WILEY | Hindawi

Research Article

Mixed Multinomial Probit Model Accommodating Flexible Covariance Structure and Random Taste Variation: An Application to Commute Mode Choice Behavior

Ke Wang [1,2] **Xin Ye** [3] **and Hongcheng Gan**[1,2]

[1]*Business School, University of Shanghai for Science and Technology, 200093 Shanghai, China*
[2]*Center for Supernetworks Research, University of Shanghai for Science and Technology, 200093 Shanghai, China*
[3]*Key Laboratory of Road and Traffic Engineering of Ministry of Education, College of Transportation Engineering, Tongji University, Shanghai 201804, China*

Correspondence should be addressed to Xin Ye; xye@tongji.edu.cn

This paper developed a mixed multinomial probit (MMNP) model with alternative error specification and random coefficients (for both generic variables and personal attributes) to accommodate flexible covariance structure and taste variation. The MMNP model can be efficiently estimated with analytic approximations of multivariate normal cumulative distribution functions, which avoid defects of simulation-based integration in the mixed multinomial logit (MMNL) model. The integral dimension of the MMNL model increases as random coefficients increase, but it only depends on the number of available alternatives in the MMNP model. Simulation experiments and empirical analysis of Shanghai commuters' mode choice behavior were undertaken to examine the performance of MMNP models. Both simulation results and empirical results show that MMNP models can well accommodate flexible covariance structures and taste variation reflected through random coefficients being associated with both generic and personal variables. Empirical results indicate that the MMNP model performs better than traditional discrete choice models, such as the multinomial logit, the cross-nested logit, MMNL, and multinomial probit models. Random coefficients of "in-vehicle time of car" and "number of companions" indicate taste heterogeneity and the identifiability of random coefficients associated with both generic and personal attributes. Pairwise positive correlations between car/taxi, bus/metro, and bus/bus and metro are to be expected. However, the positive correlation between the car and metro modes may be unique to the Chinese city, Shanghai, because of the developed metro system. Unequal error variances reflect heterogeneities in unspecified factors in commute modes' utilities. The MMNP model will offer an alternative efficient way to accommodate taste heterogeneity and flexible error covariance structure in discrete choice models. Compared with the MMNL model, the MMNP model can accommodate more random coefficients without increasing computational complexity.

1. Introduction

In general, two methods are being applied to overcome the limitation of the IIA property in the multinomial logit (MNL) model by accommodating random taste variation and correlation among alternatives in a discrete choice model. The first method adopts the mixed multinomial logit (MMNL) model, which combines random coefficients varying across individuals to represent taste variation with random error components varying across alternatives to represent correlation among alternatives [1–4]. The second method is a more complex mixed generalized extreme value (GEV) model, such as using a nested logit (NL) model to represent correlation among alternatives and random coefficients to represent taste variation [5–7]. The flexibility of these two methods in representing correlation and random taste variation comes at a cost that choice probabilities can no longer be expressed in a closed-form function. Compared with the first method, the benefit of the second one is that it has fewer dimensions of integration. The disadvantage is that

it has a more complex log-likelihood function. Garrow [8] compared these two methods using NL mixtures to approximate an NL model because, theoretically, mixture analogs can approximate any random utility model [9, 10]. The results show that the NL model cannot be well approximated by its mixture analog because it is difficult to obtain precise correlation estimates from mixture models, particularly for nests that have small correlations, and the accuracy of NL mixture coefficients is sensitive to the number of draws and number of random coefficients in the simulation. Empirical findings in Garrow [8] recommend adopting the second method to accommodate the correlation and random taste variation.

Although the second method, the mixed GEV models, has fewer dimensions of integration, the dimension of integration is the number of random coefficients in the model. With the increase of random coefficients, the integration will be more complicated with higher dimensions. Besides, GEV estimates are biased when negative correlations appear between choices [11]. However, modelers usually do not know the covariance structure in advance. Under this condition, the use of multinomial probit (MNP) models that can accommodate negative correlation will not cause the problem of estimation bias. In addition to negative correlations or covariances, an MNP model can also naturally accommodate unequal variances in its flexible error covariance structure. As well as accounting for random taste variation, this paper considers using a mixed multinomial probit (MMNP) model with random coefficients to represent random taste variation and its error terms to represent a flexible covariance structure among alternatives.

The MMNP model is rarely applied in the transportation field because of the difficulty to evaluate its likelihood function without a closed-form expression. Even if there are some applications, random error terms are usually assumed to be independently distributed [12, 13]. They only use correlated random coefficients to capture the correlation between alternatives. If random error terms are also correlated, will these correlations, as well as correlated random coefficients, be identifiable? Generally, in a mixed model, generic variables, which vary across alternatives, are assumed to take random coefficients. Are all the elements in a covariance-variance matrix of random coefficients for a set of generic variables (e.g., in-vehicle travel times across travel modes) identifiable? Can personal attributes, which do not vary across alternatives, take random coefficients? If so, how can the elements in a covariance-variance matrix of those random coefficients be identified? Some simulation experiments will be conducted to answer those questions in this paper.

The estimation of the MMNP model depends on the computation of a multivariate normal cumulative distribution (MVNCD) function. As the evaluation method of the MVNCD function improves, to avoid the above-mentioned defects of simulation techniques in mixed GEV models, several analytic approximation techniques have been developed [14]. Bhat [14] proposed four new analytic MVNCD approximation methods and compared them with other existing analytic approximation methods and simulation-based integration methods. Regardless of the dimensionality of the MVNCD function, the two-variate bivariate screening (TVBS) approach performs best in the ability to recover MNP parameters accurately and quickly. Therefore, we employed the TVBS analytic approximation evaluation of the MVNCD function to estimate the MMNP model for this study.

This paper aims to develop an MMNP model with correlated random coefficients (for both generic variables and personal attributes) and correlated error structure to accommodate a flexible covariance structure and taste variation, which can be efficiently estimated using analytic approximations of the MVNCD function. And the MMNP model will be applied to analyze commute mode choice behavior in Shanghai, China. The rest of this paper is organized as follows. The second section details the methodology of the MMNP model accommodating random taste variation and flexible covariance structures. In the third section, simulation experiments are undertaken to validate the identifiability when correlated random coefficients are associated with generic variables and personal characteristics. In the fourth section, an empirical study is conducted to explore Shanghai residents' commute mode choice behavior. Discussions and conclusions will be presented in the last section.

2. Modeling Methodology

In this section, the MMNP model will be formulated to accommodate correlated random coefficients and a flexible error covariance structure. For an I-alternative MMNP model, the utility function of alternative i ($i = 1, 2, \ldots, I$, where I denotes the total number of alternatives) is given by

$$U_i = \alpha_i + \beta_i x_i + \varepsilon_i = \alpha_i + \sum_{k=1}^{K} \beta_{ik} x_{ik} + \varepsilon_i, \tag{1}$$

where x_i is a column vector of K explanatory variables, and β_i is a row vector of corresponding coefficients. x_{ik} and β_{ik} are the kth elements of vectors x_i and β_i, respectively. $k = 1, 2, \cdots, K$. ε_i is the random error term for an alternative i. $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_I)'$ follows a multivariate normal distribution with a mean vector of zeros and variance-covariance

$$\text{matrix } \Sigma = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1I} \\ \rho_{12} & \rho_{22} & \cdots & \rho_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1I} & \rho_{2I} & \cdots & \rho_{II} \end{bmatrix}, \text{ or } \varepsilon \sim N(0, \Sigma).$$

β_{ik} is assumed to be a normally distributed random coefficient with mean b_{ik} and standard deviation σ_{ik}. That is, $\beta_{ik} = b_{ik} + \sigma_{ik} z_{ik}$, and herein z_{ik} follows a standard normal distribution. Then, Equation (1) can be rewritten as

$$\begin{aligned} U_i &= \alpha_i + \sum_{k=1}^{K} (b_{ik} + \sigma_{ik} z_{ik}) x_{ik} + \varepsilon_i \\ &= \left(\alpha_i + \sum_{k=1}^{K} b_{ik} x_{ik} \right) + \left(\sum_{k=1}^{K} \sigma_{ik} z_{ik} x_{ik} + \varepsilon_i \right) \\ &= V_i + \xi_i. \end{aligned} \tag{2}$$

where $V_i = \alpha_i + \sum_{k=1}^{K} b_{ik} x_{ik}$ is the systemic utility of alternative i, $\xi_i = \sum_{k=1}^{K} \sigma_{ik} z_{ik} x_{ik} + \varepsilon_i$ which is the total random utility combining the random error term ε_i and the random utility $\sum_{k=1}^{K} \sigma_{ik} z_{ik} x_{ik}$ from the random coefficient β_i. The covariance matrix of $\mathbf{z_k} = (z_{1k}, z_{2k}, \ldots, z_{Ik})'$ can be expressed as $\Lambda^k = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1I} \\ r_{12} & 1 & \cdots & r_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1I} & r_{2I} & \cdots & 1 \end{bmatrix}$.

Since $\xi_i = \sum_{k=1}^{K} \sigma_{ik} z_{ik} x_{ik} + \varepsilon_i$, mean values of z_{ik} and ε_i are zeros, the mean value of the total random utility ξ_i is also zero, $E(\xi_i) = 0$. Then, the variance-covariance matrix of $\xi = (\xi_1, \xi_2, \ldots, \xi_I)'$ can be calculated as

$$\Omega = \sum_{k=1}^{K} (\boldsymbol{\sigma_k} \cdot \boldsymbol{\sigma_k'}) * \Lambda^k * (\mathbf{x_k} \cdot \mathbf{x_k'}) + \Sigma, \qquad (3)$$

where $\boldsymbol{\sigma_k} = (\sigma_{1k}, \sigma_{2k}, \ldots, \sigma_{Ik})'$ and $\mathbf{x_k} = (x_{1k}, x_{2k}, \ldots, x_{Ik})'$, which are both column vectors. $\boldsymbol{\sigma_k'}$ and $\mathbf{x_k'}$ are the transpose of $\boldsymbol{\sigma_k}$ and $\mathbf{x_k}$, respectively. "$\cdot$" represents the matrix multiplication, and "$*$" represents Hadamard product or element-wise multiplication. Since $\boldsymbol{\beta_k} = \mathbf{b_k} + \boldsymbol{\sigma_k} \mathbf{z_k}$ (where $\mathbf{b_k} = (b_{1k}, b_{2k}, \cdots, b_{Ik})'$), the variance-covariance matrix of random coefficients $\boldsymbol{\beta_k} = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{Ik})'$ can be expressed as $\Psi^k = (\boldsymbol{\sigma_k} \cdot \boldsymbol{\sigma_k'}) * \Lambda^k$.

Based on the principle of random utility maximum (RUM), the probability that an alternative m is chosen can be expressed as

$$P_m = P(U_m > U_i) = P(\xi_i - \xi_m < V_m - V_i), \quad i = 1, 2, \ldots, I; i \neq m. \qquad (4)$$

Let $\xi_{im} = \xi_i - \xi_m$, $-V_{im} = V_m - V_i$ with $\forall i \neq m$. Then, the probability in Equation (4) can be transformed into

$$P_m = P(\xi_{im} < -V_{im})$$
$$= \int_{-\infty}^{-V_{1m}} \cdots \int_{-\infty}^{-V_{im}} \cdots \int_{-\infty}^{-V_{Im}} f(\xi_{1m}, \ldots, \xi_{im}, \ldots, \xi_{Im}) \quad (5)$$
$$d\xi_{1m} \ldots d\xi_{im} \ldots d\xi_{Im}, \quad i = 1, 2, \ldots, I; i \neq m.$$

where $f(\cdot)$ is the probability density function of an $(I-1)$-dimensional multivariate normal distribution (MVN).

$E(\xi_{im}) = E(\xi_i - \xi_m) = 0$. The covariance matrix of ξ_{im} ($i = 1, 2, \ldots, I; i \neq m$) is calculated as $\Omega^m = M_m \cdot \Omega \cdot M_m'$, where Ω is the variance-covariance matrix of $\xi = (\xi_1, \xi_2, \cdots, \xi_I)'$ as shown in Equation (3). And M_m is an identity matrix of $(I-1)$ size with an extra column of -1 added as the mth column [15]. For example, $M_m = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ if $m = 2$ and $I = 3$. M_m' is the transpose of M_m.

Therefore, the likelihood function of the maximum likelihood estimation procedure or probability of choosing an alternative m can be expressed as an $(I-1)$-dimensional multivariate normal cumulative distribution (MVNCD) function of $\xi_{im} \sim N(0, \Omega^m)$ as shown in

Equation (5). With the development of analytic approximations of the MVNCD function, the two-variate bi-variate screening (TVBS) approach is adopted to estimate the MMNP model in this paper. The TVBS approach can recover underlying parameters accurately and efficiently, regardless of the dimensionality involved in the MVNCD function. For more details of the TVBS approach, one can refer to Bhat [14].

Additionally, in an I-alternative MMNP model, $(I(I-1)/2 - 1)$ elements in the covariance matrix Σ of correlated error terms can be identified [15]. If correlated random coefficients are associated with generic variables, all means b and all elements in the variance-covariance matrix Ψ^k are identifiable. If correlated random coefficients are associated with a personal attribute that appears in utility functions of all the available alternatives, similar to an alternative-specific constant, only $(I-1)$ means of random coefficients for this personal attribute are identifiable. And the covariance matrix of random coefficients β is similar to that of error terms in identifiability. That is, $(I(I-1)/2 - 1)$ elements in the variance-covariance matrix of random coefficients Ψ^k can be identified. Simulation experiments will be conducted in the next section to demonstrate the number of identifiable elements in covariance matrices.

3. Simulation Experiments

Correlations among alternatives may exist in both correlated error structure and correlated random coefficients. To verify the identifiability of parameters, two simulation experiment scenarios were designed when correlated random coefficients were associated with generic variables and personal characteristics. The estimation of a model with independent random coefficients across alternatives is a special case of that with correlated random coefficients. All the correlations can be fixed at 0. Random error terms are correlated across alternatives in both simulation experiments.

In simulation experiments, we assume that all alternatives are available for each individual. Explanatory variables follow IID uniform distributions between 0 and 10. Dependent variables y_i are determined based on the RUM principle:

$$y_i = D[U_i = \text{Max}(U_1, U_2, \ldots, U_I)], \quad (i = 1, 2, 3, \ldots, I). \qquad (6)$$

If U_i is the maximum utility among all of the available I alternatives for an individual, $D[\cdot] = 1$; otherwise, $D[\cdot] = 0$.

3.1. Experimental Design

3.1.1. Correlated Random Coefficients Associated with Generic Variables. For a four-alternative MMNP model, we assumed that four random coefficients are associated with four independent generic variables x_1, x_2, x_3, and x_4. Utility functions of four alternatives are given as follows:

$$U_1 = \alpha_1 + \beta_1 x_1 + \varepsilon_1 = \alpha_1 + (b_1 + \sigma_1 z_1)x_1 + \varepsilon_1,$$
$$U_2 = \alpha_2 + \beta_2 x_2 + \varepsilon_2 = \alpha_2 + (b_2 + \sigma_2 z_2)x_2 + \varepsilon_2,$$
$$U_3 = \alpha_3 + \beta_3 x_3 + \varepsilon_3 = \alpha_3 + (b_3 + \sigma_3 z_3)x_3 + \varepsilon_3, \quad (7)$$
$$U_4 = \beta_4 x_4 + \varepsilon_4 = (b_4 + \sigma_4 z_4)x_4 + \varepsilon_4.$$

where random coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ are normally distributed with a mean vector $b = (b_1, b_2, b_3, b_4)'$ and a standard deviation vector $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)'$. Since random coefficients are assumed to be correlated across alternatives, z_1, z_2, z_3, and z_4 are correlated and follow a standard multivariate normal distribution with a mean vector $(0, 0, 0, 0)'$ and a covariance matrix

$$\Lambda = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{pmatrix}.$$ Then, the variance-covariance

matrix of $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ is calculated as $\Psi = (\sigma \cdot \sigma\prime) * \Lambda$. Random error terms ε_1, ε_2, ε_3, and ε_4 follow a standard multivariate normal distribution with a mean vector $(0, 0, 0, 0)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & \rho_{22} & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & \rho_{33} & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & \rho_{44} \end{pmatrix}.$$

All elements in the covariance matrix Ψ can be estimated in the MMNP model. Since only five elements in the error term covariance matrix Σ can be identified in a four-alternative MMNP model, we fixed all the diagonal elements at 1 and the correlation ρ_{34}. Thus, parameters to be estimated include all α values, all b values, all elements in the covariance matrix Ψ, and five covariance elements $(\rho_{12}, \rho_{13}, \rho_{23}, \rho_{14}, \rho_{24})$ in the covariance matrix Σ. ρ_{13} and ρ_{23} are set as negative to validate the ability of the MMNP model in accommodating negative correlation.

3.1.2. Correlated Random Coefficients Associated with Personal Attributes.
Generally, in a mixed model, random coefficients are assumed to be associated with generic variables, such as travel time, cost, or other level-of-service variables of different travel modes. The identifiability when random coefficients are associated with personal attributes has not been explored in previous studies. In this subsection, random coefficients of the same personal attribute are assumed to be correlated across alternatives and vary across individuals and alternatives. A three-alternative MMNP model with utility functions is considered as follows:

$$U_1 = \alpha_1 + \beta_1 x_1 + \varepsilon_1 = \alpha_1 + (b_1 + \sigma_1 z_1)x_1 + \varepsilon_1,$$
$$U_2 = \alpha_2 + \beta_2 x_1 + \varepsilon_2 = \alpha_2 + (b_2 + \sigma_2 z_2)x_1 + \varepsilon_2, \quad (8)$$
$$U_3 = \beta_3 x_1 + \varepsilon_3 = (b_3 + \sigma_3 z_3)x_1 + \varepsilon_3.$$

where random coefficients $\beta = (\beta_1, \beta_2, \beta_3)'$ follow normal distributions with a mean vector $b = (b_1, b_2, b_3)'$ and a standard deviation vector $\sigma = (\sigma_1, \sigma_2, \sigma_3)\prime$. Since random coefficients $\beta_1, \beta_2, \beta_3$ are correlated across alternatives, z_1, z_2 and z_3 follow a standard multivariate normal distribution

with a mean vector $(0, 0, 0)'$ and a covariance matrix
$$\Lambda = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}.$$ Then, the variance-covariance matrix
of $\beta = (\beta_1, \beta_2, \beta_3)'$ is calculated as $\Psi = (\sigma \cdot \sigma')* \Lambda$. Random error terms ε_1, ε_2, and ε_3 follow a standard multivariate normal distribution with a mean vector $(0, 0, 0)'$ and a

covariance matrix $\Sigma = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{12} & \rho_{22} & \rho_{23} \\ \rho_{13} & \rho_{23} & \rho_{33} \end{pmatrix}$.

In this case, random coefficients are associated with the same explanatory variable x_1 in utility functions of three alternatives. With consideration of the identification problem, one of the mean values of random coefficients (b_1, b_2 or b_3) should be fixed as a reference, similar to an alternative-specific constant. The covariance matrix of random coefficients β is similar to that of error terms in identifiability. In a three-alternative MMNP model, only two elements in the covariance matrix of error terms Σ can be identified. Therefore, all parameters to be estimated in this scenario are all α values, two of b values (b_1 was fixed), one covariance element r_{23} and one standard deviation σ_3 in the covariance matrix Ψ, and two covariance elements (ρ_{12} and ρ_{13}) in the covariance matrix of error terms Σ. r_{23} and ρ_{13} are set as negative to validate the ability of the MMNP model in accommodating negative correlation.

3.2. Simulation Results.
Simulation results of the MMNP models with correlated error structure and correlated random coefficients associated with generic variables and personal attributes are shown in Tables 1 and 2. Variances in both covariance matrices of error terms are assumed to be identical in the two experiments. Parameters are also consistently estimated for heteroscedastic error terms. For brevity, only the results of homoscedastic error terms are presented here.

According to significant t-statistics at the 0.01 level shown in Tables 1 and 2, all identifiable parameters can be estimated and are seemingly consistent with their true values, indicating that correlations among alternatives can be accommodated in both error structure and random coefficients for generic variables. In empirical studies, if correlations from only one source are considered, it may lead to bias in model estimation. Results in Table 2 show that, in addition to generic variables, personal attributes can also be specified with random coefficients. In real data, respondents may have different preferences for the same generic variables (such as travel time and cost of different travel modes), and individuals with different characteristics also have different preferences for alternative modes. And the MMNP model can well accommodate random taste variation in both generic variables and personal attributes through random coefficients. As significant t-statistics at the 0.01 level are shown in Tables 1 and 2, all negative covariance elements in covariance matrices of random error terms are correctly estimated, which indicates that the MMNP model can well accommodate various correlated error structures, even with negative correlations. It can perfectly avoid defects in previous MMNP models and mixed GEV models since

TABLE 1: Simulation results for a four-alternative MMNP model of 500,000 observations when random coefficients are associated with generic variables.

Parameters	True values	Estimates	Std. err.	t-stat
α_1	−0.5	−0.498	0.074	−6.776***
α_2	−0.4	−0.377	0.067	−5.640***
α_3	−0.3	−0.259	0.047	−5.551***
b_1	0.5	0.511	0.044	11.732***
b_2	0.3	0.306	0.027	11.563***
b_3	0.2	0.202	0.015	13.503***
b_4	0.1	0.099	0.007	13.431***
σ_1	0.2	0.206	0.018	11.658***
σ_2	0.4	0.404	0.035	11.722***
σ_3	0.6	0.606	0.051	11.833***
σ_4	0.5	0.528	0.045	11.667***
r_{12}	0.2	0.203	0.016	12.396***
r_{13}	0.4	0.379	0.015	25.196***
r_{14}	0.5	0.541	0.013	40.782***
r_{23}	0.6	0.608	0.008	74.205***
r_{24}	0.3	0.299	0.011	28.064***
r_{34}	0.4	0.401	0.010	42.107***
ρ_{11}	1.0	1.000	—	—
ρ_{12}	0.5	0.511	0.087	5.843***
ρ_{13}	−0.3	−0.341	0.105	−3.250***
ρ_{14}	0.6	0.524	0.080	6.521***
ρ_{22}	1.0	1.000	—	—
ρ_{23}	−0.4	−0.372	0.127	−2.932***
ρ_{24}	0.5	0.532	0.096	5.517***
ρ_{33}	1.0	1.000	—	—
ρ_{34}	0.3	0.300	—	—
ρ_{44}	1.0	1.000	—	—

Note. "—" the corresponding parameter is fixed; ***significant at 0.01 level.

TABLE 2: Simulation results for a three-alternative MMNP model of 5,000,000 observations when random coefficients are associated with personal attributes.

Parameters	True values	Estimates	Std. err.	t-stat
α_1	−0.5	−0.490	0.008	−63.097***
α_2	−0.4	−0.396	0.050	−7.894***
b_1	0.5	0.500	—	—
b_2	0.3	0.301	0.018	17.119***
b_3	0.2	0.205	0.007	30.905***
σ_1	0.2	0.200	—	—
σ_2	0.4	0.400	—	—
σ_3	0.6	0.592	0.029	20.669***
r_{12}	0.2	0.200	—	—
r_{13}	0.4	0.400	—	—
r_{23}	−0.6	−0.605	0.089	−6.768***
ρ_{11}	1.0	1.000	—	—
ρ_{12}	0.5	0.510	0.075	6.827***
ρ_{13}	−0.3	−0.313	0.079	−3.958***
ρ_{22}	1.0	1.000	—	—
ρ_{23}	0.4	0.400	—	—
ρ_{33}	1.0	1.000	—	—

Note. "—" the corresponding parameter is fixed; ***significant at 0.01 level.

error terms are assumed to be independent in previous MMNP models [12, 13], and negative correlations cannot be accommodated in the mixed GEV models [15], which may lead to bias in model estimation [11].

However, the sample size for ensuring the significance of all the identifiable parameters is great in MMNP models. For the four-alternative MMNP model with random coefficients associated with generic variables, the sample size is around 500,000 in our experiment to ensure the significance of all the critical coefficients. The sample size for the three-alternative MMNP model with random coefficients of personal attributes is around 5,000,000 to ensure significance. Both sample sizes are much larger than those in usual empirical studies. As per Walker [16], although model parameters are theoretically identifiable, there are issues of empirical identification, namely, whether the data can support the model specification. If the empirical data cannot allow for estimating all the theoretically identifiable parameters, some parameters may be fixed, and only parameters of particular interest need to be estimated.

To improve the efficiency of estimation and decrease the required sample size for statistically significant identification, z_1, z_2, z_3, and z_4 in the random coefficients of the four-alternative MMNP model can be set as $z_1 = z_2 = z_3 = z_4$. That is, z_1, z_2, z_3, and z_4 are linearly correlated with a correlation coefficient of 1. And they follow the standard normal distribution. Then, the required sample size for identification can be decreased to 5000. The simulation results with a sample size of 5000 observations are shown in Table 3.

4. Empirical Study

To examine the performance of the MMNP model in accommodating random taste variation and correlations among alternatives, an application is presented in this section to analyze the commute mode choice behavior of Shanghai residents.

4.1. Data and Sample Description

4.1.1. Commute Trip Data. The commute trip data used in this study were derived from a web-based travel survey of Shanghai conducted in 2018. The survey collected detailed socioeconomic and demographic information, and a complete 24-hour travel diary reported by 2033 individuals. The detailed trip data included origin and destination locations, trip beginning and ending times, travel mode, trip purpose, and the number of companions. Six travel modes are considered in this study including car, taxi, metro, bus, a combination of bus and metro (hereinafter referred to as Bus and Metro), and the nonmotorized mode (bicycling and walking). After data screening to remove records with missing data for explanatory variables of interest and travel mode choice, the final sample comprised 1743 commute trips.

4.1.2. Level-of-Service Data. The LOS characteristics of different travel modes were extracted from the zone-to-zone travel impedance matrices generated from transportation networks of Shanghai, which were integrated into Trans-CAD based on the GIS data of roads (links and nodes), bus

Table 3: Simulation results for a four-alternative MMNP model of 5,000 observations when random coefficients are associated with generic variables.

Parameters	True values	Estimates	std. err.	t-stat
α_1	−0.5	−0.792	0.192	−4.126***
α_2	−0.4	0.561	0.066	8.481***
α_3	−0.3	−0.714	0.204	−3.500***
b_1	0.5	0.326	0.040	8.154***
b_2	0.3	−0.455	0.218	−2.091**
b_3	0.2	0.188	0.034	5.488***
b_4	0.1	0.062	0.032	1.943*
σ_1	0.2	0.199	0.024	8.410***
σ_2	0.4	0.443	0.049	8.995***
σ_3	0.6	0.660	0.069	9.521***
σ_4	0.5	0.580	0.060	9.746***
r_{12}	1.0	1.000	—	—
r_{13}	1.0	1.000	—	—
r_{14}	1.0	1.000	—	—
r_{23}	1.0	1.000	—	—
r_{24}	1.0	1.000	—	—
r_{34}	1.0	1.000	—	—
ρ_{11}	1.0	1.000	—	—
ρ_{12}	0.5	0.476	0.159	2.992***
ρ_{13}	−0.3	−0.340	0.125	−2.733***
ρ_{14}	0.6	0.628	0.137	4.584***
ρ_{22}	1.0	1.000	—	—
ρ_{23}	−0.4	−1.148	0.441	−2.602***
ρ_{24}	0.5	0.430	0.206	2.089**
ρ_{33}	1.0	1.000	—	—
ρ_{34}	0.3	0.300	—	—
ρ_{44}	1.0	1.000	—	—

Note. "—" the corresponding parameter is fixed; ***significant at 0.01 level, **significant at 0.05 level, *significant at 0.1 level.

lines and stops, metro lines and stations, and residential committees. Travel time calculated from floating car data was added to each link of the road network [17, 18]. Travel times, service frequencies, and fares by the time of day were applied to each trip OD pair in the sample.

4.1.3. Sample Description.
Table 4 provides a summary of the socioeconomic and demographic characteristics of the sample. Most commuters are 20–40 years old, married, and well educated, hold a driving license, and live with family. Personal monthly income mainly falls between 4.5k and 15k RMB Yuan. More than half (62.2%) of commuters have an available private car on the travel day. Origin TAZs (or residential areas) have greater population density and lower employment density than destination TAZs (or workplaces), as expected.

Descriptive statistics of LOS characteristics are shown in Table 5. The average trip distance of nonmotorized mode is 4.204 km, which indicates that nonmotorized mode can meet most of the short-distance commute demands. In terms of average fare, the taxi has the highest fare because Shanghai taxi has a high starting price. In terms of average in-vehicle time, car, taxi, and metro are all around 20 minutes, while the bus and Bus and Metro rise obviously to 37.999 and 38.747 minutes, respectively. In the Bus and Metro mode, the in-vehicle time of the metro is longer than

that of the bus, indicating that the bus is a complementary mode to the metro. The average initial waiting time and transfer waiting time of the metro are shorter than those of bus and Bus and Metro, and they vary much less, meaning that the metro system provides a higher service frequency, and the service frequency distribution is more even among different lines. In terms of average access/egress distance, bus and Bus and Metro have substantially shorter distances than the metro because the bus has a higher station coverage rate. And metro's average access distance is slightly longer than the average egress distance presumably due to the denser distribution of metro stations in workplaces than in residential areas. The Bus and Metro mode takes a slightly greater number of transfers than bus and metro, as expected.

4.2. Empirical Results.
In this section, the empirical results of models, including the MNL, CNL (cross-nested logit model, shown in Figure 1), MMNL, MNP, and MMNP models, are presented in Table 6. And then, the data fit of the MMNP model is compared with that of the MNL, CNL, MMNL, and MNP models.

The MMNP model can accommodate both random taste heterogeneity and flexible covariance structure among alternatives. Compared to the MMNP model, the MNL, the CNL, and the MNP models cannot accommodate random coefficients. Although the MMNL model can capture the taste variation across individuals by random coefficients, it cannot accommodate the variance heterogeneity of different travel modes. The following paragraph will focus on model results of the MMNP model because coefficients of explanatory variables have the same signs and similar values.

4.2.1. Model Results of the MMNP Model.
(1) Random Coefficients. In the empirical study, both random coefficients of a generic variable and a personal attribute appear to be significant, indicating that there are variations among commuters' sensitivities to those variables.

As a generic variable, the in-vehicle time of the car has a mean coefficient of −2.662 and a standard deviation of 1.050. The negative mean value indicates that the utility of the car will decrease with the increase of in-vehicle time on average across individuals, as expected. The value of standard deviation is different from zero significantly, implying that random taste heterogeneity does exist among commuters in this LOS variable. And the value of standard deviation in this MMNP model is quite similar to that in MMNL models [19, 20].

"The number of companions" has a random coefficient with a mean value of 0.423 and a standard deviation of 0.828 in the utility function of the car mode. The positive mean value indicates that "the number of companions," on average, has a positive impact on the choice of car because commuters can share the cost while enjoying the flexibility and mobility of the car, which is more cost-effective than driving alone. The standard deviation is significantly different from zero. "The number of companions" is a kind of personal attribute because the variable remains the same in all alternatives' utility functions. Therefore, it verifies that the

TABLE 4: Sample characteristics ($N = 1743$).

Categorical variables	Categories	Percentage (%)
Gender	Male	45.439
	Female	54.561
Age	Younger than 20 years	2.811
	21–30 years	34.079
	31–40 years	46.242
	41–50 years	13.827
	Above 50 years	3.041
Marital status	Married	72.060
	Unmarried	27.940
Driving license	Own	75.961
	Does not own	24.039
Residential type	Family	87.722
	Dormitory	11.819
	Other	0.459
Education attainment	Less than high school	1.549
	High school	6.311
	Bachelor's degree	78.543
	Graduate degree	13.597
Personal monthly income (RMB yuan)	Less than 2k	2.754
	2k–4.5k	8.319
	4.5k–6k	14.974
	6k–8k	17.843
	8k–10k	20.023
	10k–15k	20.367
	15k–20k	7.975
	20k–30k	5.221
	Greater than 30 k	2.524
Private vehicles	Car	62.192
	Electric bicycle	29.260
	Bicycle	30.293
	Motorcycle	2.008

Continuous variables		Mean	S.D.
Personal income[1] (10k RMB Yuan/month)		1.018	0.638
Population density (thousand people per square kilometer)	Origin TAZ	2.538	2.178
	Destination TAZ	1.252	1.855
Employment density (thousand people per square kilometer)	Origin TAZ	1.794	1.994
	Destination TAZ	2.367	2.704

(1) The continuous variable "Personal Income" was generated from the categorical variable "Personal Monthly Income." The exchange rate of RMB Yuan to the US dollar is approximately 6.9 : 1.

random coefficient of a personal attribute can also be estimated in an empirical study.

(2) Error Covariance Structure. As shown in Table 6, the CNL, the MNP, and the MMNP models all reveal positive correlations for the pairwise alternatives of car/taxi, car/metro, bus/metro, and bus/bus and metro. In the context of single choice models (that is, when only one alternative can be chosen), such positive correlations imply a higher level of substitutability between the corresponding pairs of modes relative to assuming zero correlation (as in the MNL).

Of the pairwise correlations, those between car/taxi, bus/metro, and bus/bus and metro are to be expected. Cars and taxis are characterized by their mobility and flexibility, and both belong to the category of a private transportation mode. Meanwhile, bus and metro are both public transportation modes and have many aspects in common. However, the positive correlation between the car and metro modes, even though less than the positive correlations between car/taxi and bus/metro, is particularly interesting because one is a

private mode, while the other is a public transportation mode. This result, though perhaps likely to be unique to many Chinese cities, is probably because the Shanghai metro has been well developed and provides a convenient, high-frequency, and punctual service. Shanghai metro operates on a strict schedule, making it attractive to most commuters. The Shanghai metro operates at a speed (almost 30 km/h) even faster than the speed of road traffic during peak hours. In addition, it is difficult to find a temporary parking lot during peak hours, and it is also highly expensive to reserve a parking lot because of limited parking spaces around workplaces. This finding suggests that improvement in metro LOS can be effective in attracting commuters away from private cars to the metro and alleviate traffic congestion during peak hours.

The variance of the bus mode is also shown in Table 6. Due to the limited sample size, not all identifiable parameters appear to be significant. Variances of the other modes (car, taxi, metro, bus and metro, and nonmotorized modes)

TABLE 5: Descriptive statistics of LOS (level-of-service) attributes.

|	Attribute	Mean	S.D.
CAR	Fare (RMB yuan)	7.065	6.471
	In-vehicle time (min)	23.406	16.604
Taxi	Fare (RMB yuan)	35.490	28.880
	In-vehicle time (min)	23.406	16.604
Metro	Fare (RMB yuan)	3.938	0.948
	In-vehicle time (min)	20.469	14.714
	Initial waiting time (min)	2.164	0.317
	Transfer waiting time (min)	1.226	1.455
	Access distance (km)	1.826	1.127
	Egress distance (km)	1.327	0.978
	Number of transfers	0.568	0.669
Bus	Fare (RMB yuan)	3.458	1.707
	In-vehicle time (min)	37.999	29.897
	Initial waiting time (min)	4.281	2.751
	Transfer waiting time (min)	3.152	4.501
	Transfer walking distance (km)	0.110	0.240
	Access distance (km)	0.657	0.427
	Egress distance (km)	0.638	0.430
	Number of transfers	0.729	0.854
Bus and metro	Fare (RMB yuan)	3.610	1.626
	Total in-vehicle time (min)	38.747	18.981
	In-vehicle time allocated to metro (min)	22.060	14.584
	In-vehicle time allocated to bus (min)	16.687	14.436
	Initial waiting time (min)	3.599	2.658
	Access distance (km)	0.720	0.441
	Egress distance (km)	0.679	0.407
	Number of transfers	0.994	1.026
	Transfer waiting time (min)	2.833	3.556
	Transfer walking distance (km)	0.126	0.258
Nonmotorized mode	Trip distance (km)	4.204	5.611
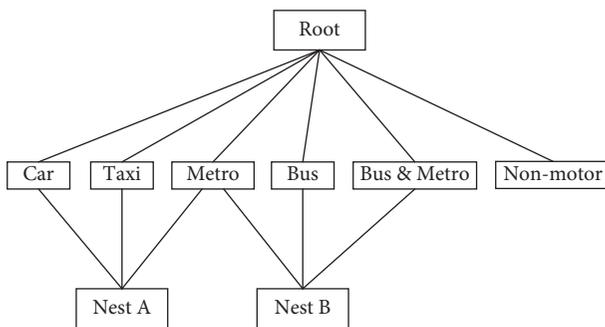
FIGURE 1: Nesting structure of the CNL model.

are fixed at 1 as a base. Consistent with the results in Bhat [21], the variance of the bus mode is greater than that of car, taxi, and metro, indicating that unobserved variables affecting the choice of bus have greater variances than those affecting car, taxi, and metro. For example, the comfort level is an unobserved variable whose values may vary considerably within bus modes because different kinds of bus services in Shanghai, such as customized buses (or demand-response transit), bus rapid transit (BRT), and regular buses, may have different occupancy rates. However, comfort varies little in the car, taxi, and metro. Therefore, the random component of the bus mode has greater variance than that of the other modes.

(3) Level-of-Service Variables. The in-vehicle time takes expected negative effects on the choice of car, taxi, metro, and bus. Taxi has the most negative coefficient value because the taxi fare ranks first among all the modes. Commuters can spend their in-vehicle time entertaining and do not have to focus on traffic conditions when they take public transit. As a result, coefficients of in-vehicle time for metro and bus are less than those for cars and taxis.

Access/egress distance reduces utilities of public transit (metro, bus, and Bus and Metro), indicating that longer access/egress distance will discourage commuters from choosing public transit. And the coefficient of access/egress distance for the metro is lower than that for bus, Bus and Metro. Because the bus has a higher station coverage rate than the metro, and the metro has a much more comfortable environment than the bus, commuters are more likely to accept the larger access/egress distance of the metro.

A negative coefficient of "the number of transfers" implies that the greater the number of transfers is, the less likely the commuters are to take the bus. Usually, the bus is not as punctual as the metro. The more the number of transfers is, the greater the probability that commuters will encounter bus delays. Besides, the transferring environment at public transit stations is crowded and uncomfortable. Commuters are usually anxiously waiting or in a hurry.

TABLE 6: Empirical results of models ($N = 1743$).

Variable	MNL Coefficient	MNL t-statistic	CNL Coefficient	CNL t-statistic	MMNL Coefficient	MMNL t-statistic	MNP Coefficient	MNP t-statistic	MMNP Coefficient	MMNP t-statistic
Alternate specific constants (nonmotorized mode is base)										
Car	−5.312	−12.38	−5.320	−12.42	−6.052	−11.23	−3.578	−5.89	−4.517	−6.58
Taxi	−4.834	−9.31	−4.840	−10.49	−4.679	−10.20	−3.023	−8.18	−2.888	−8.52
Metro	0.236	0.78	0.277	0.83	0.192	0.64	0.755	2.18	0.674	2.12
Bus	−0.226	−0.89	−0.258	−0.93	−0.255	−0.98	−0.014	−1.14	−0.607	−1.25
Bus and metro	−0.592	−1.32	−0.575	−1.19	−0.599	−1.33	−1.556	−2.47	−1.111	−1.85
Level-of-service variables										
In-vehicle time (h)										
Car										
Mean	−2.846	−6.28	−2.862	−6.68	−2.877	−6.59	−2.452	−3.88	−2.662	−4.06
Standard deviation	—	--	—	--	0.427	0.55	—	--	1.050	3.75
Taxi	−3.238	−4.90	−3.252	−4.24	−3.486	−4.28	−2.938	−4.38	−3.853	−4.31
Metro	−1.284	−2.71	−1.284	−2.78	−1.354	−2.99	−1.484	−2.53	−1.516	−2.75
Bus	−0.951	−3.05	−1.068	−3.10	−0.995	−3.28	−0.654	−1.87	−0.710	−1.84
Access/Egress distance (km)										
Metro	−0.529	−7.97	−0.552	−7.56	−0.533	−7.90	−0.501	−3.94	−0.480	−4.21
Bus	−0.550	−3.81	−0.563	−3.55	−0.561	−3.84	−0.590	−2.28	−0.690	−2.50
Bus and metro	−0.702	−3.22	−0.755	−2.86	−0.711	−2.93	−0.357	−2.14	−0.771	−2.11
The number of transfers-bus	−0.813	−5.13	−0.866	−4.71	−0.826	−5.12	−1.074	−2.86	−1.196	−2.96
Transfer waiting time (min), Bus and metro	−0.237	−4.87	−0.245	−4.71	−0.244	−4.99	−0.133	−2.66	−0.267	−3.06
Trip distance (km), nonmotorized mode	−0.256	−14.68	−0.255	−11.47	−0.261	−11.63	−0.178	−4.73	−0.185	−4.83
Sociodemographic and employment variables										
Married										
Car	0.554	3.21	0.557	3.27	0.547	2.99	0.331	2.28	0.314	1.94
Personal monthly income (10K RMB yuan)										
Car	0.812	7.32	0.828	7.13	0.841	6.97	0.620	4.09	0.718	4.73
Taxi	1.182	4.18	1.190	5.14	1.162	4.99	0.840	3.63	0.868	4.08
Metro	0.373	2.63	0.397	2.61	0.373	2.58	0.229	2.01	0.268	1.67
The number of companions										
Car										
Mean	0.234	6.15	0.232	3.62	0.494	4.69	0.149	3.43	0.423	3.75
Standard deviation	—	—	—	—	0.728	3.12	—	—	0.828	5.21
Taxi	0.341	3.07	0.339	3.46	0.174	2.17	0.191	2.48	0.150	2.07
Having a driving license										
Car	2.811	9.25	2.810	9.27	3.191	8.48	1.917	5.51	2.412	5.33
Living with family										
Car	1.315	4.29	1.310	4.44	1.556	4.58	1.026	3.70	1.323	4.70
Having available private cars on the trip day										
Metro	−0.775	−4.47	−0.819	−4.73	−0.736	−4.41	−0.896	−2.86	−0.852	−2.94
Bus	−0.861	−5.04	−0.919	−4.82	−0.818	−4.84	−1.477	−2.75	−1.509	−2.82
Having available private bikes on the trip day										
Nonmotorized mode	0.719	5.01	0.719	5.25	0.741	5.29	0.659	3.81	0.695	4.26

Table 6: Continued.

Variable	MNL		CNL		MMNL		MNP		MMNP	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Workplace employment density (1k positions/km2)										
Taxi	0.156	2.35	0.157	2.90	0.165	2.99	0.101	2.82	0.079	1.94
Metro	0.119	4.05	0.118	3.99	0.126	4.36	0.103	2.70	0.103	2.85
Bus	0.109	3.49	0.102	2.84	0.116	3.68	0.063	1.98	0.082	1.92
Error covariance Structure										
Var (Bus)	—	—	—	—	—	—	2.457	4.32	2.646	2.96
Corr (Car, taxi)	—	—	0.662	3.82	—	—	0.835	2.37	0.765	2.28
Corr (Car, metro)	—	—	0.200	3.26	—	—	0.436	2.05	0.364	2.55
Corr (Bus, metro)	—	—	0.631	7.47	—	—	0.681	2.77	0.617	3.69
Corr (Bus, bus and metro)	—	—	0.662	6.94	—	—	0.643	2.84	0.629	2.54

TABLE 7: Measures of model performance.

Goodness-of-fit measures	MNL	CNL	MMNL	MNP	MMNP
LL (β)	−1764.631	−1757.368	−1755.445	−1743.360	−1733.051
LL (0)	−2659.259	−2659.259	−2659.259	−2659.259	−2659.259
ρ^2 (0)	0.336	0.339	0.340	0.344	0.348
Adj. ρ^2 (0)	0.326	0.327	0.328	0.332	0.335
AIC	3587.261	3580.736	3572.890	3554.720	3538.103
BIC	3745.699	3761.027	3742.254	3740.474	3734.784
Non-nested test	5.71E-17	3.25E-14	1.89E-11	2.89E-07	--

Similarly, the coefficient of "transfer waiting time" is also negative for Bus and Metro.

While the nonmotorized mode mainly serves short-distance commute trips as shown in Table 5, "trip distance" takes a significantly negative coefficient, as expected.

(4) Sociodemographic and Employment Variables. Married commuters tend to drive a car because they usually need to drive their spouse to work or drive children to school during the commute trip. Commuting by car can provide maximum flexibility and convenience. Likewise, commuters living with families also prefer to use a car. "Personal monthly income" has positive coefficients in utilities of car, taxi, and metro because high-income commuters tend to use high-quality, but high-cost services. With the emerging car-sharing services, having a driving license becomes the only requirement for commuters to drive a car as the commute mode. Therefore, having a driving license will increase the likelihood of commuting by car. Compared with using a car-sharing service, "having available private cars on the trip day" will encourage commuters to drive a car to commute because commuters do not have to walk to car-sharing stations from home and return the car to designated stations. On the other hand, "having available private cars on the trip day" will decrease the likelihood of taking public transit. "Having available private bikes on the trip day" provides a more convenient opportunity for commuters to commute by the nonmotorized mode as evidenced by a positive coefficient. Higher employment densities of workplaces indicate lower parking facility densities and higher parking fees. Commuters working in these areas will be less likely to drive to work, and the likelihood of taking a taxi, bus, and metro will increase.

4.2.2. Comparisons of Data Fit with the MNL, CNL, MMNL, and MNP Models. To evaluate the data fit of the MMNP model, other four models were also estimated: a basic six-alternative MNL model, a CNL model with nesting structure shown in Figure 1, an MMNL model with random coefficients associated with the same variables (in-vehicle time of the car, and the number of companions) as in the MMNP model, and an MNP model with the same correlated error structure as in the MMNP model.

To compare the model fitting performance of the MMNP model with those of the other four models, multiple goodness-of-fit measures are calculated, including ρ^2 (0), adjusted ρ^2 (0), Akaike Information Criteria (AIC), and Bayesian Information Criterion (BIC):

$$\rho^2(0) = 1 - \frac{LL(\beta)}{LL(0)},$$
$$adj.\rho^2(0) = 1 - \frac{LL(\beta) - R}{LL(0)}. \quad (9)$$

where $LL(0)$ is the log-likelihood value of the model without any explanatory variables. $LL(\beta)$ is the log-likelihood value at convergence.

$$AIC = 2(R - LL(\beta)),$$
$$BIC = -2LL(\beta) + R \cdot \ln(N). \quad (10)$$

where R is the number of parameters estimated in the model, and N is the number of observations.

Further, to compare the logit-based models with the MMNP model using a statistical test, a nonnested test statistic is applied (see [22–24]). The adjusted log-likelihood ratio statistic of each model is first computed:

$$\overline{\rho}^2 = 1 - \frac{LL(\beta) - R/2}{LL(0)}. \quad (11)$$

The nonnested test statistic is

$$P\left(\overline{\rho}_1^2 - \overline{\rho}_2^2 > z\right) \le \Phi\left(-\sqrt{-2 \cdot LL(0) \cdot z}\right). \quad (12)$$

where Φ represents the CDF of standard normality, and z takes a positive value. That is, the probability that $(\overline{\rho}_1^2 - \overline{\rho}_2^2 > z)$ could have occurred by chance is no larger than $\Phi(-\sqrt{-2 \cdot LL(0) \cdot z})$. A small value for the nonnested test statistic indicates a small probability of erroneously selecting the incorrect model. And the model with a higher value for the adjusted likelihood statistic is preferred. The nonnested test is then used to compare the MMNP model with the other four models.

Based on ρ^2 (0), adjusted ρ^2 (0), the model with the highest value performs best. And the model with the smallest values of AIC and BIC has the best performance. As shown in Table 7, all the measures indicate the same result that the MMNP model performs better than the other four models with the highest values of ρ^2 (0) and adjusted ρ^2 (0), and the smallest AIC and BIC values. It indicates that the MMNP can not only accommodate random taste variation and flexible covariance structures among alternatives, but also perform better than the traditional discrete choice models. From the informal nonnested likelihood statistic values provided in Table 7, it can be inferred that the probability of erroneously selecting the MMNP model is literally zero.

5. Discussion and Conclusion

This paper developed an MMNP model with correlated error structure and random coefficients for both generic variables and personal attributes to accommodate flexible covariance structure and random taste variation. And the MMNP model can be efficiently estimated with analytic approximations of the multivariate normal cumulative distribution (MVNCD) function proposed by Bhat [14]. Both simulation experiments and empirical analysis of Shanghai commuters' mode choice behavior are carried out to examine the performance of the MMNP model.

Compared to the MMNP model, the MNL, the CNL, and the MNP model cannot accommodate random coefficients. Although the MMNL model can capture the taste variation across individuals by random coefficients, it cannot accommodate the variance heterogeneity of different alternatives. Simulation results show that the MMNP model can accommodate a flexible covariance structure, even with negative correlations that cannot be accommodated in mixed GEV models. However, the neglect of negative correlation in mixed GEV models may lead to estimation bias. Correlations among alternatives can exist in both error structure and random coefficients for generic variables, while most previous applications of MMNP models assumed that random error terms were independent of each other, which may also lead to bias in model estimation. In addition to correlated random coefficients for generic variables, correlated random coefficients for personal attributes can also be significantly identified.

Empirical results are consistent with simulation results in the fact that random coefficients can exist in both generic variables and personal attributes in MMNP models. As a generic variable, the in-vehicle time of the car has a negative mean coefficient and a standard deviation being significantly different from zero. "The number of companions" remains the same in all alternatives' utility functions and can be regarded as a kind of personal attribute. The random coefficient of "the number of companions" has a positive mean and a standard deviation being significantly different from zero. Simulation results show that all theoretically identifiable elements in variance-covariance matrices of correlated random coefficients can be estimated with a large enough sample size, while in empirical studies, the real data may not allow for estimating all these parameters due to limited sample size or little variation in explanatory variables. In these cases, some parameters may be fixed, and only parameters of interest can be estimated. The variance and correlation structures from the MMNP model provide useful insights. A greater variance for bus than metro indicates much higher variability in the quality of bus-related equipment, bus stop environment, and bus service relative to the more streamlined and less variable metro rail service. Interestingly, whenever correlations were allowed in the models, these turned out to be positive. Of the pairwise positive correlations, those between car/taxi, bus/metro, and bus/bus and metro are to be expected. However, the positive correlation between the car and metro modes, even though less than the positive correlations between car/taxi, bus/ metro, and bus/bus and metro, is particularly interesting

because one is a private mode, while the other is a public transportation mode. But this result, though perhaps likely to be unique to many Chinese cities, suggests that improvement in the metro level of service can be effective in attracting commuters from private cars to the metro and alleviate traffic congestion during peak hours.

Empirical results also show that the MMNP model has a better model fit than traditional discrete choice models, such as the MNL, MMNL, and MNP models, evidenced by the highest values of $\rho^2(0)$ and adjusted $\rho^2(0)$, and the smallest AIC and BIC values. And the nonnested test statistics show that the probability of erroneously selecting the MMNP model is literally zero. A model comparison could also be achieved with a synthetic population of larger size. Researchers and practitioners may not get access to large samples of available travel data in practice. To better test the performance of these models in the real world, this paper uses a small sample of real data.

Combined with analytic approximations of MVNCD functions, the MMNP model can be estimated more efficiently and quickly and avoid the high-dimensional simulation integration of using the traditional MMNL model or the mixed NL model with random coefficient to relax the response homogeneity assumption and the independence assumption of the MNL model. In summary, the MMNP model can well accommodate random taste variation and flexible covariance structure in both simulation experiments and empirical study and can be efficiently estimated by employing analytic approximations of MVNCD functions. The developed MMNP model will offer researchers and practitioners another simple and efficient way to accommodate taste heterogeneity and flexible covariance structure. In this paper, random coefficients are assumed to be normally distributed, but the possibility to develop an MMNP model based on different distributional assumptions may be explored in future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors confirm their contribution to the paper as follows. Xin Ye and Ke Wang conceptualized and designed the study; Ke Wang collected the data; Ke Wang and Xin Ye interpreted and analyzed the results; Ke Wang, Xin Ye, and Hongcheng Gan prepared the draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

References

[1] D. Brownstone and K. Train, "Forecasting new product penetration with flexible substitution patterns," *Journal of Econometrics*, vol. 89, no. 1-2, pp. 109–129, 1998.

[2] E. Cherchi and J. Ortuzar, "Alternative specific variables in non-linear utility functions: influence of correlation, homoscedasticity and taste variations," in *Proceedings of the Presented at the the 10th Intenational Conference on Travel Behaviour Research Luceme,* Switzerland, 2003.

[3] S. Hess and J. W. Polak, "Mixed Logit modelling of parking type choice behaviour," *Transportation Statistics*, pp. 77–102, 2009.

[4] D. Revelt and K. Train, *Customer-specific Taste Parameters and Mixed Logit (Working Paper)*, Department of Economics, University of Califomia, Berkeley, CL, USA, 2000.

[5] C. R. Bhat and J. Guo, "A mixed spatially correlated logit model: formulation and application to residential choice modeling," *Transportation Research Part B: Methodological*, vol. 38, no. 2, pp. 147–168, 2004.

[6] M. Chernew, G. Gowrisankaran, and D. P. Scanlon, "Learning and the value of information: evidence from health plan report cards," *Journal of Econometrics*, vol. 144, no. 1, pp. 156–174, 2008.

[7] S. Hess, M. Bierlaire, and J. Polak, "Capturing taste heterogeneity and correlation structure with mixed GEV models," in *Proceedings of the Infoscience. Presented at the 8th Annual Meeting of the Transportation Research Board8th Annual Meeting of the Transportation Research Board*, 2005.

[8] L. A. Garrow, *Comparison of Choice Models Representing Correlation and Random Taste Variation: An Application to Airline Passengers' Rescheduling Behavior*, Northwestern University, Evanston, IL, USA, 2004.

[9] S. R. Dalal and R. W. Klein, "A flexible class of discrete choice models," *Marketing Science*, vol. 7, no. 3, pp. 232–251, 1988.

[10] D. McFadden and K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.

[11] H. Dong, E. Ben-Elia, C. Cirillo, T. Toledo, and J. N. Prashker, "On negative correlation: a comparison between Multinomial Probit and GEV-based discrete choice models," *Transportmetrica: Transportation Science*, vol. 13, no. 4, pp. 356–379, 2017.

[12] C. R. Bhat and R. Sidharthan, "A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 940–953, 2011.

[13] N. Eluru, V. Chakour, M. Chamberlain, and L. F. Miranda-Moreno, "Modeling vehicle operating speed on urban roads in Montreal: a panel mixed ordered probit fractional split model," *Accident Analysis & Prevention*, vol. 59, pp. 125–134, 2013.

[14] C. R. Bhat, "New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function," *Transportation Research Part B: Methodological*, vol. 109, pp. 238–256, 2018.

[15] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, England, 2009.

[16] J. Walker, "Mixed logit (or logit kernel) model: dispelling misconceptions of identification," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1805, no. 1, pp. 86–98, 2002.

[17] R. Zhang, X. Ye, K. Wang, D. Li, and J. Zhu, "Development of commute mode choice model by integrating actively and passively collected travel data," *Sustainability*, vol. 11, no. 10, 2019.

[18] R. Zhang, K. Wang, J. Zhu, and X. Ye, "Development of mode choice model to understand why auto owners use rail for commute in Shanghai, China," in *Proceedings of the 17th COTA International Conference of Transportation Professionals*, 2017.

[19] C. R. Bhat, "Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling," *Transportation Science*, vol. 34, no. 2, pp. 228–238, 2000.

[20] J. Shen, Y. Sakata, and Y. Hashimoto, "A comparison between latent class model and mixed logit model for transport mode choice: evidences from two datasets of Japan," *Economics And Business*, vol. 6, 2006.

[21] C. R. Bhat, "Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice," *Transportation Research Part B: Methodological*, vol. 32, no. 7, pp. 455–466, 1998.

[22] M. E. Ben-Akiva and S. R. Lerman, "Discrete choice analysis: theory and application to travel demand," MIT Press, Cambridge, England, 1985.

[23] K. Wang, X. Ye, and J. Ma, "An empirical analysis of post-work grocery shopping activity duration using modified accelerated failure time model to differentiate time-dependent and time-independent covariates," *PLoS One*, vol. 13, no. 11, Article ID e0207810, 2018.

[24] K. Wang and X. Ye, "Development of alternative stochastic Frontier models for estimating time-space prism vertices," *Transportation*, vol. 48, no. 2, pp. 773–807, 2021.