[\left[s_j\left(\frac{\varepsilon_j - \mu_j(\lambda_j)}{\sigma_j(\lambda_j)}\right) < \left(\mathbf{b}'\left[\mathbf{x}_i - \mathbf{x}_j\right] + s_i\left(\frac{\varepsilon_i - \mu_i(\lambda_i)}{\sigma_i(\lambda_i)}\right)\right)\right]\ \forall\, j \neq i\right] \\
&= \text{Prob}\left[\varepsilon_j < \left[\left\{\left(\mathbf{b}'\left[\mathbf{x}_i - \mathbf{x}_j\right] + s_i\left(\frac{\varepsilon_i - \mu_i(\lambda_i)}{\sigma_i(\lambda_i)}\right)\right)\frac{\sigma_j(\lambda_j)}{s_j}\right\} + \mu_j(\lambda_j)\right]\ \forall\, j \neq i\right] \\
&= \text{Prob}\left[\eta_j < \left(t_{\lambda_j}\left[\left\{\left(\mathbf{b}'\left[\mathbf{x}_i - \mathbf{x}_j\right] + s_i\left(\frac{t_{\lambda_i}^{-1}(\eta_i) - \mu_i(\lambda_i)}{\sigma_i(\lambda_i)}\right)\right)\frac{\sigma_j(\lambda_j)}{s_j}\right\} + \mu_j(\lambda_j)\right]\right)\ \forall\, j \neq i\right]
\end{aligned} \tag{10}$$

Next, conforming with the earlier notations, define the following:

$$g_j\left|(\eta_i = g_i) = \left(t_{\lambda_j}\left[\left\{\left(\mathbf{b}'\left[\mathbf{x}_i - \mathbf{x}_j\right] + s_i\left(\frac{t_{\lambda_i}^{-1}(g_i) - \mu_i(\lambda_i)}{\sigma_i(\lambda_i)}\right)\right)\frac{\sigma_j(\lambda_j)}{s_j}\right\} + \mu_j(\lambda_j)\right]\right)\ \forall\, j \neq i\right. \tag{11}$$

Collect all the $g_j\left|(\eta_i = g_i)\ \forall\, j \neq i\right.$ into a vector $\mathbf{g}_{-i}\left|\left(\eta_i = g_i\right)\right.$.

Then, we may write:

$$\begin{aligned}
\text{Prob}(i) &= \int_{g_i=-\infty}^{+\infty} \text{Prob}\left[\left(\boldsymbol{\eta}_{-i} < \mathbf{g}_{-i}\right)\right|\left(\eta_i = g_i\right)\left.\right]\phi(g_i)dg_i \\
&= \int_{g_i=-\infty}^{+\infty} \left\{\Phi_{I-1}\left[\left[\left(\mathbf{g}_{-i}\left|\left(\eta_i = g_i\right) - \mathbf{B}_{-i}(g_i)\right.\right)\boldsymbol{\omega}_{-ii}^{-1}\right]; \left(\boldsymbol{\omega}_{-ii}^{-1}\mathbf{C}_{-ii}\,\boldsymbol{\omega}_{-ii}^{-1}\right)\right]\right\}\phi(g_i)dg_i,
\end{aligned} \tag{12}$$

The above probability expression can be evaluated using Bhat's (2018) analytic approximation for the $(I-1)$-dimensional MVNCD function within the one-dimensional integral, followed by a Gauss-Hermite quadrature procedure to evaluate the one-dimensional integral over the standardized normal density function. A maximum likelihood estimation procedure is adopted for estimation of the parameter set $(\mathbf{b}', \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}', \text{Vech}(\mathbf{R}), \mathbf{s})'$, where Vech(.) vectorizes the upper diagonal elements (ignoring the diagonal elements) of the correlation matrix it operates on, and $\mathbf{s} = (s_1, s_2, ..., s_I)'$. To ensure that all elements of $\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}$ are within the bound of 0 to 2, we use the following elementwise parameterization for the vector $\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}}$:

$$\lambda_i = \frac{2}{1 + \exp(-\lambda_i^*)} \text{ for } i = 1, 2, ..., I. \tag{13}$$

Similarly, to honor the identification constraint that $\sum_{i=1}^{I} s_i^2 = 1$, we parameterize each s_i as:

$$s_i = \sqrt{\frac{\exp(s_i^*)}{\sum_{j=1}^{I} \exp(s_j^*)}} \text{ for } i = 1, 2, ..., I . \tag{14}$$

Once the model is estimated with the λ_i^* and s_i^* values, we run a final iteration with the implied values for the parameters λ_i and s_i .

A few other notes are in order here relating to the estimation procedure in the case of the MNYJ with fixed coefficients. As indicated earlier in Equation (4), for a given value of λ_i during the maximum likelihood iterations, $\mu_i(\lambda_i)$ and $\sigma_i(\lambda_i)$ may be computed using one-dimensional Gauss-Hermite quadrature. In our simulation runs discussed in Section 3, we used a 92-point quadrature that is still very fast and makes the standardization highly accurate. The issue with the quadrature evaluation of the outer integral (that is, the integral in Equation (12)) is more tricky. A sensitivity analysis was undertaken in our simulation runs with respect to the number of Gauss–Hermite quadrature nodes used for this outer one-dimensional integral evaluation. Specifically, values of 10, 20, 30, 50, and 92 quadrature points were examined. The results indicated that moving from 10 to 20 nodes produced a substantial improvement in parameter recovery, with the estimates becoming much closer to the true parameter values. Increasing the quadrature order further from 20 to 30 provided additional improvement, although the gains were noticeably smaller than those observed when moving from 10 to 20 nodes. Comparisons between 30 and 50 nodes revealed very little difference in the resulting parameter estimates whenever the 50-node specification successfully converged. However, occasional convergence failures began to emerge at 50 nodes. When the quadrature order was increased to 92 nodes, convergence failures became substantially more frequent, and the numerical stability of the covariance matrix computations deteriorated considerably. These findings suggest that, for the MNYJ model, increasing the outer integral quadrature order beyond a moderate level does not necessarily improve estimation performance. While higher-order quadrature rules provide a more accurate approximation to the underlying integral, they also evaluate the integrand more finely in the tails of the kernel distribution ε_i of the chosen alternative as contained in the nonlinear inverse Yeo–Johnson transformations $\varepsilon_i = t_{\lambda_i}^{-1}(\eta_i)$. This, together with multivariate normal cumulative distribution function evaluations embedded within the integrand itself, can expose regions of the likelihood surface with very weak curvature, sharper ridges, more local roughness, and overall increased numerical irregularity when higher-order quadrature is used, particularly in the correlation and shape-parameter directions. The net result is near-singular Hessians, unstable covariance matrix calculations, and convergence difficulties, even when lower-order quadrature rules already provide sufficiently accurate approximations to the likelihood function. Overall, the correlation block is numerically fragile under sharper outer quadrature (high quadrature nodes), and a lower quadrature order can provide a more stable approximation without materially changing the fitted likelihood. Based on the overall tradeoff between numerical stability, convergence behavior, and parameter recovery, a quadrature order of 30 was selected for the simulation analysis of Section 3 in the current paper.

2.3. The MNYJ Model with Random Coefficients

The MNYJ model can be extended in a straightforward manner to include asymmetric (inverse YJ transformation based) random coefficients on exogenous variables. Further, we can also relax the assumption that random coefficients are distributed independent of the kernel error terms (or independent of any error components if error components are the vehicle to engender correlations across utilities).[1] In this regard, our model not only extends the traditional mixed logit and MNP models to asymmetric kernel error terms with a full covariance matrix, but also allows asymmetric random coefficients as well as correlation effects between the random coefficient vector and the kernel error vector.

The random coefficients formulation begins with the equation below:

$$U_i = \mathbf{b}'\mathbf{x}_i + \upsilon\tilde{\boldsymbol{\beta}}'\mathbf{x}_i + s_i\varsigma_i = \mathbf{b}'\mathbf{x}_i + \upsilon\left[\boldsymbol{\beta} - \left(\boldsymbol{\tau}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})\right)\right]\left[\boldsymbol{\theta}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})\right]^{-1}\mathbf{x}_i + s_i\varsigma_i, \text{ with } \tilde{\boldsymbol{\beta}} = \left[\boldsymbol{\beta} - \left(\boldsymbol{\tau}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})\right)\right]\left[\boldsymbol{\theta}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})\right]^{-1}, \quad (15)$$

with υ being a $(E \times E)$ diagonal matrix containing the scale associated with each element of $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta} = \mathbf{t}^{-1}_{\tilde{\lambda}_{\boldsymbol{\beta}}}(\boldsymbol{\gamma})$ being an inverse YJ-transformed asymmetric random coefficient ($\boldsymbol{\gamma}$ is a standardized normal vector), $\left(\boldsymbol{\tau}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})\right)$ being a vector of means due purely to the inverse YJ transformation, and $\boldsymbol{\theta}(\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}})$ being the diagonal standard deviation matrix due to the inverse YJ transformation. As in the previous section, the use of $\tilde{\boldsymbol{\beta}}$ rather than $\boldsymbol{\beta}$ in the utility function above ensures that the inverse YJ transformation parameter vector $\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\beta}}$ solely captures shape.[2] Next, let the correlation matrix of $\boldsymbol{\delta} = \left(\boldsymbol{\gamma}', \boldsymbol{\eta}'\right)'$ be $\boldsymbol{\Theta}$ (an $[(E+I)\times(E+I)]$ matrix). Define the following:

- An $[(I-1)\times(E+I)]$ mask matrix $\tilde{\mathbf{D}}_i$ that holds zeros in all rows of the first E columns, followed by an $[(I-1)\times(I-1)]$ identity matrix supplemented by a column of zeros in the $(E+i)^{th}$ column,
- An $[(E+I)\times(E+1)]$ matrix $\tilde{\mathbf{G}}_i$ with an identity matrix in the first E rows and first E columns, an entry of '1' in the $(E+i)^{th}$ row and last column, and zero elsewhere.

[1]There is nothing from a theoretical standpoint that prevents non-kernel random coefficients from being correlated with kernel error terms, although the assumption of independence between these two stochastic components is almost universally invoked in the earlier mixing literature. This convention arises primarily from computational and identification considerations rather than from any behavioral necessity. Allowing such correlations substantially increases the complexity of the likelihood function and can create numerical instability in simulation-based estimation. Conceptually, random coefficients capture unobserved heterogeneity in sensitivities to observed explanatory variables, while kernel error terms capture residual utility variation associated with omitted attributes, latent perceptions, contextual influences, measurement error, and other unobserved determinants of utility. While these two forms of stochasticity are not fundamentally separable from a behavioral standpoint, maintaining independence between them provides a useful and tractable decomposition between unobserved taste heterogeneity in observed-variable sensitivities and residual unobserved utility variation.

[2]Note that Equation (15) assumes an inverse YJ transformation for the random coefficients, which implies an unbounded coefficient. If the analyst desires a strictly bounded coefficient such as strictly positive or strictly negative for specific coefficients, log-normal specifications can be introduced in the usual way while allowing for inverse YJ-transformed coefficients on other variables.

$$\text{Let } \tilde{\mathbf{A}}_i = \left(\tilde{\mathbf{G}}_i'\boldsymbol{\Theta}\tilde{\mathbf{G}}_i\right),\ \tilde{\mathbf{B}}_{-i}(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i) = \left[\left(\tilde{\mathbf{D}}_i\boldsymbol{\Theta}\tilde{\mathbf{G}}_i\right)\tilde{\mathbf{A}}_i^{-1}(\tilde{\mathbf{g}}',g_i)'\right],$$
$$\tilde{\mathbf{C}}_{-ii} = \left[\left(\tilde{\mathbf{D}}_i\boldsymbol{\Theta}\tilde{\mathbf{D}}_i'\right) - \left(\tilde{\mathbf{D}}_i\boldsymbol{\Theta}\tilde{\mathbf{G}}_i\right)\tilde{\mathbf{A}}_i^{-1}\left(\tilde{\mathbf{D}}_i\boldsymbol{\Theta}\tilde{\mathbf{G}}_i\right)'\right], \text{ and} \tag{16}$$

define $\tilde{\boldsymbol{\omega}}_{-ii}$ as a diagonal $[(I-1)\times(I-1)]$ matrix holding the square root of the diagonal elements of $\tilde{\mathbf{C}}_{-ii}$.

Define the following:

$$g_j\left|\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right. = \left(t_{\lambda_j}\left[\left\{\left(\left(\mathbf{b}+\boldsymbol{\upsilon}\left\{\left[\mathbf{t}_{\tilde{\lambda}_\beta}^{-I}(\tilde{\mathbf{g}})-\left(\boldsymbol{\tau}(\tilde{\boldsymbol{\lambda}}_\beta)\right)\right]\left[\boldsymbol{\theta}(\tilde{\boldsymbol{\lambda}}_\beta)\right]^{-1}\right\}\right)'\left[\mathbf{x}_i-\mathbf{x}_j\right]+s_i\left(\frac{t_{\lambda_i}^{-1}(g_i)-\mu_i(\lambda_i)}{\sigma_i(\lambda_i)}\right)\right)\frac{\sigma_j(\lambda_j)}{s_j}\right\}+\mu_j(\lambda_j)\right]\right)\ \forall\, j\neq i \tag{17}$$

Collect all the $g_j\left|\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right.\ \forall\, j\neq i$ into a vector $\mathbf{g}_{-i}\left|\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right.$.

Then we may write the probability of choice of alternative *i* as:

$$\begin{aligned}\text{Prob}(i) &= \int_{\tilde{\mathbf{g}}=-\infty}^{\infty}\int_{g_i=-\infty}^{+\infty}\text{Prob}\left[\left(\boldsymbol{\eta}_{-i}<\mathbf{g}_{-i}\right)\right|\left.\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right]\left(\phi_{E+1}\left[(\tilde{\mathbf{g}}',g_i)',\mathbf{A}_i\right]\right)dg_i\,d\tilde{\mathbf{g}}\\ &= \int_{\tilde{\mathbf{g}}=-\infty}^{\infty}\int_{g_i=-\infty}^{+\infty}\left\{\Phi_{I-1}\left[\left[\left\{\mathbf{g}_{-i}\left|\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right.-\tilde{\mathbf{B}}_{-i}\left(\boldsymbol{\gamma}=\tilde{\mathbf{g}},\eta_i=g_i\right)\right.\right\}\tilde{\boldsymbol{\omega}}_{-ii}^{-1}\right];\left(\tilde{\boldsymbol{\omega}}_{-ii}^{-1}\tilde{\mathbf{C}}_{-ii}\,\tilde{\boldsymbol{\omega}}_{-ii}^{-1}\right)\right]\right\}\left(\phi_{E+1}\left[(\tilde{\mathbf{g}}',g_i)',\mathbf{A}_i\right]\right)dg_i\,d\tilde{\mathbf{g}},\end{aligned} \tag{18}$$

where $\phi_{E+1}[.,.]$ represents the (*E*+1)-dimensional multivariate normal density function. The integration of the (*E*+1) infinite integral in Equation (18) over the multivariate normal density function can be undertaken using Halton sequences as proposed by Bhat (1999), and subsequently published as Bhat (2001) and disseminated by Train (1999) with explicit acknowledgment of Bhat (1999) as the source origin. A maximum likelihood estimation procedure is again adopted for estimation of the parameter set $(\mathbf{b}',\boldsymbol{\lambda}_\beta',\text{diag}(\boldsymbol{\upsilon}),\boldsymbol{\lambda}_\varepsilon',\text{Vech}(\boldsymbol{\Theta}),\mathbf{s}')'$, with diag(.) being an operator that extracts the diagonal elements of the matrix it operates on. Extensive testing during the empirical analysis suggested that 125 Halton draws were adequate, with the parameter estimates stabilizing and exhibiting little variation beyond this number of draws. At the same time, stable convergence for all our simulation runs was obtained at this order of the number of Halton draws.

2.4. MNYJ Model Identification

2.4.1 The Multinomial (I ≥ 3) Case

To examine identification issues, we start with the independent MNYJ model (with the correlation matrix $\mathbf{R}$ underlying the vector $\boldsymbol{\eta}$ being the identity matrix) and with fixed coefficients. As in any discrete choice model, only utility differences matter. Taking differences with respect to the first alternative in Equation (9), let

$$\tilde{U}_i = U_i - U_1 = V_i - V_1 + \varpi_i, \text{ where } V_i = \mathbf{b}'\mathbf{x}_i \text{ and } \varpi_i = s_i\varsigma_i - s_1\varsigma_1. \tag{19}$$

The usual location normalization is needed (so that one alternative needs to be the base for any alternative-invariant exogenous variable as well as for the constant). Similarly, a global scale normalization is needed because multiplying the utilities by any positive constant does not change

the relative utility ordering, which is the reason for the normalization $\sum_{i=1}^{I} s_i^2 = 1$. Then, we get the following:

$$Var(\varpi_i) = s_i^2 - s_1^2, \text{ and } \text{cov}(\varpi_i, \varpi_j) = s_1^2,\ i \neq 1, j \neq 1. \tag{20}$$

Thus, the off-diagonal elements of the differenced covariance matrix identify s_1^2, which leaves the (I–1) structural variances s_i^2 $(i \neq 1)$. When $I \geq 3,$ all of these can be uniquely determined from the (I–1) diagonals of the utility-differenced covariance matrix up to the global scale normalization.[3] Next, define the third marginal cumulant (skewness measure) of the transformed errors as $\kappa_i = E[\tilde{\varepsilon}_i^3]$. The third marginal cumulants of the differenced error terms satisfy:

$$E[\varpi_i^3] = s_i^3 \kappa_i - s_1^3 \kappa_1 \tag{21}$$

The expressions for the mixed third cumulant with one index repeated as well as the fully mixed third cumulant are given by an identical expression as follows:

$$E[\varpi_i^2, \varpi_j] = -s_1^3 \kappa_1,\ i \neq j; E[\varpi_i, \varpi_j, \varpi_j] = -s_1^3 \kappa_1, \quad i \neq j \neq k\ . \tag{22}$$

The above result emerges because under independence, all mixed cumulants vanish unless all three factors come from the same underlying structural shock. The only shock common to every differenced utility is the base shock ς_1. So every off-diagonal third cumulant isolates the base-alternative asymmetry, from which κ_1 gets identified. Then, from Equation (21), each κ_i $(i \neq 1)$ gets identified from the marginal third cumulant of the differenced error terms. Thus, all third cumulants κ_i are identified when $I \geq 3$. Then, since the mapping from the YJ parameter λ_i to the shape distribution of $t_{\lambda_i}(\eta_i)$ is locally one-to-one (because of the normalization in Equation (5)), the λ_i parameters of each alternative are generically identified. That is, the joint distribution of utility differences contains mixed higher-order moments/cumulants that can separate the base alternative's asymmetry from the other alternatives' asymmetries.

[3]An important point here. Even in the independent (but non-identical) MNP, one can technically estimate the structural variances up to the global scale normalization. The problem though is that, in the MNP, the differenced kernel error distribution of ϖ_i is also Gaussian (multivariate normal), which is fully characterized by the second moments (that is, the differenced covariance matrix). As Bunch (1991) cautions, if one estimates the differenced covariance matrix, this preferred covariance matrix may not correspond to a valid independent structural covariance system with strictly positive structural variances. But this situation can be circumvented in the independent MNP by starting off with a heteroscedastic (independent) structural error covariance matrix maintaining strict positivity of each variance term (up to a global scale normalization), as we propose in our formulation that also immediately accommodates the global scale normalization. However, the net result in estimation can be a much worse model fit than what would be obtained if one simply focused on the differenced error covariance matrix (because the implied differenced covariance matrix cannot span the full admissible covariance space, and the MNP has no additional means to compensate for the resulting poor fit). The independent MNYJ model inherits the same second-moment restrictions under structural independence. However, unlike MNP, the MNYJ specification introduces additional flexibility through alternative-specific asymmetry and the resulting higher-order moment structure. Consequently, substitution patterns that would require flexible covariance in the Gaussian MNP framework may instead be partially accommodated through asymmetric marginal behavior, making the independent MNYJ model behaviorally richer and less restrictive than the independent MNP despite the same diagonal covariance restriction. In this context, though, the real benefit of the MNYJ is in the case of the non-diagonal structural covariance matrix specification, as discussed later.

In summary, in the independent case and for $I \geq 3$, the joint distribution of utility differences contains sufficient information to recover all the structural scale parameters s_i $(i = 1, 2, ..., I)$ and the YJ shape parameters λ_i $(i = 1, 2, ..., I)$ after imposing one location normalization and one global scale normalization. This identification is generic, in the sense that it holds for almost all parameter values except in special cases, most notably when all the λ_i terms are uniformly equal to one (when the independent MNYJ collapses to the independent MNP). Even beyond the case of all the λ_i parameters being uniformly equal to one, empirical identification can become tenuous under other circumstances. In particular, if all the λ_i parameters approach the symmetric Gaussian case (even if not exactly all equal to one), the identifying information from asymmetry diminishes. In this region, the model becomes close to the MNP, and the separation between the second and third moments of the error terms becomes weak. As a result, estimation may exhibit large variability in the shape parameters, even though the model remains generically identified. Similarly, the likelihood can become nearly flat if one or more alternatives have very low market shares or have kernel error variances that are relatively miniscule. In the first case of low market shares, the deterministic component would dominate too strongly, pushing the probability of some alternatives close to one and other alternatives close to zero, making many observations become nearly deterministic and reducing the information available for identifying the scale, shape, and dependence structure of the kernel errors. In the second case, the shape contribution can become so tiny as to effectively provide little asymmetry to aid in the identification of the shape parameters. Another instance when identification can become fragile is if the shape parameters (after standardization as introduced in our model) and scale parameters of two or more alternatives are very close to one another. In this instance, the difference in utilities of these two alternatives becomes close to being symmetric, which can cause estimation instability. In all of these cases, the situation may be diagnosed by examining the local behavior of the likelihood function in the neighborhood of the optimum. In particular, profile likelihood plots for individual parameters provide a useful diagnostic tool: flat or weakly curved profile likelihoods indicate that the likelihood changes very little over a broad range of parameter values, suggesting weak identification or limited information in the data for those parameters.

In the more interesting and much richer specification of non-independent kernel errors with a full copula-based structural correlation structure for **R**, the same broad logic as above survives, but the algebra is less clean because mixed moments now also contain correlation contributions. Still, asymmetry gives information beyond second moments, so the joint law of the differenced utilities is no longer summarized by a Gaussian covariance matrix. This allows the generic identification of the structural correlation matrix rather than only the covariance of differenced errors. In essence, the combination of alternative-specific nonlinear transformations and Gaussian copula dependence generates a joint distribution of utility differences that depends on higher-order features beyond covariance, enabling generic identification for $I \geq 3$. Of course, weak identification may be achieved under any of the conditions already discussed for the independent case. In the non-independent case, the identification situation can become further fragile exactly because marginal asymmetry (along with marginal scale of the kernel errors) and the dependence among the kernel errors all jointly shape the distribution of utility differences. Thus, when two or more alternatives have very similar standardized shape parameters, their marginal distributions exhibit similar skewness and tail behavior, making it more difficult for the likelihood function to distinguish between dependence arising from the correlation structure and similarity arising from nearly identical marginal asymmetry. As a result, correlation and shape effects may become

partially substitutable, leading to weak local curvature of the likelihood surface, instability in the estimated covariance matrix, and inflated standard errors for both the correlation and shape parameters. Distinct alternative-specific shape parameters help mitigate this issue by generating more differentiated higher-order behavior across alternatives, thereby improving empirical identification of the correlation structure.

Finally, the introduction of random coefficients does not introduce any substantial additional complications, especially for continuous random variables that vary across alternatives and individuals.

2.4.2 The Binary (I = 2) Case

For the binary case, we discuss identification only for the independent case because the resulting model does not allow the separate identification of λ parameters and scales. In particular, with only two alternatives, only a single utility difference is observed: $\varpi_2 = s_2\tilde{\varepsilon}_2 - s_1\tilde{\varepsilon}_1$. Then, $Var(\varpi_2) = s_1^2 + s_2^2$ and $E[\varpi_2^3] = s_2^3\kappa_2 - s_1^3\kappa_1$. These give us two equations, but, even with the global scale normalization of $s_1^2 + s_2^2 = 1$, there are three remaining unknowns (κ_1, κ_2, and one of the scale parameters) that cannot be estimated from two equations. Also, while higher-order moments such as the fourth and fifth moments add information in the binary case, they are functions of a single scalar utility difference and do not fully resolve the decomposition into alternative-specific components without additional restrictions. That is, the mapping from structural parameters to the distribution of the utility difference is not one-to-one. Specifically, distinct parameter triples (κ_1, κ_2, and one of the scale parameters) can generate identical third and fourth moments of the utility difference. This arises because the inverse Yeo–Johnson transformation induces a symmetry in the higher-order moments around the Gaussian benchmark. For instance, shape parameters equidistant from unity produce opposite skewness but identical kurtosis after standardization. By appropriately reallocating scale across the two alternatives, it is possible to offset these changes in skewness and preserve the overall distribution of the utility difference. Thus, even when higher-order moments such as skewness and kurtosis are taken into account, the binary case does not provide sufficient information to uniquely recover alternative-specific scale and shape parameters. At the same time, in this binary case, imposing equality of the shape parameters across alternatives does not resolve the identification problem. When the kernel error distributions are identical across alternatives, the resulting utility difference is symmetrically distributed, even if the underlying marginal distributions are asymmetric. As a result, all odd moments (such as the third moment) of the utility difference vanish, eliminating the primary source of identifying information for the shape parameters. Consequently, the model collapses to a symmetric error structure, and the common shape parameter is not identified. Effectively then, in the binary case, both the cases of having separate shape parameters across both alternatives as well as having equal shape parameters across both alternatives are not identified.

A coherent and invariant approach in the binary setting is to parameterize the distribution of the utility difference directly, using a single asymmetry parameter with a normalized scale. Under this reduced-form representation, reversing the direction of differencing simply changes the sign of the asymmetry parameter, preserving observational equivalence. However, this reduced-form asymmetry parameter is not equivalent to alternative-specific structural shape parameters. Thus, while binary choice data can identify the asymmetry of the utility difference, they do not provide sufficient information to separately recover the alternative-specific kernel-error shape or

scale parameters. This contrasts with the multinomial case, where cross-alternative differences generate additional moment conditions to estimate all scales and all shape parameters up to the global scaling normalization.

3. SIMULATION ANALYSIS

The simulation analysis undertaken here is centered around two objectives. The first is to empirically demonstrate the identifiability of the (a) inverse YJ-transformed kernel structural error distributions, (b) the full correlation matrix for the structural error terms, and (c) the ability of the estimation process to recover all parameters in the MNYJ model. The second is to illustrate the problems that may arise from ignoring skew/heavy-tailed kernel error terms and how these may incorrectly manifest themselves in restrictive models.

3.1. Experimental Set-Up

The MNYJ model is versatile, and the recent analytic-based evaluation of the MVNCD function (Bhat, 2018), combined with the use of Gaussian quadrature or Halton sequences for the mixing integration (Bhat, 1999, 2001), makes it feasible to estimate such a model. In this simulation, to focus attention on the YJ kernel errors, we consider the case without random coefficients, because the case with random coefficients is a form of mixing that has already been studied substantially in the literature. In particular, we consider three alternatives and ignore constants in the utilities, with the focus being on the parameters on the independent variables and the kernel error term distributions. Two exogenous variables are considered, the first being a continuous variable that varies across alternatives and the second being a dummy variable that is alternative-invariant. For the first independent variable, the values for the three alternatives are generated from univariate normal distributions as follows: $x_{i1} \sim N(\mu_i, \sigma_i^2)$, $i = 1,2,3$, with $(\mu_1, \mu_2, \mu_3) = (0.0, 0.5, 1.0)$ and $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1.5^2, 1.25^2, 1.0^2)$. For the second dummy variable, independent values are drawn from the standard uniform distribution. If the value drawn is less than 0.5, the value of '0' is assigned for the dummy variable. Otherwise, the value of '1' is assigned. A sample of 6,000 realizations of the exogenous variables is generated corresponding to 6,000 individuals. Once generated, these independent variable values are held fixed in the rest of the simulation exercise.

The $\mathbf{b} = (b_1, b_2, b_3)$ vector is specified as $\mathbf{b} = (-0.50, 0.25, 0.50)$, with $b_1 = -0.5$ being the generic coefficient on the continuous exogenous variable, and the remaining elements of $\mathbf{b}$ representing the alternative-specific coefficients on the alternative-invariant dummy exogenous variable for the second and third alternatives (with the coefficient on the first alternative set to zero for identification). The YJ parameter vector for the kernel error terms is set as $\boldsymbol{\lambda}_{\varepsilon} = (\lambda_1, \lambda_2, \lambda_3) = (0.25, 0.55, 1.45)$. This configuration provides two alternatives with right-skewed kernel errors and one with a left-skewed kernel error, all sufficiently distinct from the Gaussian benchmark. The implicit Gaussian copula correlation matrix $\mathbf{R}$ is a 3×3 matrix, which we assume as follows:

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.35 & 0.20 \\ 0.35 & 1.00 & 0.30 \\ 0.20 & 0.30 & 1.00 \end{bmatrix} \tag{23}$$

The upper diagonal elements of the above matrix are the ones to be estimated: $\text{Vech}(\mathbf{R}) = \{R_{12} = 0.35, R_{13} = 0.20, R_{23} = 0.30\}$. Finally specify the vector

$\mathbf{s} = (s_1 = 0.792, s_2 = 0.500, s_3 = 0.350)$, so that $\sum_{i=1}^{I} s_i^2 = 1$. The parameter set to be estimated is the vector $(\mathbf{b}', \boldsymbol{\lambda}_\varepsilon', \text{Vech}(\boldsymbol{\Theta}), \mathbf{s}')'$, which contains 12 parameters.

3.2 Data Generation and Performance Metrics

The simulation experiments assume underlying "true" values for the many parameters as just discussed in the previous section, followed by the generation of data sets for estimation. In particular, using the underlying parameters, first compute $\mathbf{xb}$. Next draw Q =6,000 independent sets of draws from the standard multivariate normal distribution with correlation matrix $\mathbf{R}$ for $\boldsymbol{\eta}$. For each set of draws (corresponding to an individual), compute the kernel error realization vector $\boldsymbol{\varepsilon} = \mathbf{t}_{\lambda_\varepsilon}^{-1}(\boldsymbol{\eta})$. From these, one can compute $\boldsymbol{\varsigma} = \left[\boldsymbol{\varepsilon} - \boldsymbol{\mu}(\boldsymbol{\lambda}_\varepsilon)\right]\left(\boldsymbol{\Sigma}(\boldsymbol{\lambda}_\varepsilon)\right)^{-1}$, and further compute the utility for each alternative using Equation (5). Finally, for each of the 6,000 observation units, select the alternative with the highest utility and declare that as the chosen alternative.

The data generation procedure above is repeated 250 times with different realizations for $\boldsymbol{\eta}$ to generate 250 different data sets. The proposed MNYJ model is estimated on each of these different data sets, as is the traditional MNP model with normally distributed kernel error terms. A number of different measures of performance are computed from the estimation results to assess the ability of the models to recover the "true" parameters and their standard errors. This assessment procedure is as follows:

(1) Estimate model parameters and standard errors for each of the 250 data sets generated using each of the two model specifications (for a total of 500 estimations).

(2) Compute the mean (across the 250 datasets) estimate for each model parameter and compute the **absolute percentage (finite sample) bias** (APB) of the estimator for each model specification:

$$APB(\%) = \left|\frac{\text{mean estimate - true value}}{\text{true value}}\right| \times 100. \tag{24}$$

(3) Similarly, for each model specification, compute the standard deviation (across the 250 data sets) for each model parameter as the **finite sample standard deviation or FSSD** (essentially, this is the empirical standard error). Also, compute the median standard error for each model parameter across the data sets and label this as **the asymptotic standard error or ASE (the standard error for each run is computed using the sandwich covariance matrix)**.

(4) For each model specification, compute the APB associated with the ASE of the estimator as:

$$APBASE(\%) = \left|\frac{\text{ASE-FSSD}}{\text{FSSD}}\right| \times 100. \tag{25}$$

(5) For each parameter, compute the **coverage probability (CP)**, which is the empirical probability that a confidence interval contains the true parameter, as below:

$$\text{CP} = \frac{1}{N}\sum_{r=1}^{N} I\left[\hat{\beta}_X^r - t_\alpha * \text{se}(\hat{\beta}_X^r) \le \beta_X \le \hat{\beta}_X^r + t_\alpha * \text{se}(\hat{\beta}_X^r)\right] \tag{26}$$

where, CP is the coverage probability, $\hat{\beta}_X^r$ is the estimated value of the parameter in dataset *r* and $\text{se}(\hat{\beta}_X^r)$ is the estimated standard error for the same parameter, β_X is the true value of the parameter, $I[.]$ is an indicator function which takes a value of 1 if the argument in the bracket is true (otherwise 0), and t_α is the t-statistic value for a given confidence level $(1-\alpha)\times 100$. We compute CP values for 95% nominal coverage probability ($\alpha = 0.05$).

(6) Compare the performance of the restrictive MNP model formulation with the proposed model using a likelihood ratio (LR) test. The LR statistic is computed for each data set separately, and compared with the chi-squared table value with the appropriate degrees of freedom. Then, the number of times (out of the 250 estimations corresponding to the 250 data sets) that the LR value rejects the MNP model in favor of the proposed MNYJ model is calculated, along with the mean likelihood value across the 250 estimations for the MNP and MNYJ models.

3.3. Simulation Results

3.3.1 Ability of MNYJ to Recover Model Parameters

Table 1 presents the simulation results. We first focus on demonstrating the identifiability of all parameters in the MNYJ model, as shown in the upper panel of Table 1. Overall, the results indicate excellent finite-sample performance of the MNYJ model in recovering all categories of parameters, including the main coefficients, correlation terms, scale parameters, and YJ transformation parameters characterizing structural kernel error asymmetry. The mean parameter estimates are uniformly very close to their true values across all parameters. In particular, the absolute percentage biases (APB) (see third row of the upper panel of Table 1) are generally below 1%, with the largest APB being 3.465% for one of the correlation-related parameters. The coverage probabilities (CPs) in the fourth row provide another assessment of the distribution of the parameter estimates around the true values based on the probability that the true values are contained within the 95% confidence interval. As shown in Table 1, the 95% CPs for the MNYJ model are consistently close to the 95% target, with all estimated CPs falling between 90% and 97%.

The precision of the MNYJ estimator is also notable. The finite sample standard deviations (FSSD) are small across all parameters, and the corresponding asymptotic standard errors (ASE) closely track the empirical FSSDs. This is further confirmed by the absolute percentage bias of the ASE (APBASE), which remains modest for most parameters. While a few parameters exhibit higher APBASE values, these remain within reasonable bounds and do not materially affect inference.

Taken together, these results clearly demonstrate that the MNYJ model is able to accurately and precisely recover the true parameters of the data-generating process, including those relating to the asymmetric kernel error distributions.

3.3.2 Effect of Ignoring Asymmetry in the Random Coefficient and/or Kernel Distribution

The lower panel of Table 1 reports the corresponding results for the traditional MNP model, which restricts the kernel error terms to follow a multivariate normal distribution. Note that there are two fewer correlation parameters and one fewer scale, as only the differenced covariance matrix of the kernel errors is estimated. Further, while these estimated MNP values for the differenced covariance matrix are reported, there are no corresponding "true values" with which to compare them, because the data generation was based on the Gaussian copula correlation matrix **R** and the scale values for the non-Gaussian structural kernel error terms in the MNYJ model. Thus, unlike

the main coefficients, the estimated differenced covariance matrix parameters in the MNP model do not have direct counterparts in the data-generating process. Consequently, the focus of the comparison between the MNYJ and MNP models is on the main coefficients.

The results clearly show the poor performance of the MNP model compared with the MNYJ model. The MNP model displays substantial bias in several key parameters. APB values for the main coefficients all exceed 5%, with one exceeding 24%. These high levels of bias sharply contrast with the negligible biases observed under the correctly specified MNYJ model. Similar poor performance may be observed based on the CPs, with 95% CPs for the MNP model dropping to as low as 1.6%. Thus, while the MNP model is somewhat more precise (with generally smaller FSSD and ASE values), the bias in the estimates is clear, with the estimates generally outside the 95% confidence interval of the true parameters. Overall, the use of a normal distribution for the kernel errors when there is asymmetry in the kernel parameters can lead to seriously mis-estimated distributions and poor estimates of true model parameters.

Regarding model fit, the MNYJ model also consistently outperforms the MNP model. The mean log-likelihood across the 250 datasets is lower for the MNP model than for the MNYJ model. Furthermore, the likelihood ratio (LR) tests overwhelmingly favor the proposed MNYJ model, with the restrictive MNP specification rejected in 100% of the simulation runs. Overall, these findings clearly demonstrate that ignoring asymmetry in the kernel error distribution leads to distorted estimates, unreliable inference, and inferior model fit. In contrast, the MNYJ framework provides a flexible and robust modeling structure that successfully captures both correlation and distributional asymmetries, leading to substantial gains in estimation accuracy and inference reliability.

4. MODEL APPLICATION DEMONSTRATION

To demonstrate the practical application and empirical importance of adopting the MNYJ model, we apply the model to analyze future vehicle ownership decisions and compare the model with several restricted versions. For this application we also include two random coefficients, as discussed later.

4.1 Sample Data

The data for this analysis is drawn from the TOMNET Transformative Transportation Technologies (T4) Survey. The overarching objective of the T4 survey is to understand how emerging mobility technologies, particularly autonomous vehicles (AVs) and mobility-on-demand services such as ridehailing, may reshape individual travel behavior, vehicle ownership decisions, and long-term transportation outcomes. The full deployment of the T4 survey was conducted in late 2019 across four major metropolitan regions in the southern United States: Phoenix (AZ), Austin (TX), Atlanta (GA), and Tampa (FL). Data collection relied on a combination of email and mail-based recruitment using random address-based sampling as well as Facebook advertisements (for a complete description of the survey and sampling approach, see Khoeini et al., 2018). The final combined sample across all four metropolitan regions consisted of 3,465 responses. After cleaning and removing obviously erroneous responses, a total of 3,202 individuals were retained for the analysis.

Using the T4 survey data we focus on a stated preference question about future intentions regarding the adoption/use of automated vehicles (AVs). Respondents were asked about their next car purchase decision in a scenario where AVs were available for purchase or to use via ridehailing

services. To provide additional market environment context, respondents were asked to assume that half of the vehicles on the streets were AVs. The three alternatives presented were:

1. Buy a regular vehicle
2. Buy an AV
3. Don't buy a vehicle and use AV ridehailing/rental services

Alternative attributes included a fixed monthly cost for the purchase of a regular vehicle (ranging from $200 to $500) and the purchase of an AV (ranging from $150 to $625), a variable per-mile usage cost for all three options (ranging from $0.25 to $0.75 for each purchased option and from $1.50 to $3.00 for the ridehailing option) and an average waiting time for ridehailing (ranging from 3 to 9 minutes), each of which was varied based on a random block design (see Khoeini et al., 2019 for additional details regarding the stated preference question design).[4] Participants were then asked to provide their preference from among the three alternatives in two separate scenarios. For the sake of simplicity in the context of the proposed cross-sectional YJ model, we consider only the first scenario presented to each individual.

Based on the attribute levels presented, 55.7% of respondents indicated that they would buy a regular vehicle, 32.6% indicated that they would buy an AV, and 11.7% indicated that they would use a AV ridehailing service. These aggregate results indicate that there is a good deal of skepticism still about autonomous vehicles, even for a future situation where they make up a good portion of vehicles on the road. The following section examines heterogeneity in such preferences based on the MNYJ model in comparison to several restricted models.

4.2 Model Results

Four different models were estimated in the current analysis: (1) the standard multinomial probit (MNP) model that assumes symmetric and non-skewed normal kernel error distributions and no random coefficients, (2) the multinomial probit model with random coefficients (or MNPRC), (3) the multinomial Yeo-Johnson (MNYJ) model that allows for skewed kernel error distributions but no random coefficients, and (4) the MNYJ model with random cost coefficients (MNYJRC).

The emphasis of our empirical efforts in this paper is to demonstrate the application of the proposed model rather than necessarily on substantive interpretations and policy implications. But, within the context of the data available, the exogenous variable specification developed was based on considering a wide range of exogenous variable combinations and including categorical variables in their most disaggregate form first before progressively combining categories as appropriate based on statistical tests. Also, to compare and contrast the effects of exogenous variable across the utility distributional forms, we include, in all four models, any variable that was found to be statistically significant at the 90% confidence level in at least one of the four models. This also allows us to employ likelihood ratio tests to compare each restricted model with the proposed MNYJRC model.

The model estimation results for each of the four models are presented in Table 2. The remainder of this section discusses the model results, focusing on the comparison of the results across the four models. Note that caution needs to be exercised when comparing the magnitudes of any variable coefficient across the different model coefficients, because the error distributions

[4]Variable costs in the context of conventional non-AV vehicles typically refer to usage-dependent operating expenses such as fuel, maintenance, repair, insurance, parking, and toll costs. In the context of future owned AVs, variable costs may additionally include electricity or charging costs, AV software and connectivity fees, sensor/computing maintenance costs, autonomous-system service subscriptions, and other recurring mobility-related operating expenditures incurred on a per-trip or per-mile basis.

are different. Therefore, we also present elasticity effects following the model results to demonstrate the different magnitude effects across the four models.

4.2.1 Service Attributes

As noted above, three service attributes were included in the experimental design, capturing a fixed cost of vehicle ownership, a per-mile variable cost using each option, and an expected waiting time among those using the ridehailing option. Interestingly, waiting time was found not to be statistically significant using any of the model specifications. This may be because of the relatively small range of waiting times considered (from 3 to 9 minutes) in the stated preference design and/or because respondents focused on their own existing experience of ridehailing services alongside the costs rather than closely considering this aspect of the ridehailing experience. However, the two cost-related variables were statistically significant in all four models. A variety of continuous distributional forms were tested in the random coefficients specifications for the cost coefficients, including the unbounded normal and YJ distributions and the bounded (to the negative interval) log-normal distribution. In these tests, and for both the MNPRC and MNYJRC formulations, and for both the fixed and variable cost coefficients, the log-normal cost coefficients provided the best data fit. The log-normal distribution provides skewed distributions while maintaining strict negativity for the cost coefficients. An implicit Gaussian copula was also tested to capture unobserved dependence effects across the two cost coefficients as well as among the cost coefficients and the kernel error terms. Of these, the latter correlations did not turn out to be statistically significant in the random coefficient specifications, and so were not considered. Finally, separate coefficients were tested for the fixed costs for regular vehicles and AVs as well as for the variable costs for regular vehicles, AVs, and ridehailing. However, no significant differences were found between these mode-specific cost coefficients, with model fit improvements being negligible compared to the specification of a single generic fixed cost coefficient and a single generic variable cost coefficient. As may be observed from Table 2, higher costs (either monthly or per-mile) are a deterrent to the selection of the corresponding alternative (note that fixed cost is included in hundreds of dollars per month and variable cost is included in dollars per mile). For the random coefficient models, the median coefficient value is presented in the table for the cost coefficients, while the scale presented corresponds to the standard deviation of the corresponding normal distribution. The median of the cost coefficients are in the same order of magnitude across the many models, even if not directly comparable. The marginal rate of substitution between the fixed and variable cost also suggests remarkably similar trade-offs across the many models, with respondents considering a $1/mile increase in variable cost as about equivalent to about a $150/month increase in fixed cost (for instance, in the MNP model, the ratio of the coefficient on variable cost (in $/mile) to fixed cost (in 100s of $/month) is 1.486; that is respondents consider a 1$/mile increase in variable costs as about equivalent to 1.486*100=$148.6/month increase in fixed costs). At the U.S. average of about 1000 miles per month that a vehicle is driven, this implies that $1000 of variable cost has the same disutility as $150 of fixed cost, or that fixed costs are about 6.7 times (1000/150) more important than variable costs in vehicle purchase decisions. That is, respondents place substantially greater behavioral weight on fixed monthly ownership costs than on variable per-mile operating costs. This finding is broadly consistent with the vehicle-choice literature, which has repeatedly shown that consumers tend to place disproportionately greater weight on upfront or recurring ownership costs than on future operating expenditures. One explanation is that fixed costs are more immediate, salient, and cognitively transparent, while variable operating costs are distributed over time and are often

discounted behaviorally in standard lifecycle-cost calculations (see, for example, Allcott and Wozny, 2014; Gillingham et al., 2021; Huse and Koptyug, 2022; Petrov et al., 2025).

The implied scale (standard deviation) of the lognormal distribution may be computed using well-known expressions for the log-normal distribution (see, for example, Johnson et al., 1994). These turn out to be 0.0034 for the fixed cost coefficient and 0.0111 for the variable cost coefficient in the MNPRC model, and 0.0068 for the fixed cost coefficient and 0.0221 for the variable cost parameter in the MNYJRC model. In both the random coefficient models, the scale is consistently larger for the variable cost coefficient than for the fixed cost parameter. This is to be expected because variable operating expenditures depend strongly on anticipated vehicle usage patterns, travel intensity, and lifestyle characteristics, all of which can vary substantially across households. In the AV context, this heterogeneity in variable cost responsiveness may be further amplified by uncertainty regarding future AV usage, charging behavior, software-service expenditures, and substitution with mobility-as-a-service alternatives. In contrast, fixed ownership costs are generally more transparent, stable, and uniformly experienced across individuals. In fact, our results show very little heterogeneity to fixed costs, given that the ratio of the median value to the standard deviation for the fixed cost coefficient is 12.64 in both the MNPRC and MNYJRC models (this ratio is the same in both the models because the standard deviation of the underlying normal distribution is estimated to be identically 0.079 in both models). The corresponding ratio is of the order of 5.7 in both the models for the variable cost coefficient.

The log-normal distributions for the cost parameters imply left-skewed coefficient distributions, indicating that while many individuals exhibit moderate levels of cost sensitivity, a smaller subset displays substantially higher sensitivity to cost. The two cost parameters are also estimated to be positively related, suggesting that some individuals are intrinsically more cost sensitive overall. However, the estimated Gaussian copula correlation of the log-normal cost distributions is notably higher in the MNPRC model than in the MNYJRC model. One possible explanation is that, under the symmetry restriction imposed by the MNPRC specification, the model compensates for its inability to capture asymmetric heterogeneity by attributing more of the extreme cost sensitivity to correlation between the two cost coefficients. In contrast, the MNYJRC model is able to accommodate skewness directly through the asymmetric kernel-error structure, reducing the need for the correlation parameter to absorb such effects. These results suggest that failing to account for distributional asymmetry may lead analysts to infer greater uniformity in fixed and variable cost sensitivities across decision-makers, while allowing for skewness reveals a more differentiated pattern of responses to fixed and variable costs.

4.2.2 Individual Characteristics

All four models indicate that older individuals, women, and individuals with less than a Bachelor's degree of formal educational attainment are less inclined toward AV ownership relative to younger individuals, men, and individuals with a Bachelor's degree or higher. These results are consistent with findings from previous studies (see, for example, Charness et al., 2018; Asmussen et al., 2020; Rahman and Thill, 2026). Similarly, all the models suggest that nonwhite individuals may be more interested in AV ridehailing than in the purchase of either a new regular vehicle or a new AV. This likely reflects existing differences in vehicle ownership constraints and preferences, as minority racial groups tend to have a greater openness toward mobility-as-a-service alternatives while also facing more substantial barriers to vehicle ownership (Klein and Smart, 2017; Brown, 2019; Molloy et al., 2024). However, while the MNP models suggest that women are less interested in AV ridehailing and that individuals with higher levels of formal educational attainment are more

interested in AV ridehailing, these effects are not as statistically significant in the MNYJ models (and actually are well below any typically used level of statistical significance for variable retainment). This suggests that some of the variation attributed to the observed gender and education variables in the MNP specification may instead reflect more general asymmetries in the latent utility distributions that are not adequately represented under the symmetry restriction imposed by the MNP model.

4.2.3 Household Characteristics

The effects of household characteristics in Table 2 indicate that individuals in the highest household income category ($100,000 or more) are more inclined to adopt AVs in ownership mode than those in lower income categories, and individuals in higher income households tend to rely less on AV ridehailing services relative to those in lower income households. Notably, this is one area where the inclusion of the random coefficients for cost-related variables appears to have a substantial impact. Specifically, the two models without random coefficients indicate that those in the highest income category ($100,000 or more) actually exhibit a higher utility for ridehailing than those in the middle income category ($50,000 - $99,999), as evidenced by the smaller magnitude of the coefficient associated with the highest income group relative to the middle income group in both the MNP and MNYJ models. However, once random coefficients are introduced, the estimated income effects align more closely with behavioral expectations, with the magnitudes of the coefficients increasing monotonically with household income. This suggests that accounting for unobserved heterogeneity in cost sensitivities is important for properly disentangling the effects of income on AV ownership and ridehailing preferences.

Another notable difference across the models pertains to the effect of presence of children in the household. The MNP models indicate that the presence of children does not have a statistically significant effect on AV ridehailing utility even at the 60% confidence level, while the variable negatively influences AV ridehailing utility at over the 90% confidence level in the MNYJ models. The latter result is intuitively appealing, given the potential safety and supervision concerns parents may associate with AV ridehailing services for their children (Ahmed and Hyland, 2023; De Leonardis et al., 2024).

4.2.4 Constants and Error Distribution Parameters

The alternative-specific constants shown closer to the bottom of Table 2 primarily absorb residual systematic utility differences across alternatives not otherwise captured in the specification, including effects associated with differing ranges and distributions of explanatory variables across alternatives. Together with the remaining components of the model, they help account for the aggregate observed choice patterns across alternatives. Even so, the strong negative constant on the AV ridehailing mode is clearly noticeable, reflecting the relatively low choice share of this alternative.

The elements of the estimated Gaussian copula correlation matrix **R** are presented next in Table 2. Both the MNPRC and MNP models include a significant positive correlation for the non-diagonal element of the differenced covariance matrix (defined with respect to utility differences from the first “regular vehicle” alternative). However, the structural interpretation of this differenced correlation is not straightforward, because many structural covariance matrices can imply the same differenced covariance matrix. If one were to assume that the variance associated with the “regular vehicle” alternative is negligible and that this alternative is uncorrelated with the other two alternatives (as is sometimes done when interpretation is desired), the implication would

be that the utilities associated with AV ownership and AV ridehailing are positively correlated. In contrast, the MNYJRC and MNYJ models indicate a significant positive correlation between regular vehicle ownership and AV ownership, together with significant negative correlations between regular vehicle ownership and AV ridehailing and between AV ownership and AV ridehailing. These richer structural interpretations associated with the copula correlation elements in the MNYJ models would not have been possible within the MNP framework, representing an important advantage of the MNYJ formulation.

The scale parameters are shown below the correlations in Table 2. In the MNP model, the scales refer to the differenced kernel error terms (with differencing taken relative to the first "regular vehicle" alternative) and are therefore not directly interpretable at the structural utility level (recall from earlier that the scale normalization used in this paper differs from the conventional normalization adopted in MNP models; specifically the sum of the squared scale values equals one in our normalization). The scales from the MNYJ models, on the other hand, refer directly to the structural kernel errors and therefore admit direct interpretation. The results from the MNYJ models reveal that the scale associated with the AV ridehailing alternative is substantially larger than that for the other alternatives, perhaps because the logistical and operational mechanics associated with AV ridehailing services (such as vehicle availability, waiting times, routing reliability, pickup coordination, or system trustworthiness) remain more uncertain than in the case of privately owned regular or AV vehicles.

Finally, for the MNYJ models, the YJ shape parameters are shown in the bottom row of the table. A visual translation of the implications of these shape parameters for the distributions of the standardized kernel error terms ς_i in Equation (9) is provided in Figure 2 (we plot the distributions of the standardized ς_i terms as opposed to the overall kernel error terms $s_i\varsigma_i$ to maintain focus on a comparative discussion of the shape characteristics). The results suggest a relatively small leftward skew for the regular vehicle ownership alternative, a larger leftward skew for the AV ownership alternative, and a clear rightward skew for the AV ridehailing alternative. These findings are consistent with the behavioral motivations discussed earlier, suggesting the presence of a relatively small group of individuals who are strongly opposed to AV ownership and a relatively small group who are especially attracted to AV ridehailing. More generally, the results provide empirical evidence that the latent utility distributions associated with the three alternatives are not symmetric, reinforcing the value of accommodating alternative-specific asymmetry in kernel error distributions.

4.3 Data Fit Measures

We compare the proposed MNYJRC model with the three restricted formulations using a series of fit metrics, as shown in Table 3. First, we compute adjusted likelihood ratio indices for each model relative to both the log-likelihood at zero and the log-likelihood at constants. These indices may be computed as:

$$\bar{\rho}_c^2 = 1 - \frac{L(\hat{\theta}) - M_c}{L(c)} \quad \text{and} \quad \bar{\rho}_0^2 = 1 - \frac{L(\hat{\theta}) - M_0}{L(0)} \tag{11}$$

where $L(\hat{\theta})$ is the log-likelihood at convergence, $L(0)$ is the log-likelihood at zero, $L(c)$ is the log-likelihood at constants, M_0 is the total number of parameters in the model, and M_c is the number of non-constant parameters in each model. The proposed model outperforms all three restricted models based on both adjusted likelihood ratio indices (see the fifth and sixth numeric

rows of Table 3). Next, we compare the restricted models with the proposed MNYJRC model using formal likelihood ratio tests. All three restricted models are rejected in favor of the proposed model at well over the 95% confidence level, with both the MNP models being rejected at essentially the 100% confidence level. Finally, we compare the average probability of a correct prediction across the four models. These statistics once again highlight the improved predictive performance of the MNYJRC model relative to the restricted models.

4.4 Elasticity Effects and Policy Implications

While some differences across the models are readily apparent from the coefficients presented in Table 2, and the model fit of the proposed MNYJRC model outperforms the MNP models in overall fit, the coefficient values provide only the effects on the underlying utilities for each alternative, not the actual magnitude (or even direction) of effects on choice probabilities. Further, and as noted above, because of differences in the shapes and scales of the random error distributions for each model, the magnitudes of the coefficients are not readily comparable. Thus, to provide a better sense of the effects of the exogenous variables on the probability of selecting each alternative, we compute aggregate elasticity effects.

Specifically, for each categorical exogenous variable, all individuals are first assigned to a base level of the variable (for example, assigning all individuals to "Male") while retaining the original values for all other exogenous variables. Each of the four models is used to generate predicted probabilities for each individual choosing each alternative, and aggregate shares are computed as the averages of these probabilities across individuals. Next, for the same exogenous variable, all individuals are assigned to a treatment level (for the gender example, assigning all individuals to "Female"), again retaining all other exogenous variable values, and the shares are recomputed. The percentage change in the shares for each alternative is reported in Table 4. For categorical exogenous variables, the elasticity effects correspond to changes from the lowest to the highest levels. For the continuous cost variables, the reported "arc" elasticity effects correspond to a 20% increase from the originally presented cost levels.

As may be observed from the table, there is broad alignment across the four models in both the direction and magnitude of the cost effects, although the estimated effects are generally somewhat smaller in magnitude in the MNYJ model. Focusing on a comparison between the proposed MNYJRC model and the two MNP models, notable differences emerge for the AV fixed cost variable and the AV ridehailing cost variable. Specifically, the MNP models overestimate the negative sensitivity toward AV ownership associated with increases in AV fixed costs, while underestimating the negative sensitivity toward AV ridehailing in response to an increase in AV ridehailing cost. The implication is that AV fixed cost pricing changes (AV ridehailing per-mile cost changes) have a somewhat smaller (larger) effect on AV ownership (AV ridehailing adoption) decisions than implied by the MNP models. For example, if a transportation agency were to consider subsidies for AV ownership with the goal of shifting travelers away from conventional vehicle ownership, the MNYJ models suggest that AV ownership uptake resulting from such subsidies would likely be lower than that projected by the MNP models. Conversely, subsidies for AV ridehailing are likely to have a stronger positive effect on AV ridehailing adoption than projected by the MNP models.

Regarding individual and household characteristics, several notable differences also emerge across the models. First, the MNYJ models indicate that the effects of gender and educational attainment on AV ridehailing use are substantially overestimated in the MNP models. In contrast, the presence of children has a considerably larger negative effect on AV ridehailing

uptake in the MNYJ models (-20.40% in the MNYJ and -17.77% in the MNYJRC) than in the MNP models (-8.57% in the MNP and -8.39% in the MNPRC). From a policy and service design perspective, this suggests that reliance on the MNP results would overestimate AV ridehailing adoption among households with children, highlighting the need to more specifically improve the appeal of AV ridehailing services for such households. Potential measures could include child-seat availability, guaranteed door-to-door service, or bundled family-oriented ridehailing plans. Overall, the elasticity results demonstrate that the specification of the kernel error distribution meaningfully affects the estimated impacts of several variables, and that ignoring the skewed nature of the error distribution can lead not only to poorer model fit, but also to materially different behavioral implications and policy conclusions.

5. CONCLUSION

This paper proposes and demonstrates the application of the MNYJ model, a flexible, computationally tractable, and parsimonious multinomial choice formulation that accommodates asymmetric and skewed kernel error distributions across alternatives. The model is built on the inverse Yeo–Johnson transformation of latent normal error terms, enabling a wide range of asymmetric distributional shapes while remaining analytically and computationally manageable. An implicit Gaussian copula is used to tie the transformed marginals together across alternatives, enabling the estimation of the full set of structural scales and correlation matrix, rather than the scales and correlation matrix of the differenced kernel error terms, as in the traditional MNP formulation. This formulation results in improvements in model fit and estimation accuracy, as well as improving the interpretability of the structural parameters governing cross-alternative error dependence and asymmetry. The model extends naturally to include asymmetric random coefficients that may be correlated with one another and with the kernel error terms.

Using a simulation exercise, we demonstrate the ability of the model to recover all parameters of the data-generating process, including the full structural correlation matrix of the kernel errors. The simulation analysis provides strong evidence of low estimation bias and generally accurate finite-sample parameter recovery across all parameter categories, including the asymmetric shape parameters and the structural correlation matrix. Coverage probabilities are generally close to their nominal 95% targets. By contrast, the traditional MNP model, when estimated on data generated from asymmetric kernel error processes, produces substantially biased parameter estimates and coverage probabilities that fall well below nominal levels. These simulation results reinforce the importance of accounting for asymmetric kernel error distributions, as incorrect symmetry assumptions can lead not only to poorer model fit, but also to materially distorted behavioral inference.

We also demonstrate the model using an empirical application involving future vehicle ownership intentions, specifically the choice among purchasing a regular vehicle, purchasing an AV, or relying on AV ridehailing services. The proposed MNYJRC model consistently outperforms all restricted alternatives based on standard fit measures, and the elasticity analysis reveals that the choice of kernel error distribution has meaningful implications for the estimated effects of several variables. Most notably, the effects of gender, formal educational attainment, and the presence of children differ substantially across the MNP and MNYJ formulations, with important implications for transportation planning and policy design. Further, while the differenced covariance implications of the MNYJ and MNP models are broadly similar, the MNYJ models additionally recover the structural utility-level error correlation matrix, enabling direct interpretation of cross-alternative dependence at the level of the original utilities rather than only

in the differenced space. Finally, the estimated YJ shape parameters reveal meaningful asymmetries across the three alternatives, including a moderate leftward skew for regular vehicle ownership, a larger leftward skew for AV ownership, and a pronounced rightward skew for AV ridehailing, highlighting the importance of allowing for these asymmetric kernel error structures in multinomial choice settings.

Several directions for future research remain. First, the MNYJ framework may be extended to richer panel, spatial, and dynamic choice settings in which asymmetric preference evolution and state dependence may play an important role. Second, additional work is needed on computationally efficient covariance estimation and inference procedures for highly flexible asymmetric choice models, particularly in the presence of weak curvature and simulation-based likelihood evaluation. Third, future research may explore more general copula structures beyond the implicit Gaussian copula considered here, allowing for asymmetric dependence and tail dependence across alternatives. Finally, the proposed framework may prove particularly useful in emerging transportation settings involving autonomous mobility, shared mobility services, and evolving vehicle ownership paradigms, where behavioral asymmetries and heterogeneous substitution patterns are likely to be especially important.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Department of Transportation through the Center for Understanding Future Travel Behavior and Demand (TBD) (Grant No. 69A3552344815 and No. 69A3552348320). The authors are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

Ahmed, T., Hyland, M., 2023. Exploring the Role of Ride-Hailing in Trip Chains. Transportation 50, 959–1002. https://doi.org/10.1007/s11116-022-10269-w

Allcott, H., Wozny, N., 2014. Gasoline Prices, Fuel Economy, and the Energy Paradox. Rev. Econ. Stat. 96, 779–795. https://doi.org/10.1162/REST_a_00419

Asmussen, K., Mondal, A., Bhat, C.R., 2020. A Socio-Technical Model of Autonomous Vehicle Adoption Using Ranked Choice Stated Preference Data. Transp. Res. Part C Emerg. Technol. 121. https://doi.org/10.1016/j.trc.2020.102835

Azzalini, A., 2005. The Skew-normal Distribution and Related Multivariate Families. Scand. J. Stat. 32, 159–188. https://doi.org/10.1111/j.1467-9469.2005.00426.x

Bhat, C.R., 2024. Transformation-Based Flexible Error Structures for Choice Modeling. J. Choice Model. 53, 100522. https://doi.org/10.1016/j.jocm.2024.100522

Bhat, C.R., 2020. Consumer choice modeling: The promises and the cautions, in: Mapping the Travel Behavior Genome. Elsevier, pp. 63–80. https://doi.org/10.1016/B978-0-12-817340-4.00005-X

Bhat, C.R., 2018. New Matrix-Based Methods for the Analytic Evaluation of the Multivariate Cumulative Normal Distribution Function. Transp. Res. Part B Methodol. 109, 238–256. https://doi.org/10.1016/j.trb.2018.01.011

Bhat, C.R., 2001. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. Transp. Res. Part B Methodol. 35, 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X

Bhat, C.R., 1999. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. Work. Pap. Dep. Civ. Eng. University Texas at Austin.

Bhat, C.R., 1995. A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. Transp. Res. Part B Methodol. 29, 471–483. https://doi.org/10.1016/0191-2615(95)00015-6
Bhat, C.R., Mondal, A., Pinjari, A.R., 2025. A Flexible Non-Normal Random Coefficient Multinomial Probit Model: Application to Investigating Commuter's Mode Choice Behavior in a Developing Economy Context. Transp. Res. Part B Methodol. 195, 103186. https://doi.org/10.1016/j.trb.2025.103186
Brown, A., 2019. Redefining Car Access: Ride-Hail Travel and Use in Los Angeles. J. Am. Plann. Assoc. 85, 83–95. https://doi.org/10.1080/01944363.2019.1603761
Bunch, D.S., 1991. Estimability in the Multinomial Probit Model. Transp. Res. Part B Methodol. 25, 1–12. https://doi.org/10.1016/0191-2615(91)90009-8
Castillo, E., Menéndez, J.M., Jiménez, P., Rivas, A., 2008. Closed Form Expressions for Choice Probabilities in the Weibull Case. Transp. Res. Part B Methodol. 42, 373–380. https://doi.org/10.1016/j.trb.2007.08.002
Charness, N., Yoon, J.S., Souders, D., Stothart, C., Yehnert, C., 2018. Predictors of Attitudes Toward Autonomous Vehicles: The Roles of Age, Gender, Prior Knowledge, and Personality. Front. Psychol. 9. https://doi.org/10.3389/fpsyg.2018.02589
Daganzo, C., 2014. Multinomial Probit: The Theory and Its Application to Demand Forecasting. Elsevier.
De Leonardis, D., Levi, S., Benedick, A.K., Eisenhauer, E., Ferg, R., Petraglia, E., 2024. Child Passenger Safety Perception and Practices in Ride-Sharing Vehicles (No. DOT HS 813 532). National Highway Traffic Safety Administration.
Dias, C., Abdullah, M., Lovreglio, R., Sachchithanantham, S., Rekatheeban, M., Sathyaprasad, I.M.S., 2022. Exploring Home-to-School Trip Mode Choices in Kandy, Sri Lanka. J. Transp. Geogr. 99, 103279. https://doi.org/10.1016/j.jtrangeo.2022.103279
Dubey, S., Bansal, P., Daziano, R.A., Guerra, E., 2020. A Generalized Continuous-Multinomial Response Model with a T-Distributed Error Kernel. Transp. Res. Part B Methodol. 133, 114–141. https://doi.org/10.1016/j.trb.2019.12.007
Fosgerau, M., Bierlaire, M., 2009. Discrete Choice Models with Multiplicative Error Terms. Transp. Res. Part B Methodol. 43, 494–505. https://doi.org/10.1016/j.trb.2008.10.004
Gallaugher, M.P.B., McNicholas, P.D., Melnykov, V., Zhu, X., 2020. Skewed Distributions or Transformations? Modelling Skewness for a Cluster Analysis. https://doi.org/10.48550/arXiv.2011.09152
Gillingham, K.T., Houde, S., van Benthem, A.A., 2021. Consumer Myopia in Vehicle Purchases: Evidence from a Natural Experiment. Am. Econ. J. Econ. Policy 13, 207–238. https://doi.org/10.1257/pol.20200322
Guo, J., Feng, T., Timmermans, H.J.P., 2020. Co-Dependent Workplace, Residence and Commuting Mode Choice: Results of a Multi-Dimensional Mixed Logit Model with Panel Effects. Cities 96, 102448. https://doi.org/10.1016/j.cities.2019.102448
Huse, C., Koptyug, N., 2022. Salience and Policy Instruments: Evidence from the Auto Market. J. Assoc. Environ. Resour. Econ. 9, 345–382. https://doi.org/10.1086/716878
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Volume 1. John Wiley & Sons.
Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision Under Risk. Econometrica 47, 263–292.

Khoeini, S., Pendyala, R.M., Capasso da Silva, D., Lee, Y., Dias, F., Salon, D., Circella, G., Maness, M., 2019. Attitudes Towards Emerging Mobility Options and Technologies – Phase 2: Pilot and Full Survey Deployment. Teaching Old Models New Tricks (TOMNET) Transportation Center, Arizona State University.

Khoeini, S., Pendyala, R.M., Capasso da Silva, D., Lee, Y., Dias, F., Salon, D., Circella, G., Maness, M., 2018. Attitudes towards Emerging Mobility Options and Technologies – Phase 1: Survey Design. Teaching Old Models New Tricks (TOMNET) Transportation Center, Arizona State University.

Klein, N.J., Smart, M.J., 2017. Car Today, Gone Tomorrow: The Ephemeral Car in Low-Income, Immigrant and Minority Families. Transportation 44, 495–510. https://doi.org/10.1007/s11116-015-9664-4

Krueger, R., Bierlaire, M., Gasos, T., Bansal, P., 2023. Robust Discrete Choice Models with T-Distributed Kernel Errors. Stat. Comput. 33, 2. https://doi.org/10.1007/s11222-022-10182-3

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15, 447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5%3C447::AID-JAE570%3E3.0.CO;2-1

Molloy, Q., Garrick, N., Atkinson-Palombo, C., 2024. Black Households Are More Burdened by Vehicle Ownership than White Households. Transp. Res. Rec. 2678, 163–173. https://doi.org/10.1177/03611981241231968

Nagler, J., 1994. Scobit: An Alternative Estimator to Logit and Probit. Am. J. Polit. Sci. 38, 230–255. https://doi.org/10.2307/2111343

Paleti, R., 2019. Discrete Choice Models with Alternate Kernel Error Distributions. J. Indian Inst. Sci. 99, 673–681. https://doi.org/10.1007/s41745-019-00128-6

Patil, G.R., Basu, R., Rashidi, T.H., 2020. Mode Choice Modeling Using Adaptive Data Collection for Different Trip Purposes in Mumbai Metropolitan Region. Transp. Dev. Econ. 6, 9. https://doi.org/10.1007/s40890-020-0099-z

Petrov, I., Carroll, J., Denny, E., 2025. Communicating Car Costs: Results from Stated and Revealed Preference Experiments. Transp. Res. Part Transp. Environ. 147, 104921. https://doi.org/10.1016/j.trd.2025.104921

Peyhardi, D.J., 2020. Robustness of Student link function in multinomial choice models. J. Choice Model. 36, 100228. https://doi.org/10.1016/j.jocm.2020.100228

Prentice, R.L., Gloeckler, L.A., 1978. Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data. Biometrics 34, 57–67. https://doi.org/10.2307/2529588

Rahman, Md.M., Thill, J.-C., 2026. Who Is Inclined to Buy an Autonomous Vehicle? Empirical Evidence from California. Transportation 53, 71–104. https://doi.org/10.1007/s11116-024-10490-9

Stukel, T.A., 1988. Generalized Logistic Models. J. Am. Stat. Assoc. 83, 426–431. https://doi.org/10.1080/01621459.1988.10478613

Train, K., 2000. Halton Sequences for Mixed Logit. Tech. Pap. Dep. Econ. Univ. Calif. Berkeley.

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

Vichiensan, V., Wasuntarasook, V., Malaitham, S., Fukuda, A., Rujopakarn, W., 2025. Willingness to Pay for Station Access Transport: A Mixed Logit Model with Heterogeneous Travel Time Valuation. Sustainability 17. https://doi.org/10.3390/su17156715

White, H., 1982. Maximum Likelihood Estimation of Misspecified Models. Econometrica 50, 1–25. https://doi.org/10.2307/1912526

Wu, Y., Yildirimoglu, M., Zheng, Z., 2026. Modeling car owners' stated preferences for electric micro-mobility ownership: Evidence from a mixed logit approach. Transp. Policy 181, 104063. https://doi.org/10.1016/j.tranpol.2026.104063

Yeo, I., Johnson, R.A., 2000. A New Family of Power Transformations to Improve Normality or Symmetry. Biometrika 87, 954–959. https://doi.org/10.1093/biomet/87.4.954

Table 1: Simulation Results

MNYJ Model Performance											
Metric	Main Coefficients			Correlation Terms			Scale Parameters*		YJ Parameters		
	b_1	b_2	b_3	Θ_1	Θ_2	Θ_3	s_2	s_3	λ_1	λ_2	λ_3
True Values	-0.500	0.250	0.500	0.350	0.200	0.300	0.500	0.350	0.250	0.550	1.450
Mean Parameter Estimate	-0.499	0.248	0.497	0.348	0.207	0.294	0.496	0.352	0.253	0.554	1.454
APB	0.285	0.971	0.671	0.617	3.465	2.076	0.813	0.533	1.072	0.747	0.249
CP	93.200	92.000	91.200	96.400	94.400	92.800	96.000	95.600	90.400	96.800	94.000
FSSD	0.019	0.032	0.038	0.050	0.074	0.067	0.039	0.041	0.062	0.122	0.071
ASE	0.021	0.035	0.043	0.049	0.076	0.065	0.040	0.043	0.060	0.136	0.068
APBASE	8.276	7.908	11.310	1.358	3.301	3.444	0.878	4.720	3.741	11.278	4.611
Mean Likelihood						-5321.395					
MNP Model Performance											
Metric	Main Coefficients			Correlation Terms			Scale Parameters		YJ Parameters		
	b_1	b_2	b_3	Θ_1	Θ_2	$\tilde{\Theta}_3$	s_2	$\tilde{s}_3$	λ_1	λ_2	λ_3
True Values	-0.500	0.250	0.500	--	--	--	--	--	--	--	--
Mean Parameter Estimate	-0.470	0.310	0.584	--	--	0.731	--	0.716	--	--	--
APB	5.973	24.078	16.890	--	--	--	--	--	--	--	--
CP	47.200	26.000	1.600	--	--	--	--	--	--	--	--
FSSD	0.015	0.024	0.022	--	--	0.027	--	0.009	--	--	--
ASE	0.014	0.022	0.020	--	--	0.025	--	0.010	--	--	--
APBASE	1.671	7.863	6.268	--	--	8.273	--	10.362	--	--	--
Mean Likelihood						-5355.6					
Likelihood Ratio Test Performance						100%					

* The final scale parameter (s_1) is implicitly identified through the scale normalization.

Table 2: Vehicle Ownership Model Results for MNP and MNYJ Models

Variable (base)	MNP						MNPRC						MNYJ						MNYJRC					
	Regular Vehicle		Autonomous Vehicle		No Vehicle / AV Ridehailing		Regular Vehicle		Autonomous Vehicle		No Vehicle / AV Ridehailing		Regular Vehicle		Autonomous Vehicle		No Vehicle / AV Ridehailing		Regular Vehicle		Autonomous Vehicle		No Vehicle / AV Ridehailing	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Service Attributes																								
Fixed Cost (in $100/month)																								
Median			-0.037	-2.252					-0.043	2.067					-0.064	-3.446					-0.086	-2.334		
Standard Deviation									0.079	2.022											0.079	1.692		
Variable Cost (in $/mile)																								
Median			-0.055	-2.156					-0.063	-2.125					-0.085	-1.868					-0.127	-2.304		
Standard Deviation									0.172	1.934					--						0.170	1.841		
Correlation between Fixed and Variable Costs			--						0.880	1.754					--						0.569	1.552		
Individual Characteristics																								
Age (18 - 44)																								
45 - 64	--		-0.056	-2.186	--		--		-0.065	-2.455	--		--		-0.119	-3.772	--		--		-0.136	-2.316	--	
65 or older	--		-0.085	-2.215	--		--		-0.100	-2.489	--		--		-0.180	-4.899	--		--		-0.210	-2.334	--	
Gender (Male)																								
Female	--		-0.038	-2.129	-0.089	-1.702	--		-0.045	-2.373	-0.090	-1.690	--		-0.078	-2.766	-0.035	-0.256	--		-0.092	-2.310	-0.104	-0.928
Education (Less than Bachelor's)																								
Bachelor's degree or higher	--		0.019	1.761	0.086	1.665	--		0.022	1.889	0.086	1.501	--		0.040	1.788	0.036	0.347	--		0.046	1.838	0.144	1.207
Race (White)																								
Nonwhite	--		--		0.146	2.268	--		--		0.146	2.282	--		--		0.189	1.711	--		--		0.229	1.838
Household Characteristics																								
Income (Less than $25,000)																								
$25,000 - $49,999	--		--		-0.295	-2.968	--		--		-0.291	-2.951	--		--		-0.773	-2.429	--		--		-0.453	-2.373
$50,000 - $99,999	--		--		-0.454	-4.893	--		--		-0.452	-4.899	--		--		-0.854	-3.487	--		--		-0.701	-2.388
$100,000 or more	--		0.022	1.657	-0.428	-5.370	--		0.026	1.781	-0.521	-5.315	--		0.053	1.838	-0.808	-3.395	--		0.055	1.847	-0.847	-3.391
Presence of Children (No)																								
Yes	--		--		-0.051	-0.785	--		--		-0.050	-0.765	--		--		-0.192	-1.868	--		--		-0.191	-1.717
Alternative Constants	--		0.008	0.628	-0.764	-8.095	--		0.009	0.595	-0.755	-7.891	--		-0.001	-0.034	-0.540	-2.094	--		0.016	0.384	-0.671	-2.728
Estimated Correlation Matrices																								
Regular Vehicle	--						--						1.000	--					1.000	--				
Autonomous Vehicle	--		1.000	--			--		1.000	--			0.437	1.955	1.000	--			0.473	1.743	1.000	--		
No Vehicle / AV Ridehailing	--		0.539	1.745	1.000	--	--		0.535	1.793	1.000	--	-0.596	-2.330	-0.659	-2.639	1.000	--	-0.698	-2.390	-0.669	-2.030	1.000	--
Scales	--		0.129	--	0.992	135.009	--		0.151	--	0.989	111.636	0.301	--	0.199	4.965	0.933	12.609	0.350	--	0.219	3.324	0.911	12.334
YJ Parameters	--		--		--		--		--		--		1.128	1.700	1.433	2.301	0.254	-7.485	1.154	1.562	1.452	1.955	0.286	2.262

Table 3: Model Fit Comparison

Metric	**MNP**	**MNPRC**	**MNYJ**	**MNYJRC**
Log-likelihood at convergence	-2863.14	-2855.86	-2838.39	-2833.47
Number of non-constant parameters	16	19	22	25
Log-likelihood at zero		-3517.76		
Log-likelihood at constants		-3016.60		
Adjusted likelihood ratio index (zero)	0.1810	0.1822	0.1863	0.1868
Adjusted likelihood ratio index (constants)	0.0456	0.0470	0.0518	0.0524
Likelihood Ratio Test between MNP and MNYJRC		$LR = 59.34 > \chi^2_{(9,0.05)} = 16.92$		
Likelihood Ratio Test between MNPRC and MNYJRC		$LR = 44.80 > \chi^2_{(6,0.05)} = 12.59$		
Likelihood Ratio Test between MNPYJ and MNYJRC		$LR = 9.84 > \chi^2_{(3,0.05)} = 7.81$		
Average Probability of a Correct Prediction	0.4645	0.4650	0.4656	0.4661

Table 4: Elasticity Effects

Exogenous Variable			MNP			MNPRC			MNYJ			MNYJRC		
			Regular Vehicle	Autonomous Vehicle	No Vehicle / Ridehailing	Regular Vehicle	Autonomous Vehicle	No Vehicle / Ridehailing	Regular Vehicle	Autonomous Vehicle	No Vehicle / Ridehailing	Regular Vehicle	Autonomous Vehicle	No Vehicle / Ridehailing
Service Attributes														
Effect of 20% increase to:														
Regular Vehicle Fixed Cost			-11.20	18.68	1.45	-11.13	18.44	1.76	-9.36	14.95	1.91	-11.04	17.96	2.59
Regular Vehicle Variable Cost			-2.51	4.19	0.33	-2.52	4.18	0.40	-1.88	2.98	0.44	-2.46	4.01	0.56
Autonomous Vehicle Fixed Cost			9.90	-17.85	2.44	9.78	-17.77	2.82	8.35	-14.44	1.29	8.44	-15.27	2.12
Autonomous Vehicle Variable Cost			2.35	-4.24	0.57	2.35	-4.26	0.67	1.74	-3.03	0.33	2.25	-4.09	0.70
Ridehailing Variable Cost			0.32	0.95	-4.16	0.38	1.10	-4.89	0.53	0.63	-4.09	0.52	1.12	-5.62
Individual and Household Characteristics														
Variable	**Base**	**Treatment**												
Age	< 45	65+	45.73	-51.02	9.50	45.38	-51.06	11.12	45.42	-49.43	7.20	45.46	-51.24	12.20
Gender	Male	Female	20.20	-23.46	-10.49	20.16	-23.50	-10.17	18.40	-24.02	-0.04	19.50	-24.52	-2.07
Education	Less than Bachelor's	Bachelor's Degree	-9.03	12.81	13.77	-9.00	12.86	13.41	-8.16	14.99	1.85	-8.98	12.95	7.91
Race	White	Nonwhite	-2.08	-5.82	28.09	-2.15	-5.79	28.35	-1.83	-3.52	22.00	-2.34	-4.83	26.34
Income	< $25,000	$100,000+	-0.39	53.93	-58.85	-0.26	53.50	-58.82	0.73	49.12	-53.94	0.67	50.66	-59.11
Presence of Children	No	Yes	0.67	2.00	-8.57	0.68	1.93	-8.39	2.96	3.11	-20.40	1.73	3.74	-17.74

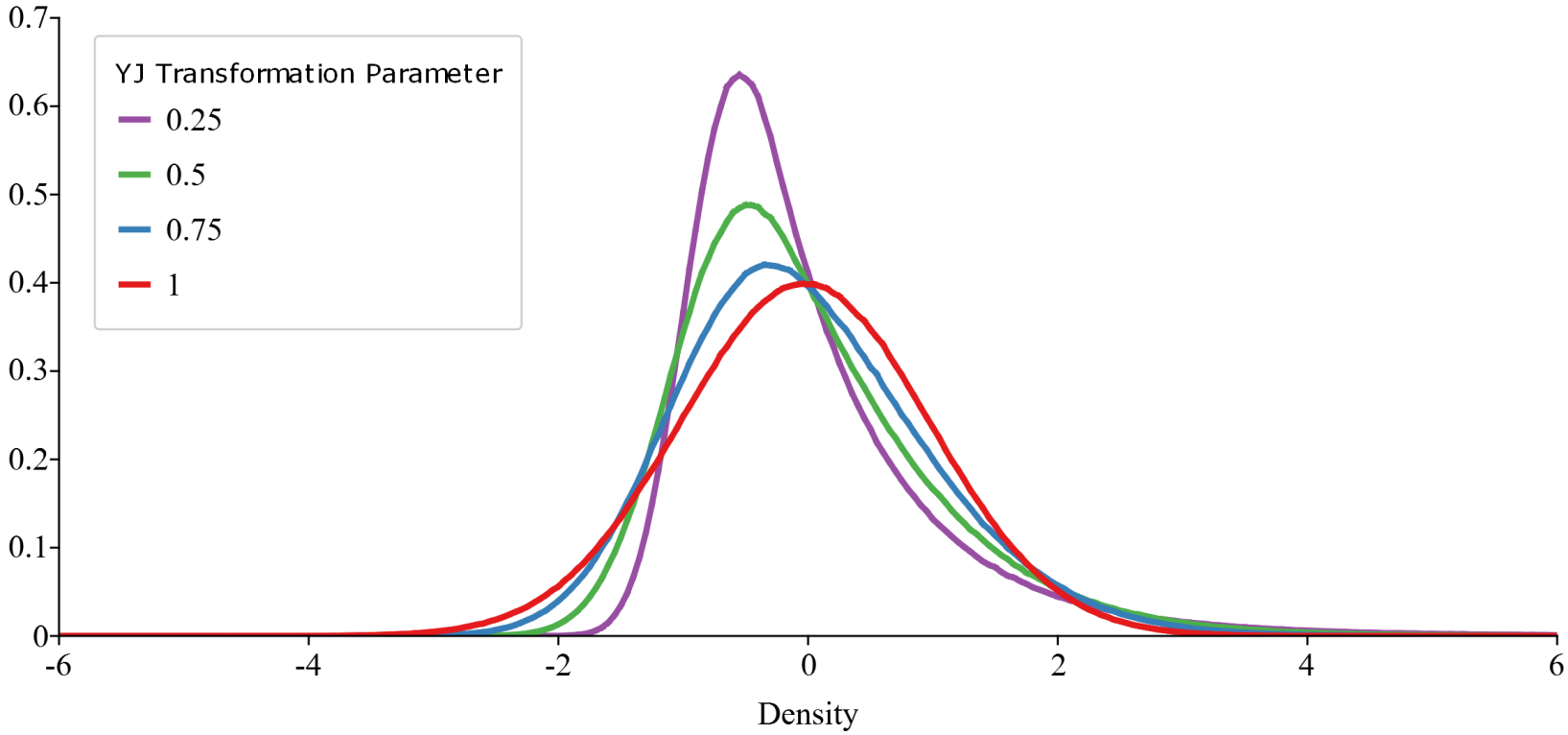

Figure 1: Density of Standardized Asymmetric Error Terms with Varying YJ Transformation Parameters

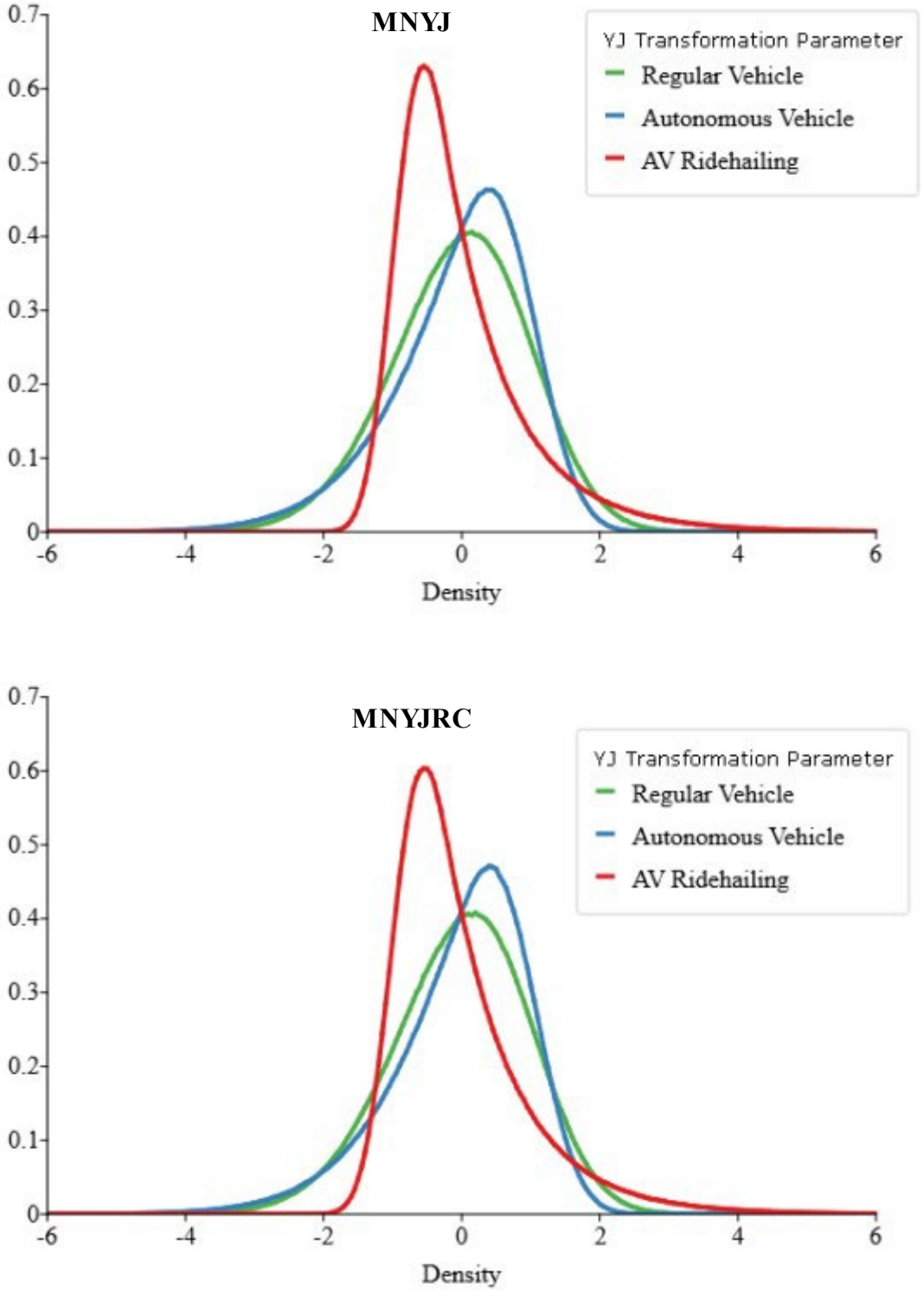

Figure 2: Distributions of Standardized Kernel Errors Based on Estimated Shape Parameters