WILEY | Hindawi

Research Article

Development of Integrated Choice and Latent Variable (ICLV) Models Using Matrix-Based Analytic Approximation and Automatic Differentiation Methods on TensorFlow Platform

Jie Ma [ID]**, Xin Ye** [ID]**, and Kun Huang** [ID]

Key Laboratory of Road and Traffic Engineering of Ministry of Education College of Transportation Engineering, Tongji University, Shanghai, China

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

This study further explores the multinomial probit-based integrated choice and latent variable (ICLV) models. The LDLT matrix-based analytic approximation methods, including Mendell and Elston (ME) method, bivariate ME (BME) method, and two-variate bivariate screening (TVBS) method, were adapted to calculate the multivariate cumulative normal distribution (MVNCD) function in the ICLV model because of the better performances in accuracy and computational time. Integrated with the composite marginal likelihood (CML) estimation approach, the ICLV model based on high-dimensional integration can be estimated accurately within a reasonable time. In this study, some three-alternative and four-alternative ICLV models are simulated to examine their abilities to recover model parameters. It is found that the parameter estimates and standard error estimates are acceptable for both models and the computational time is expected to decrease using tensor data structures on the TensorFlow platform. For the four-alternative ICLV models, the TVBS method has the highest level of accuracy. The BME method is also a good alternative to TVBS if computational time is of great concern. The application of the automatic differentiation (AD) technique in the model can free researchers from coding analytical gradients of log-likelihood functions and thereby greatly reduce the workload of researchers.

1. Introduction

In the discrete choice model (DCM), the choice depends not only on the transport characteristics and the observable characteristics of decision-makers but also on the unobservable characteristics, namely latent variables. Among them, psychological and attitudinal factors as parts of unobservable latent variables have been proved to be significant when being integrated into DCMs [1].

One alternative to incorporate latent variables into discrete choice models is the integrated choice and latent variable (ICLV) model [2, 3], which can be viewed as a variation of the traditional structural equation model (SEM) with the advantage to involve the underlying choice process based on psychological concepts. The ICLV model is a popular research topic in the field of transport choice studies since the early 2000s, and the application of this model has been increasing in recent years [3–7]. The earlier ICLV models are generally developed on the basis of the mixed logit model, of which the error terms follow an independent and identically Gumbel distribution. These logit kernel-based ICLV models ignore the correlation among the utility of alternatives and the correlation among the error terms of latent variables. Meanwhile, the models have several deficiencies in the aspects of the estimation method (i.e., maximum simulated likelihood (MSL) approach [8, 9]), the convergence of model estimation with multiple latent variables [7], and computational time of multidimensional integration. Regarding the limitations of the above models, Bhat and Dubey [10] firstly proposed an multinomial probit (MNP) kernel-based ICLV formulation method, which can accommodate a large number of latent variables, and

performances in convergence and computational efficiency have been improved. In particular, the maximum approximate composite marginal likelihood (MACML) inference approach is integrated to estimate the model parameters, which further simplifies the estimation process. As the number of latent variables is independent of the integral dimensions of the log-likelihood function and only relates to the number of alternatives of choice models, the computational time of this method is relatively short. Besides, the ordered and continuous response indicators for the latent variables can be involved in the model easily. Up to now, the new MNP-based ICLV model has not been extensively applied [11, 12] and remains a broad prospect for practice.

At present, it is critical to compute the multivariate cumulative normal distribution (MVNCD) function of the proposed ICLV model, which can be further improved. In some earlier studies, the simulation techniques are used for evaluating the MVNCD functions, of which the most common method is using a Geweke–Hajivassiliou–Keane (GHK) simulator [13]. However, since there is a large time cost of the simulation method with the same level of accuracy, this method may be replaced by the analytic approximation method. Mendell and Elston (ME) proposed a univariate conditioning approach earlier [14], and then, the researchers made a series of improvements. In the above probit-based ICLV model, the MACML approach of Bhat [15] is adopted, which combines the Switzer–Solow–Joe (SSJ) approximation for the MVNCD function with the CML inference approach for MNP models. Here, we intend to apply a new LDLT matrix-based analytic approximation method proposed by Bhat in 2018 [16] to evaluate the MVNCD function, which is demonstrated to have better performance than the SSJ. The method can effectively reduce difficulties in convergence and covariance matrix computation that occur routinely in the maximum-likelihood estimation of choice models with analytically intractable likelihood functions. Meanwhile, applying the LDLT decomposition method to the algorithms also assists in ensuring no substantial speed reduction in each iteration.

In the process of the maximum-likelihood estimation (MLE) for models, programming analytical gradients are generally needed to improve the speed and accuracy at convergence. However, since the derivation and coding of analytical gradients are usually complicated and time-consuming for complex models, the introduction of automatic differentiation (AD) into the model estimation process enables the acquisition of gradients without coding effort, which significantly reduces the workload of researchers. The AD method takes advantage of the fact that every computer program will execute a sequence of elementary arithmetic operations and basic functions, no matter how complicated it is. The MLE procedure usually adopts a reverse mode of AD, which is first published in 1976 by Linnainmaa [17]. Compared with numerical differentiation, it has obvious advantages in terms of calculating higher-order derivatives and partial derivatives with multiple inputs. In the past decades, the theory of AD has been considerably developed. Nevertheless, it has limited applications in the field of econometrics and statistics up to now [18, 19], which

provides opportunities for further investigation in this respect.

Furthermore, the graph computation technique based on tensors represented as multidimensional arrays in the TensorFlow platform can avoid loop operations based on the two-dimensional matrix data structure used in conventional computational platforms. TensorFlow should have an advantage in computational speed over conventional platforms when the LDLT-based analytic approximation method is applied within a sample. On the one hand, the graph computation technique based on tensors can ensure that all the calculation process in LDLT-based analytic approximation is completely realized at a speed of C language codes, rather than that of scripting language codes to operate matrices repeatedly in multiple loops. On the other hand, in MLE procedures using matrix-based analytic approximations, a tensor with more dimensions is desirable because each individual observation in the sample needs a matrix-based operation for its log-likelihood computation and such computation needs to be repeated as many times as the sample size. Thus, a computational platform with a tensor data structure can avoid repeated matrix operations in a large loop and further contribute to computational speed improvement. In this study, some efforts will be made to explore the feasibility of MNP-based ICLV model estimation using LDLT-based analytic approximation and tensor data structure on the TensorFlow platform.

The rest of this study will be organized as follows. The second section elaborates on the MNP-based ICLV model and LDLT-based analytic approximation method and then introduces the AD mechanism. In the third section, the simulation experiments are designed for validating the three-alternative and four-alternative ICLV models. The next section presents the evaluation results of the MVNCD function and the estimation results for the ICLV model with different analytic approximation methods (ME, BME, and TVBS). Conclusions and discussions are summarized in the last section.

2. Methodology

2.1. MNP-Based ICLV Model. The MNP-based integrated choice and latent variable (ICLV) model proposed by Bhat and Dubey in 2014 [10] contains the following three components. Assume there are L latent variables $\mathbf{z}^*$ (index $l = 1, 2, \ldots, L$), G ordinal indicator variables $\mathbf{y}$ (index $g = 1, 2, \ldots G$), I alternatives' utilities $\mathbf{U}$ (index $i = 1, 2, \ldots I$), and Q individuals (index $q = 1, 2, \ldots Q$) in this model. To simplify the presentation, the matrix form is taken in the formulas below and the index of the variables is not shown.

2.1.1. Latent Variable Structural Equation Model.

$$\mathbf{z}^* = \boldsymbol{\alpha}\mathbf{w} + \boldsymbol{\eta}, \boldsymbol{\eta} \sim N[0_{\mathbf{L}}, \Gamma], \tag{1}$$

where $\mathbf{w}$ is a $(\widetilde{D} \times 1)$ vector of observed covariates affecting $\mathbf{z}^*$ (not including a constant) and $\boldsymbol{\alpha}$ is a corresponding $(L \times \widetilde{D})$ matrix of coefficients (i.e., exogenous variable loadings on $\mathbf{z}^*$). $\boldsymbol{\eta}$ is a $(L \times 1)$ vector of errors assumed to be standard

multivariate normally distributed: $\eta \sim N[0_L, \Gamma]$, where Γ is a correlation matrix. Each η_l of η is assumed to be the standard normal distribution. For the structural equation model, the same exogenous variables $\mathbf{w}$ can be used for different latent variables z_l^*.

2.1.2. Latent Variable Measurement Equation Model

$$\mathbf{y} = \boldsymbol{\delta} + \mathbf{d}\mathbf{z}^* + \boldsymbol{\xi}, \psi_{\text{low}} < \mathbf{y} < \psi_{\text{up}}, \boldsymbol{\xi} \sim N[0_G, \boldsymbol{\Sigma}], \qquad (2)$$

where $\mathbf{y}$, $\boldsymbol{\delta}$, and $\boldsymbol{\xi}$ are $(G \times 1)$ vectors that represent observed ordered indicator-dependent variables, intercept in measurement equation, and error items in measurement equation, respectively. $\mathbf{d}$ is a $(G \times L)$ matrix of coefficients representing the effect of latent variables $\mathbf{z}^*$ on observed indicators $\mathbf{y}$. Also, let $\boldsymbol{\Sigma}$ be the correlation matrix of $\xi = (\xi_1, \xi_2, \ldots, \xi_G)$, in which the identification of diagonal is not considered. Here, the normalization on the error terms $\boldsymbol{\xi}$ is needed for identification, as in the usual ordered-response model. The ordered indicator $\mathbf{y}$ can be horizontally partitioned into corresponding ordinal categories by setting thresholds. As per Bhat in 2014, ψ_{low} is a $(G \times 1)$ vector that stacks the lower thresholds $\psi_{g,n^{g-1}}$ ($g = 1, 2, \ldots, G$) and ψ_{up} is another vector that stacks the upper thresholds ψ_{g,n_g} ($g = 1, 2, \ldots, G$).

2.1.3. Choice Model (Multinomial Probit Model)

$$\mathbf{U} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\gamma}(\boldsymbol{\varphi}\mathbf{z}^*) + \boldsymbol{\varepsilon} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\mathbf{z}^* + \boldsymbol{\varepsilon},$$
$$\text{where } \boldsymbol{\lambda} = \boldsymbol{\gamma}\boldsymbol{\varphi}, \boldsymbol{\varepsilon} \sim N[0_I, \boldsymbol{\Lambda}], \qquad (3)$$

where $\mathbf{U}$ is $(I \times 1)$ vector of alternative utilities, $\mathbf{x}$ is $(I \times D)$ matrix of exogenous variables in the choice model, and $\boldsymbol{\varepsilon}$ is $(I \times 1)$ utility error vector that followed multivariate normal distribution whose variance-covariance matrix is $\boldsymbol{\Lambda}$. In the equation, $\boldsymbol{\varphi}$ is a $\sum_{i=1}^{I} N_i \times L$ matrix of variables interacting with latent variables and $\boldsymbol{\gamma}$ is an $I \times \sum_{i=1}^{I} N_i$ matrix of coefficients capturing the effects of latent variables and their interactions with exogenous variables. In a case without interactions, $\boldsymbol{\lambda}(I \times L)$ matrix of coefficients can be specified when $\boldsymbol{\lambda} = \boldsymbol{\gamma}\boldsymbol{\varphi}$ and the matrix $\boldsymbol{\varphi}$ is reduced to an identity matrix of size L. Since only the covariance matrix of error differences is estimable, $I(I-1)/2 - 1$ elements can be identified in $\boldsymbol{\Lambda}$ after normalizing the covariance matrix.

The MNP-based ICLV model framework can be presented in Figure 1.

In the process of the model estimation, $\mathbf{z}^*$ in equations (2) and (3) can be substituted into the right side of equation (1) to further simplify expressions. We assume that the error vectors $\boldsymbol{\xi}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\eta}$ are independent of each other. The following system can be obtained:

$$\begin{aligned}
\mathbf{y} &= \boldsymbol{\delta} + \mathbf{d}(\boldsymbol{\alpha}\mathbf{w} + \boldsymbol{\eta}) + \boldsymbol{\xi} \\
&= \boldsymbol{\delta} + \mathbf{d}\boldsymbol{\alpha}\,\boldsymbol{\omega} + \mathbf{d}\boldsymbol{\eta} + \boldsymbol{\xi}, \\
\mathbf{U} &= \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\mathbf{z}^* + \boldsymbol{\varepsilon} \qquad\qquad (4) \\
&= \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\lambda}(\boldsymbol{\alpha}\mathbf{w} + \boldsymbol{\eta}) + \boldsymbol{\varepsilon} \\
&= \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\boldsymbol{\alpha}\mathbf{w} + \boldsymbol{\lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}.
\end{aligned}$$

Consider the $(G + I)$ vector $\mathbf{YU}_{G+I} = [\mathbf{y}_G', \mathbf{U}_I']'$. The following is defined:

$$\mathbf{B} = \left[\frac{\boldsymbol{\delta} + \mathbf{d}\boldsymbol{\alpha}\mathbf{w}}{\mathbf{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\boldsymbol{\alpha}\mathbf{w}}\right],$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{d}\Gamma\mathbf{d}' + \boldsymbol{\Sigma} & \mathbf{d}\Gamma\boldsymbol{\lambda}' \\ \boldsymbol{\lambda}\Gamma\mathbf{d}' & \boldsymbol{\lambda}\Gamma\boldsymbol{\lambda}' + \boldsymbol{\Lambda} \end{bmatrix}. \qquad (5)$$

Then, $\mathbf{YU} \sim \mathbf{MVN}_{G+I}(\mathbf{B}, \boldsymbol{\Omega})$.

To estimate the model, we need to define matrix $\mathbf{M}$ of size $(G + I - 1) \times (G + I)$ firstly to transform the choice model into the utility difference form. When the alternative i is chosen, the matrix $(G \times G)$ in $\mathbf{M_i}$ with the first G rows and first G columns is an identity matrix, and the rest matrix $(I - 1) \times I$ in $\mathbf{M_i}$ is an identity matrix of $(I - 1)$ size with an extra column of -1 added as the ith column. For example, when the alternative $i = 2$ is chosen, $\mathbf{M_i} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$ if $G = 2$ and $I = 3$. Next, let $\mathbf{Yu} \sim \mathbf{MVN}_{G+I-1}(\widetilde{\mathbf{B}}, \widetilde{\boldsymbol{\Omega}})$, where $\widetilde{\mathbf{B}} = \mathbf{MB}$, $\widetilde{\boldsymbol{\Omega}} = \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'$. After that, the matrix dimension is reduced to $(G + I - 1)$.

To integrate the multivariate OP model and MNP model to estimate the parameters, the upper and lower threshold vectors are defined: $\widetilde{\psi}_{\text{low}} = [\psi_{\text{low}}', (-\infty_{I-1})']$ and $\widetilde{\psi}_{\text{up}} = [\psi_{\text{up}}', (0_{I-1})']$, where both $-\infty_{I-1}$ and 0_{I-1} are $(I - 1) \times 1$-column vectors that represent negative infinities and zeros, respectively. Thus, the likelihood function of the ICLV model may be written as follows:

$$L(\boldsymbol{\theta}) = \Pr\left[\widetilde{\psi}_{\text{low}} \le \widetilde{u} \le \widetilde{\psi}_{\text{up}}\right] = \int_{D_{\widetilde{u}}} f_{G+I-1}(\widetilde{\mathbf{u}}|\widetilde{\mathbf{B}}, \widetilde{\boldsymbol{\Omega}})d u, \qquad (6)$$

where $D_{\widetilde{u}} = \left\{\widetilde{\mathbf{u}}: \widetilde{\psi}_{\text{low}} \le \widetilde{\mathbf{u}} \le \widetilde{\psi}_{\text{up}}\right\}$ is the integration domain and $f_{G+I-1}(.)$ is the multivariate normal density function of dimension $(G + I - 1)$. The integral dimension of the above likelihood function does not increase with the increase in latent variables $\mathbf{z}^*$.

To further proceed, the composite marginal likelihood (CML) approach can be used to simplify the model estimation process when the full likelihood function may be
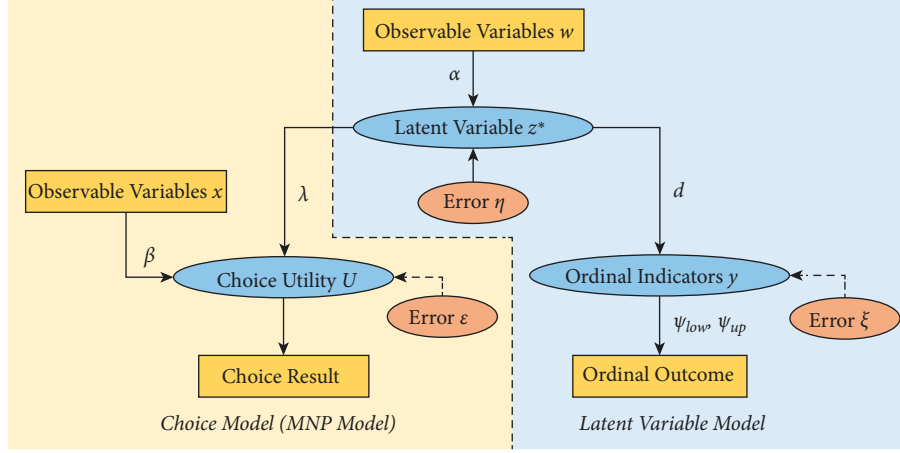
FIGURE 1: MNP-based ICLV model framework.

infeasible to evaluate. The CML estimator is the one that maximizes the compounded probability of all pairwise events, which obtains much easier-to-compute, lower-dimensional marginal likelihood. In the above case, the pairwise marginal likelihood function may be written for the ICLV model as follows:

$$L_{\text{CML}}(\boldsymbol{\theta}) = \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr\left(j_g = n_g, j_{g'} = n_{g'} \right) \right) \times \left(\prod_{g=1}^{G} \Pr\left(j_g = n_g, i = m \right) \right), \tag{7}$$

where j_g represents the index for the ordinal outcome category for the gth ordinal variables. Then, consider $\mathbf{S_g}$ (similar to $\mathbf{M}$), which is a $I \times (G + I - 1)$ selection matrix to select the gth ordered variables and all alternatives of the choice model. Let $\Phi_E(:\,;\boldsymbol{\Delta}^*)$ be the multivariate standard normal cumulative matrix function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$ ($\boldsymbol{\Delta}^* = \boldsymbol{\omega}_\Delta^{-1} \boldsymbol{\Delta} \boldsymbol{\omega}_\Delta^{-1}$, where $\boldsymbol{\omega}_\Delta$ is the diagonal matrix of the standard deviation of the covariance matrix $\boldsymbol{\Delta}$). The $L_{\text{CML}}(\boldsymbol{\theta})$ function is expanded in detail as follows:

$$L_{\text{CML}}(\boldsymbol{\theta}) = \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \left[\begin{array}{c} \Phi_2\left(\vartheta_{g,\text{up}}, \vartheta_{g',\text{up}}, v_{gg'} \right) - \Phi_2\left(\vartheta_{g,\text{up}}, \vartheta_{g',\text{low}}, v_{gg'} \right) \\ -\Phi_2\left(\vartheta_{g,\text{low}}, \vartheta_{g',\text{up}}, v_{gg'} \right) + \Phi_2\left(\vartheta_{g,\text{low}}, \vartheta_{g',\text{low}}, v_{gg'} \right) \end{array} \right] \right)$$
$$\times \left(\prod_{g=1}^{G} \Phi_I\left[\boldsymbol{\omega}_{\breve{\Omega}_{\mathbf{g}}}^{-1} \left\{ \breve{\boldsymbol{\psi}}_{\mathbf{g,up}} - \breve{\mathbf{B}}_{\mathbf{g}} \right\}; \breve{\boldsymbol{\Omega}}_g^* \right] - \Phi_I\left[\boldsymbol{\omega}_{\breve{\Omega}_{\mathbf{g}}}^{-1} \left\{ \breve{\boldsymbol{\psi}}_{\mathbf{g,low}} - \breve{\mathbf{B}}_{\mathbf{g}} \right\}; \breve{\boldsymbol{\Omega}}_g^* \right] \right), \tag{8}$$

where $\vartheta_{g,\text{up}} = (([\widetilde{\boldsymbol{\psi}}_{\text{up}}]_g - [\widetilde{\mathbf{B}}]_g)/\sqrt{[\widetilde{\boldsymbol{\Omega}}]_{gg}})$, $\vartheta_{g,\text{low}} = (([\widetilde{\boldsymbol{\psi}}_{\text{low}}]_g - [\widetilde{\mathbf{B}}]_g)/\sqrt{[\widetilde{\boldsymbol{\Omega}}]_{gg}})$, $v_{gg'} = ([\widetilde{\boldsymbol{\Omega}}]_{gg'}/\sqrt{[\widetilde{\boldsymbol{\Omega}}]_{gg} * [\widetilde{\boldsymbol{\Omega}}]_{g'g'}})$, $\breve{\boldsymbol{\psi}}_{\mathbf{g,low}} = \mathbf{S_g}([\boldsymbol{\psi}_{\text{low}}', (0_{I-1})']')$, $\breve{\boldsymbol{\psi}}_{\mathbf{g,up}} = \mathbf{S_g}\widetilde{\boldsymbol{\psi}}_{\text{up}}$, $\breve{\mathbf{B}}_{\mathbf{g}} = \mathbf{S_g}\widetilde{\mathbf{B}}$, $\breve{\boldsymbol{\Omega}}_{\mathbf{g}} = \mathbf{S_g}\widetilde{\boldsymbol{\Omega}}\mathbf{S_g}'$.

In equations (7) and (8), the CML function can be divided into two components: the first component corresponds to each pair of ordinal indicators (of which the integral dimension of MVNCD is 2), while the second component corresponds to each pair of the choice outcome and an ordinal indicator outcome (of which the integral dimension is equal to I). Accordingly, the maximum integral dimension of the MVNCD function in equation (8) does not exceed I. It is pivotal to solve the multidimensional integral in the process of model estimation. The matrix-based analytic approximation method proposed by Bhat [16] can be applied to evaluate the MVNCD function, as detailed in a later section. At last, accounting for the index q for individuals, the pairwise marginal log-likelihood function is defined as $\log L_{\text{CML}}(\boldsymbol{\theta}) = \sum_{q=1}^{Q} \log L_{CML,q}(\boldsymbol{\theta})$.

The pairwise estimator $\boldsymbol{\theta}_{\mathbf{CML}}$ obtained by maximizing the logarithm of the pairwise marginal likelihood function with respect to the vector $\boldsymbol{\theta}$ is consistent and asymptotically normally distributed with asymptotic mean $\boldsymbol{\theta}$ and co-variance matrix given by the inverse of Godambe's [20] sandwich information matrix $G(\boldsymbol{\theta})$ [21]:

$$\begin{aligned} V_{\text{CML}}(\boldsymbol{\theta}) &= [G(\boldsymbol{\theta})]^{-1} \\ &= [H(\boldsymbol{\theta})]^{-1} [J(\boldsymbol{\theta})] [H(\boldsymbol{\theta})]^{-1}, \end{aligned} \tag{9}$$

where $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ can be estimated straightforwardly at the CML estimate ($\widehat{\boldsymbol{\theta}}_{\mathbf{CML}}$):

$\widehat{H}(\widehat{\boldsymbol{\theta}}) = -[\sum_{q=1}^{Q} (\partial^2 \log L_{\mathrm{CML},q}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \ \partial\boldsymbol{\theta}')]_{\widehat{\boldsymbol{\theta}}_{\mathrm{CML}}}$ and

$$\widehat{J}(\widehat{\boldsymbol{\theta}}) = \sum_{q=1}^{Q} \left[\left(\frac{\partial \log L_{\mathrm{CML},q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right) \left(\frac{\partial \log L_{\mathrm{CML},q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \right) \right]_{\widehat{\boldsymbol{\theta}}_{\mathrm{CML}}}. \quad (10)$$

2.2. Matrix-Based Analytic Approximation Method.

The computation of the MVNCD function can be applied for the multidimensional integral of the ICLV model above. In the previous ICLV model of Bhat, the MACML approach in 2011 [15] is adopted. In this section, a new matrix-based analytic approximation method proposed by Bhat in 2018 [16] can replace the MACML approach to evaluate the MVNCD function for acquiring better estimation results [22, 23].

There are five methods for approximating the MVNCD function mentioned and summarized in the literature of Bhat in 2018. As per Bhat, the TVBS method is recommended as the evaluation approach for the MVNCD function with the highest accuracy, and the BME method has a significant speed advantage over these evaluation approaches. Besides, as the foundation of the series of analytic approximation methods, the ME method can also be included in the evaluation system in this study. Thus, these three typical methods (ME, BME, and TVBS) will be selected for elaboration below. Conceptually, the ME method is based on univariate conditioning, while BME and TVBS are based on a bivariate conditioning mechanism.

Let $\mathbf{e} = (e_1, e_2, \ldots, e_K)$ be a multivariate normally distributed random vector with zero means, variances of 1, and a correlation matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{12} & 1 & \cdots & \rho_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1K} & \rho_{2K} & \cdots & 1 \end{bmatrix}$ $(K > 2)$, where K is the dimensionality of the MVNCD function. The K-dimensional MVNCD function can be expressed as follows: $\Phi_K(e_1, e_2, \ldots, e_K: \boldsymbol{\Sigma}) = P(\mathbf{e} < \mathbf{V}) = P(e_1 < V_1, e_2 < V_2, \ldots, e_K < V_K)$. To facilitate the understanding, assume that $\mathbf{t}$ represents $\mathbf{e} < \mathbf{V}$ and t_n represents $e_1 < V_1$, and then, $\Phi_K(\mathbf{e}: \boldsymbol{\Sigma}) = P(\mathbf{e} < \mathbf{V}) = P(\mathbf{t}) = P(t_1, t_2, \ldots, t_K)$.

The joint probability of the MVNCD function of the three methods (ME, BME, and TVBS) may be expressed as follows. All the joint probability may be written as the product of a marginal probability and conditional probabilities, and the normal distribution can be used to approximate the skew normal distribution for conditional probabilities. In the equations below, the superscript n of $t^{(n)}$ is the number of truncations; $n_{\max}$ is the maximum number of truncations, and a lower $n_{\max}$ means a higher accuracy of the analytic approximation method. Let N be an intermediate parameter used to measure K: K is equal to N for the ME method, and K is discussed as being odd ($K = 2N - 1$) or even ($K = 2N$) for the BME (or TVBS) method. The expected values μ_k and variances σ_k^2 of single-sided truncations in the following equations can be obtained by the properties of truncated multivariate normal distribution, which is not detailed here (refer to Section 2.1 of Bhat in 2018).

2.2.1. The ME Method ($N \geq 2$).

$$
\begin{aligned}
P(t) &= P(t_1, t_2, \ldots, t_N), \\
&= P(t_1) \times \prod_{n=1}^{N-1} P(t_{n+1} | t_1 t_2, \ldots, t_n), \\
&\approx P(t_1) \times \prod_{n=1}^{N-1} P(t_{n+1}^{(n)}), \\
&\approx \Phi(V_1) \times \prod_{n=1}^{N-1} \Phi\left(\frac{V_{n+1} - \mu_{n+1}}{\sigma_{n+1}} \right).
\end{aligned}
\quad (11)
$$

The method relies on a single-sided truncation of a multivariate normal distribution in which some variables are truncated, while others are not. After a single variable is truncated, the residual variables follow a skew normal distribution in the ME method. $P(t_{n+1}^{(n)})$ is defined as the univariate conditional probabilities of t_{n+1} after n times truncations, which follows a univariate normal distribution (being used for approximating the skew normal distribution). Thus, the above $P(\mathbf{t})$ can be transformed into the product of recursive univariate normal cumulative distribution function (CDF) (represented by $\Phi(.)$). In equation (11), $K = N$ and $n_{\max} = N - 1$.

For instance, when $K = 6$, the probability of multivariate normal distribution can be written as follows: $P(t_1, t_2, t_3, t_4, t_5, t_6) \approx P(t_1) \times P(t_2^{(1)}) \times P(t_3^{(2)}) \times P(t_4^{(3)}) \times P(t_5^{(4)}) \times P(t_6^{(5)}) = P(e_1 < V_1) \times P(e_2^* < V_2 | e_1 < V_1) \times P(e_3^* < V_3 | e_2^* < V_2) \times P(e_4^* < V_4 | e_3^* < V_3) \times P(e_5^* < V_5 | e_4^* < V_4) \times P(e_6^* < V_6 | e_5^* < V_5)$, where e_k^* follows the skew normal distribution after $(k-1)$ times truncations. Thereinto, t_2 was truncated once and $P(t_2^{(1)}) = P(e_2^* < V_2 | e_1 < V_1) \approx \Phi((V_2 - \mu_2)/\sigma_2)$. $z_1 = e_1 | e_1 < V_1$ is defined, and it can be deduced that the expected values $\mu_2 = \rho_{12} E(z_1)$ and variances $\sigma_2^2 = \rho_{12}^2 \mathrm{Var}(z_1) + 1 - \rho_{12}^2$, where $E(z_1) = \mu_1 + \sigma_1 \lambda_1 = \lambda_1$, $\mathrm{Var}(z_1) = \sigma_1^2(1 + \lambda_1 V_1 - \lambda_1^2) = 1 + \lambda_1 V_1 - \lambda_1^2$, and $\lambda_1 = -(\phi(V_1))/\Phi(V_1)$. Similar to the calculation principle of $\Phi((V_2 - \mu_2)/\sigma_2)$, $\Phi((V_{n+1} - \mu_{n+1})/\sigma_{n+1})$ in equation (11) can be calculated by iteratively updating the expected value and covariance matrix. In the equation above, the maximum number of truncations $n_{\max}$ is 5.

2.2.2. The Bivariate ME (BME) Method ($N \geq 2$). When $K = 2N$ is even,

$$
\begin{aligned}
P(t) &= P(t_1, t_2, \ldots, t_{2N}), \\
&= P(t_1, t_2) \times \prod_{n=1}^{N-1} P(t_{2n+1}, t_{2n+2} \mid t_1, t_2, \ldots, t_{2n}), \\
&\approx P(t_1, t_2) \times \prod_{n=1}^{N-1} P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}), \\
&\approx \Phi_2(V_1, V_2, \rho_{12}) \times \prod_{n=1}^{N-1} \Phi_2\left(\frac{V_{2n+1} - \mu_{2n+1}}{\sigma_{2n+1}}, \frac{V_{2n+2} - \mu_{2n+2}}{\sigma_{2n+2}}, \rho_{2n+1, 2n+2}^* \right).
\end{aligned}
\tag{12}
$$

When $K = 2N - 1$ is odd,

$$
\begin{aligned}
P(t) &= P(t_1, t_2, \ldots, t_{2N-1}), \\
&= P(t_1, t_2) \times \prod_{n=1}^{N-2} P(t_{2n+1}, t_{2n+2} \mid t_1, t_2, \ldots, t_{2n}) \times P(t_{2N-1} \mid t_1, t_2, \ldots, t_{2N-2}), \\
&\approx P(t_1, t_2) \times \left(\prod_{n=1}^{N-2} P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}) \right) \times P(t_{2N-1}^{(2N-2)}), \\
&\approx \Phi_2(V_1, V_2, \rho_{12}) \times \prod_{n=1}^{N-2} \Phi_2\left(\frac{V_{2n+1} - \mu_{2n+1}}{\sigma_{2n+1}}, \frac{V_{2n+2} - \mu_{2n+2}}{\sigma_{2n+2}}, \rho_{2n+1, 2n+2}^* \right) \times \Phi\left(\frac{V_{2N-1} - \mu_{2N-1}}{\sigma_{2N-1}} \right).
\end{aligned}
\tag{13}
$$

The BME method is in accordance with a bivariate truncation scheme, assuming that the marginal distribution of the third and fourth untruncated variables, given the first two are truncated, remains bivariate normal, and so on. In the equation above, similar to the ME method, $P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)})$ uses the bivariate normal distribution to approximate the bivariate skew normal distribution. The conditional probability $P(\mathbf{t})$ of this method for MVNCD function is a product of bivariate normal CDF $\Phi_2(.)$ when the dimension is even and a product of $\Phi_2(.)$ and $\Phi(.)$ when the dimension is odd. Here, $\rho_{2n+1,2n+2}^*$ is the updated correlation coefficient between error terms after bivariate truncation, and the calculation principle of the expected values μ_k and variances σ_k^2 is similar to the ME method. As the extension of the ME method, the BME method truncates two dimensions each time and $n_{\max}$ times in total (also does the TVBS method), where $n_{\max} = N - 1$ in equations (12) and (13).

Similarly, when $K = 6$, the $P(\mathbf{t})$ using the BME method can be expressed as follows: $P(t_1, t_2, t_3, t_4, t_5, t_6) \approx P(t_1, t_2) \times P(t_3^{(2)}, t_4^{(2)}) \times P(t_5^{(4)}, t_6^{(4)})$. Thereinto, t_3 and t_4 are truncated once ($n = 1$), and t_5 and t_6 are truncated twice ($n = 2$). The maximum number of truncations $n_{\max}$ is 2, which is smaller than the $n_{\max}$ of the ME method. This indicates that the accuracy of the bivariate truncation scheme is better than the univariate truncation scheme.

2.2.3. The Two-Variate Bivariate Screening (TVBS) Method.
When $K = 2N \, (N \geq 2)$ is even, &ecmath;

$$P(t) = P(t_1, t_2, \ldots, t_{2N}),$$

$$= P(t_1, t_2, t_3, t_4) \times \prod_{n=1}^{N-2} \frac{P(t_{2n+1}, t_{2n+2}, t_{2n+3}, t_{2n+4} | t_1, t_2, \ldots, t_{2n})}{P(t_{2n+1}, t_{2n+2} | t_1, t_2, \ldots, t_{2n})},$$

$$\approx P(t_1, t_2, t_3, t_4) \times \prod_{n=1}^{N-2} \frac{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}, t_{2n+3}^{(2n)}, t_{2n+4}^{(2n)})}{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)})},$$

$$= P(t_1, t_2, t_3) \times \left(\prod_{n=1}^{N-2} \frac{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}, t_{2n+3}^{(2n)})}{P(t_{2n+1}^{(2n)})} \right) \times \frac{P(t_{2N-1}^{(2N-2)}, t_{2N}^{(2N-2)})}{P(t_{2N-1}^{(2N-2)})},$$

$$\approx \Phi_3(V_1, V_2, V_3 : \mathbf{0}, \Sigma[1:3, 1:3]) \times \prod_{n=1}^{N-2} \frac{\Phi_3(V_{2n+1}, V_{2n+2}, V_{2n+3} : \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Omega}})}{\Phi((V_{2n+1} - \mu_{2n+1})/\sigma_{2n+1})}$$

$$\times \frac{\Phi_2(((V_{2N-1} - \mu_{2N-1})/\sigma_{2N-1}), ((V_{2N} - \mu_{2N})/\sigma_{2N}), \rho_{2N-1,2N}^*)}{\Phi((V_{2N-1} - \mu_{2N-1})/\sigma_{2N-1})}.$$

$$(14)$$

When $K = 2N - 1 (N \geq 3)$ is odd,

$$P(t) = P(t_1, t_2, \ldots, t_{2N-1}),$$

$$= P(t_1, t_2, t_3, t_4) \times \left(\prod_{n=1}^{N-3} \frac{P(t_{2n+1}, t_{2n+2}, t_{2n+3}, t_{2n+4} | t_1, t_2, \ldots, t_{2n})}{P(t_{2n+1}, t_{2n+2}, t_1, t_2, \ldots, t_{2n})} \right) \times \frac{P(t_{2N-3}, t_{2N-2}, t_{2N-1} | t_1, t_2, \ldots, t_{2N-4})}{P(t_{2N-3}, t_{2N-2} | t_1, t_2, \ldots, t_{2N-4})},$$

$$\approx P(t_1, t_2, t_3, t_4) \times \left(\prod_{n=1}^{N-2} \frac{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}, t_{2n+3}^{(2n)}, t_{2n+4}^{(2n)})}{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)})} \right) \times \frac{P(t_{2N-3}^{(2N-4)}, t_{2N-2}^{(2N-4)}, t_{2N-1}^{(2N-4)})}{P(t_{2N-3}^{(2N-4)}, t_{2N-2}^{(2N-4)})},$$

$$= P(t_1, t_2, t_3) \times \prod_{n=1}^{N-3} \frac{P(t_{2n+1}^{(2n)}, t_{2n+2}^{(2n)}, t_{2n+3}^{(2n)})}{P(t_{2n+1}^{(2n)})},$$

$$\approx \Phi_3(V_1, V_2, V_3 : \mathbf{0}, \Sigma[1:3, 1:3]) \times \prod_{n=1}^{N-2} \frac{\Phi_3(V_{2n+1}, V_{2n+2}, V_{2n+3} : \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Omega}})}{\Phi((V_{2n+1} - \mu_{2n+1})/\sigma_{2n+1})}.$$

$$(15)$$

Similarly, the TVBS method combines the bivariate truncation scheme, in which the bivariate distribution is assumed to be rectangle-screened normal (RSN) distribution [24]. The corresponding conditional CDF at each step can be expressed as a ratio of a four-variate normal CDF $\Phi_4(.)$ and a two-variate normal CDF $\Phi_2(.)$. Since the calculation of four-variate normal CDF is time-consuming, the TVBS method adopts an approximation method by taking the trivariate CDF of the first three variates (P_{123}). Given the first three variables, the CDF of the skew normal distribution of the fourth variable is denoted as $P_{4|123}$. Accordingly, a four-variate normal CDF is obtained: $P_{1234} = P_{123} * P_{4|123} = P_{123} * P_{34|123}/P_{3|12}$. Further, the MVNCD function is composed of the product of $\Phi_3(.)$ (denotes trivariate

normal CDF) and $\Phi_3(.)/\Phi(.)$ when K is an odd number, and the product of $\Phi_3(.)$, $\Phi_3(.)/\Phi(.)$, and $\Phi_2(.)/\Phi(.)$ when K is odd. In equations (14) and (15), $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Omega}}$ represent the expected value and covariance matrix updated after bivariate truncation, respectively. Additionally, $n_{\max} = N - 1$ (K is even) and $n_{\max} = N - 2$ (K is odd) for the TVBS method.

When $K = 6$, the $P(\mathbf{t})$ using the TVBS method can be expressed as follows: $P(t_1, t_2, t_3, t_4, t_5, t_6) \approx P(t_1, t_2, t_3) \times P(t_3^{(2)}, t_4^{(2)}, t_5^{(2)})/P(t_3^{(2)}) \times P(t_5^{(4)}, t_6^{(4)})/P(t_5^{(4)})$, where $N = 3$ and $n_{\max} = 2$. Here, $t_3, t_4,$ and t_5 conducted one truncation ($n = 1$) in $P(t_3^{(2)}, t_4^{(2)}, t_5^{(2)})/P(t_3^{(2)})$, while t_5 and t_6 conducted two truncations ($n = 2$) in $P(t_5^{(4)}, t_6^{(4)})/P(t_5^{(4)})$.

Furthermore, the matrix-based analytic approximation method applies an LDLT decomposition method for the correlation matrix followed by rank 1 (ME) or rank 2 (BME

and method) updates of the factorization. The implementation is much easier to code and more computationally efficient.

As we know from the above discussion, the multidimensional integral for the analytical approximation method of MNVCD can be completely converted into integrals with no more than three dimensions. At present, the CDF of bivariate normal (BVN) distribution and trivariate normal (TVN) distribution is generally realized by numerical methods based on the Legendre–Gauss quadrature integral approximation. This numerical integration method adopted in most existing packages is relatively old with low levels of accuracy and speed. In this study, referring to the study of Genz in 2004 [25], a new numerical integration method is used to replace the original method. For BVN probabilities, we adopted the algorithm proposed by Drezner and Wesolowsky [26] with a higher-order $O(x^4)$ Taylor expansion; for TVN probabilities, the second Plackett formula method studied by Gassmann [27] is applied (not be detailed here). Accordingly, after the above algorithms are applied along with matrix operations, the performance of both functions is improved in terms of accuracy and speed.

2.3. Automatic Differentiation (AD). Automatic differentiation (AD) is a chain rule-based technique for evaluating the derivatives with respect to the input variables of functions defined by a computer program [28]. There are two different AD strategies: the forward mode and the reverse mode. The advantage of the reverse mode over the forward mode will be highlighted when the number of parameters is large, which is more suitable for the MLE procedure.

For a simple example, consider a binary probit model: $U = a + bx + \varepsilon, y = I(U > 0)$, and $\varepsilon \sim N[0, 1]$. The logarithmic likelihood function of the model can be written as $LL(\theta) = LL(a, b) = \ln\{\Phi[(2y - 1)(a + bx)]\}$. In the reverse mode of AD, the derivative is computed with respect to each sub-expression recursively in the chain rule:

$$
\begin{aligned}
\frac{\partial LL}{\partial \theta} &= \frac{\partial LL}{\partial v_1} \frac{\partial v_1}{\partial \theta} \\
&= \left(\frac{\partial \mathbf{y}}{\partial v_2} \frac{\partial v_2}{\partial v_1}\right) \frac{\partial v_1}{\partial \theta} \\
&= \left(\left(\frac{\partial \mathbf{y}}{\partial v_3} \frac{\partial v_3}{\partial v_2}\right) \frac{\partial v_2}{\partial v_1}\right) \frac{\partial v_1}{\partial \theta} \\
&= \ldots.
\end{aligned}
\tag{16}
$$

The derivative of LL with respect to v_i is defined as $\bar{v}_i = \partial LL / \partial v_i$, and the reverse accumulation with computation graph is illustrated in Figure 2. In this figure, the partial

derivatives are estimated as follows: $\partial LL / \partial a = \bar{a} = (2y - 1) * \varphi[(2y - 1) * (a + bx)]/\Phi[(2y - 1) * (a + bx)]$ and $\partial LL / \partial b = \bar{b} = x * (2y - 1) * \varphi[(2y - 1) * (a + bx)]/\Phi[(2y - 1) * (a + bx)]$. Thereinto, $\varphi(.)$ has a closed form, and $\Phi(.)$ can be approximated by the numerical integration method.

TensorFlow (https://www.tensorflow.org/), an open-source framework for large-scale distributed numerical computation, is widely applied in the areas of machine learning and deep neural network during early explorations. The main feature of the TensorFlow framework is based on a data flow graph and uses tensor as the core component to denote data. The framework builds a gradient graph corresponding to the original computation graph by backward propagation and chain rule of derivative values, which is more mature and flexible than other programming modes in the application of automatic gradient. Due to the advantages of TensorFlow, we applied it for developing econometric models in this study.

AD has obvious advantages when it is applied to the MLE of parameters in econometric models. Using this technique, modelers do not need to derive and code the complex analytic gradient expression of the log-likelihood function, which greatly reduces the workload. When the CML approach is used to estimate model parameters, the Hessian matrix and Jacobian matrix of estimated parameters need to be calculated. The errors and complexity of the numerical differentiation method increase when second-order derivatives are computed, and the speed is low in calculating partial derivatives of a function with respect to many inputs. Therefore, the use of automatic differentiation instead of numerical differentiation can aid in attaining more accurate estimation results with less computational time.

3. Simulation Experiments

To evaluate the ability of the proposed ICLV model to recover parameters from finite samples, it is necessary to conduct Monte Carlo simulation experiments. In the study of Bhat [10], four ordered indicator variables and three alternatives are concerned. In this section, referring to Bhat's experiment and further increasing the number of alternatives, we applied the LDLT matrix-based analytic approximation method for a higher integral dimension in simulation experiments.

Here, we design simulation experiments and discuss the identifiability of the model system based on two assumptions (excluding those mentioned in the last section): (1) the number of indicators L will not be less than the number of latent variables G (i.e., each latent variable at least has one indicator, which only loads on that variable but does not load on any other latent variable); (2) there is no common variable between each row of $\mathbf{x}$ matrix and the $\mathbf{w}$ vector. To explain the model more clearly and concisely, the following ICLV model is designed as an example. Consider that the
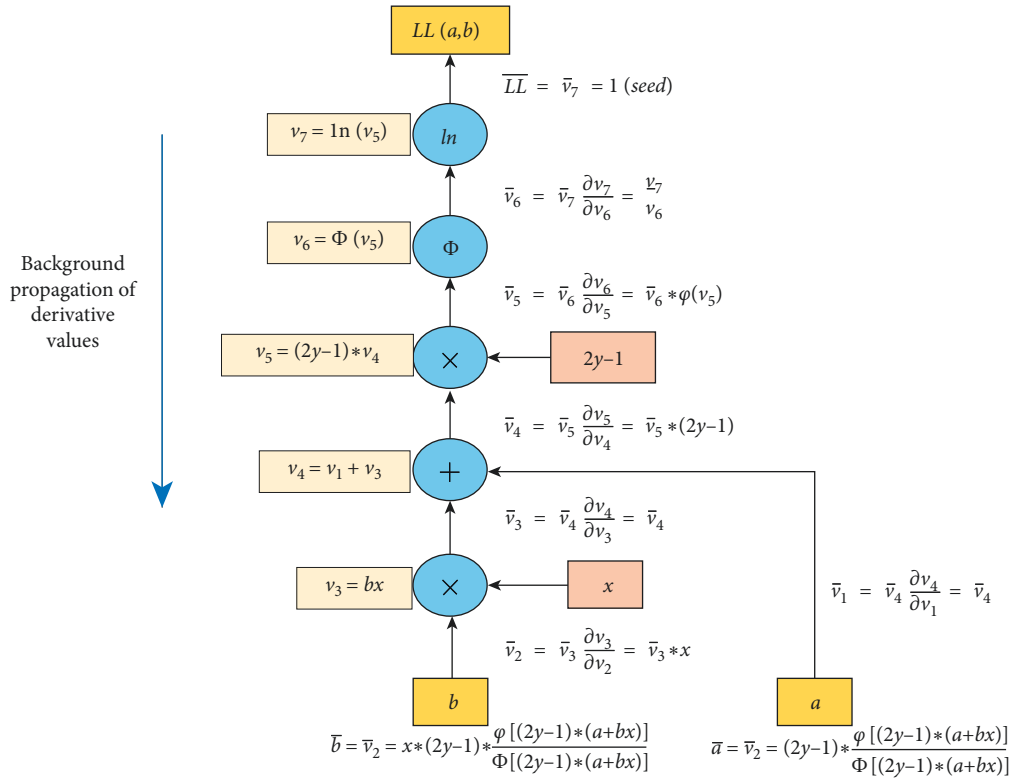
FIGURE 2: Example of reverse-mode AD in the computation graph.

ICLV model is composed of four latent variables z_i^*, four ordered indicator variables y_g, and three alternatives' utilities U_i, where the structural equation model of the latent variables in equation (1) can be represented as follows:

$$\begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0.0 & \alpha_2 & 0.0 & 0.0 \\ 0.0 & \alpha_3 & 0.0 & \alpha_4 & 0.0 \\ \alpha_5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & \alpha_6 & 0.0 & 0.0 & \alpha_7 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}. \quad (17)$$

Assume that each ordered variable has only one threshold ψ_g (i.e., $(\psi_g = \psi_{g,2})$), and the four-variate ordered probit model for the latent variable measurement model of equation (2) is as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} + \begin{bmatrix} d_1 & 0.0 & 0.0 & 0.0 \\ 0.0 & d_2 & 0.0 & 0.0 \\ 0.0 & 0.0 & d_3 & 0.0 \\ 0.0 & 0.0 & 0.0 & d_4 \end{bmatrix} \begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} \quad (18)$$

Then, the choice model with three alternatives of equation (3) can be specified as follows:

$$\begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & tt_1 & tc_1 \\ 1 & 0 & tt_2 & tc_2 \\ 0 & 1 & tt_3 & tc_3 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_{tt} \\ \beta_{tc} \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 \\ \lambda_1 & 0.0 & \lambda_3 & 0.0 \\ 0.0 & \lambda_2 & 0.0 & \lambda_4 \end{bmatrix} \begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}, \quad (19)$$

where tt_i and tc_i are the exogenous explanatory variables in the choice model.

In the above model, the correlation matrix Σ of the multivariate ordered probit model is fixed as an identity matrix, and the correlation matrix Γ of latent variables and the covariance matrix Λ of the choice model can be defined as follows:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{bmatrix}, \qquad (20)$$

$$\Lambda = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}.$$

There are $(L * (L - 1)/2)$ parameters in Γ and $(I * (I - 1)/2 - 1)$ parameters in Λ that need to be estimated. The covariance matrix Ω of the whole model after considering z^* can be expressed as follows:

$$\Omega = \begin{bmatrix} d\Gamma d' + \Sigma & d\Gamma\lambda' \\ \lambda\Gamma d' & \lambda\Gamma\lambda' + \Lambda \end{bmatrix}$$

$$= \begin{bmatrix} d_1^2 + 1 & d_1 d_2 r_{12} & d_1 d_3 r_{13} & 0 & d_1\lambda_1 + d_1\lambda_3 r_{13} & d_1\lambda_2 r_{12} + d_1\lambda_4 r_{14} \\ \lambda_1^2 + \lambda_3^2 + 2\lambda_1\lambda_3 r_{13} + \sigma_{22} & d_2 d_4 r_{24} & d_3 d_4 r_{14} & 0 & d_2\lambda_1 r_{12} + d_2\lambda_3 r_{23} & d_2\lambda_2 + d_2\lambda_4 r_{24} \\ \lambda_1\lambda_2 r_{12} + \lambda_1\lambda_4 r_{14} + \lambda_2\lambda_3 r_{23} + \lambda_3\lambda_4 r_{34} + \sigma_{23} & d_3^2 + 1 & d_3 d_4 r_{24} & 0 & d_3\lambda_1 r_{13} + d_3\lambda_3 & d_3\lambda_2 r_{23} + d_3\lambda_4 r_{34} \\ \lambda_2^2 + \lambda_4^2 + 2\lambda_2\lambda_4 r_{24} + \sigma_{33} & d_2 d_3 r_{23} & d_3 d_4 r_{34} & 0 & d_4\lambda_1 r_{14} + d_4\lambda_3 r_{34} & d_4\lambda_2 r_{24} + d_4\lambda_4 \\ d_2^2 + 1 & \cdot & d_4^2 + 1 & \sigma_{11} & \sigma_{12} & \sigma_{13} \end{bmatrix}. \qquad (21)$$

Applying the CML approach of in equations (7) and (8), the number of identifiable parameters in the Ω is 16 $(6 + 8 + 2)$ in the above sample. Thereinto, $G * (G - 1)/2$ parameters of the matrix $d\Gamma\, d' + \Sigma$ can be identified (the covariance matrix needs to normalize the diagonal elements to 1 in advance), $G * (I - 1)$ parameters of $d\Gamma\lambda'$ can be identified where $(I - 1)$ parameters can be identified in each row at most, and $(I * (I - 1)/2 - 1)$ parameters can be identified in the matrix $\lambda\Gamma\lambda' + \Lambda$ as mentioned in the section of methodology.

As shown in Figure 3, we conduct the simulation experiments of two models: one contains three alternatives of the choice model and the other contains four alternatives, with the same settings in the latent variable structural equation model and measurement equation model in which r_{12}, r_{14}, r_{23}, and r_{34} are fixed at the value of zero. For model 1 of three alternatives, the simulation experiment is undertaken 50 times for a sample size of 10,000 with different realizations of **YU** vector to generate 50 different datasets. The estimator is then applied to each dataset to estimate data specific values for the 31 parameters, and the numerical integration method for calculating the CDF of TVN distribution is used for the MVNCD function; for model 2 of

four alternatives, we perform the above data generation progress once with a large sample size of 100,000 and then carry out the Monte Carlo simulation experiments on the same dataset using ME, BME, and TVBS methods, respectively. The analytic approximation method of four-variate normal CDF $P(t_1, t_2, t_3, t_4)$ can be written as follows: $P(t_1) \times P(t_2^{(1)}) \times P(t_3^{(2)}) \times P(t_4^{(3)})$ for the ME method, $P(t_1, t_2) \times P(t_3^{(2)}, t_4^{(2)})$ for the BME method, and $P(t_1, t_2, t_3) \times P(t_3^{(2)}, t_4^{(2)})/P(t_3^{(2)})$ for the TVBS method. The 37 parameters can be estimated in the ICLV model.

4. Model Estimation Results

4.1. The MVNCD Function Evaluation Results. To evaluate the accuracy of computing the MVNCD function for different dimensions, we generate 1000 groups of random numbers, which consist of four sets of 250 MVNCD evaluations each, corresponding to (a) low correlations, high MVNCD values, (b) low correlations, low MVNCD values, (c) high correlations, high MVNCD values, and (d) high correlations, low MVNCD values. On the one hand, to generate the datasets of low and high MVNCD values, the upper integration limits can be drawn from $[0, +\sqrt{K}]$ for the

(1) Latent variable structural equation model

$z_1^* + 0.5 * w_1 + 0.6 * w_3 + \eta_1$
$z_2^* + 0.5 * w_2 + 0.6 * w_4 + \eta_2$
$z_3^* + 0.3 * w_1 \qquad\qquad + \eta_3$
$z_4^* + 0.3 * w_2 - 0.4 * w_5 + \eta_4$

$$\Gamma = \begin{bmatrix} 1.0 & 0.0 & 0.6 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.6 \\ 0.6 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.6 & 0.0 & 1.0 \end{bmatrix}$$

(2) Latent variable measurement equation model

$y_1 = -1.0 + 0.6\, z_1^* + \xi_1,\ \psi_1 = 1.5$
$y_2 = -1.0 + 0.6\, z_2^* + \xi_2,\ \psi_2 = 1.5$
$y_3 = -1.0 + 0.6\, z_3^* + \xi_3,\ \psi_3 = 1.5$
$y_4 = -1.0 + 0.6\, z_4^* + \xi_4,\ \psi_4 = 1.5$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(3) Choice mode

Model 1: 3 alternatives:

$U_1 = \qquad\qquad -1.0 * tt_1 - 0.8 * tc_1 \qquad\qquad\qquad\qquad + \varepsilon_1$
$U_2 = -0.5 - 1.0 * tt_2 - 0.8 * tc_2 + 0.5 * z_1^* + 0.5 * z_3^* + \varepsilon_2$
$U_3 = -1.0 - 1.0 * tt_3 - 0.8 * tc_3 + 0.2 * z_2^* + 0.2 * z_4^* + \varepsilon_3$

$$\wedge = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.6 \\ 0.0 & 0.6 & 1.36 \end{bmatrix}$$

Model 2: 4 alternatives:

$U_1 = \qquad\qquad -1.0 * tt_1 - 0.8 * tc_1 \qquad\qquad\qquad\qquad + \varepsilon_1$
$U_2 = -0.5 - 1.0 * tt_2 - 0.8 * tc_2 + 0.5 * z_1^* + 0.5 * z_3^* + \varepsilon_2$
$U_3 = -1.0 - 1.0 * tt_3 - 0.8 * tc_3 + 0.2 * z_2^* + 0.2 * z_4^* + \varepsilon_3$
$U_4 = -0.8 - 1.0 * tt_4 - 0.8 * tc_4 + 0.2 * z_2^* + 0.5 * z_3^* + \varepsilon_4$

$$\wedge = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.6 & 0.6 \\ 0.0 & 0.6 & 1.36 & 0.6 \\ 0.0 & 0.6 & 0.6 & 1.36 \end{bmatrix}$$
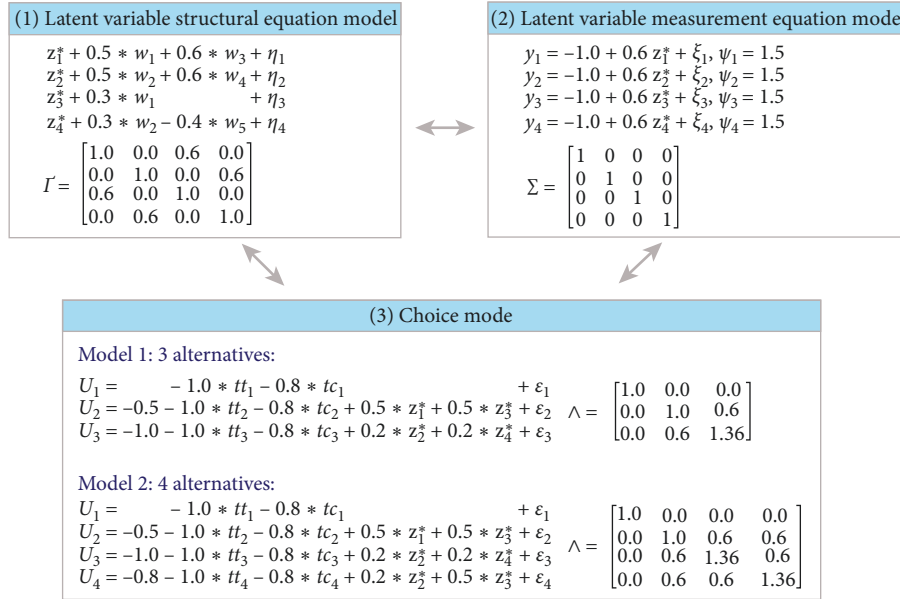
FIGURE 3: Parameter settings of the ICLV model (for three and four alternatives).

cases of high MVNCD values, while it can be drawn from $[-\sqrt{K}/2, +\sqrt{K}\,]$ for the cases of low MVNCD values. On the other hand, to generate the datasets of low correlations and high correlations, the positive-definite covariance matrices **C** can be calculated by $\mathbf{C} = \mathbf{RR}' + \delta \times \text{diag}(\mathbf{ru})$, where **R** represents a matrix of $K \times K$ random univariate standard normal variates and $\text{diag}(\mathbf{ru})$ is a diagonal matrix with the $K \times 1$ vector **ru** of standard uniform random variates on the diagonal. δ is a scalar to determine the relative magnitude of the diagonal elements relative to the non-diagonal elements. We can use the two different values of δ to distinguish the matrix with low correlations ($\delta = 10$) and high correlations ($\delta = 0$).

The MVNCD evaluation results from the different methods are presented in Tables 1 and 2. Here, we use the Gaussian quadrature integral based on 400 nodes as the exact value when $K \le 3$ and use the value calculated by the CDFMVNe function of the Gauss (up to an accuracy of $1e - 6$) as the exact value when $K > 3$. Meanwhile, the mean absolute error (MAE) and the mean absolute percentage error (MAPE) are chosen to evaluate the accuracy level of the MVNCD evaluation results. The "%MAE > 0.005" represents the percentage in which the MAE was over 0.005, and "%MAPE > 2" represents the percentage in which the MAPE was over 2.

From Table 1, the following findings can be obtained. Compared with the calculated results by the cdfBvn and cdfTvn functions in the GAUSS (https://www.aptech.com/wp-content/uploads/2016/05/GAUSS_16_PDF_UG.pdf), the accuracy levels (shown as the MAE) of BVN and TVN probability (based on the improvement in Section 2.3) calculated with the assistance of TensorFlow have been improved from e^{-10} to e^{-15} and from e^{-10} to e^{-11}, respectively, with a slight loss in time. When the integration dimension K is 2 or 3, as shown in Table 1, the accuracy level of the MAE for the ME and BME methods is e^{-3}, which is far lower than that of bivariate and trivariate normal distribution function calculated by numerical integration. Similar

conclusion can be obtained by choosing the MAPE as an index. Therefore, numerical integration should be adopted to calculate BVN or TVN values.

When $K \ge 4$, we employed the LDLT matrix-based analytic approximation method for the evaluation on Gauss and TensorFlow platforms. From Table 2, when the number of dimensions "K" increases, the performance of the three methods can be indicated by "MAE," "MAPE," "% MAE > 0.005," and "%MAPE > 2." The TVBS method is taken as an example, and there is a variation as the dimension "K" goes from 4 to 15, with "MAE" reducing from 0.00117 to 0.00037, "%MAE > 0.005" reducing from 5.8% to 0.1%, and "%MAPE > 2" increasing from 8.4% to 32.0%. As shown in Figure 4(a), "%MAE > 0.005" gradually decreases and "%MAPE > 2" gradually increases with the increase in dimension of all three methods. To explain the intrinsic reason, we checked the data and found that the overall average value of 1000 MVNCD values calculated by the method of generating random number would decrease with the increase in dimension. Due to the lower MVNCD values in the higher dimension, the relative error (MAPE) may become larger with the higher dimension. Meanwhile, as expected, with the increase in K, the time required by the analytic approximation gradually increased for all the methods. In addition, when we compare the three analytical approximation methods with each other at the same dimension, it can be found that the time consumed by the BME method is the shortest. In general, the time consumed by BME is less than half of that by the ME method and that by the TVBS method falls in the middle. Compared with the ME method, the error of the BME and TVBS methods is relatively low, as both of them are based on the bivariate truncation scheme. Overall, the TVBS is the best in terms of accuracy in evaluating individual MVNCD functions. These findings are consistent with Bhat's study (2018) and mutually corroborate.

Table 1: BVN and TVN evaluation results for the different methods.

Dim	Measure	Numerical integration methods		Analytic approximation methods (TensorFlow)		
		Gauss	TF	ME	BME	TVBS
$K = 2$ (BVN)	MAE	$4.8^*e - 10$	$1.9^*e - 15$	0.0037	—	—
	%MAE > 0.005	0.0	0.0	18.8	—	—
	MAPE	$8.6^*e - 8$	$1.3^*e - 12$	1.28	—	—
	%MAPE > 2	0.0	0.0	15.2	—	—
	Time (s)	0.001	0.002	0.003	—	—
$K = 3$ (TVN)	MAE	$8.8^*e - 10$	$6.1^*e - 11$	0.0027	0.0018	—
	%MAE > 0.005	0.0	0.0	18.1	12.4	—
	MAPE	$1.9^*e - 7$	$2.9^*e - 8$	1.34	1.00	—
	%MAPE > 2	0.0	0.0	16.7	11.0	—
	Time (s)	0.005	0.005	0.006	0.003	—

Note. "—" indicates that the corresponding analytic approximation methods cannot calculate the values of BVN or TVN.

Table 2: MVNCD evaluation results for the different analytic approximation methods.

Dim	Measure	ME		BME		TVBS	
		Gauss	TF	Gauss	TF	Gauss	TF
$K = 4$	MAE	0.00226	0.00226	0.00152	0.00152	0.00117	0.00117
	MAPE	0.59	0.59	0.40	0.40	0.26	0.26
	%MAE > 0.005	15.2	15.2	9.8	9.8	5.8	5.8
	%MAPE > 2	18.0	18.0	12.8	12.8	8.4	8.4
	Time (s)	0.052	0.010	0.033	0.004	0.039	0.007
$K = 5$	MAE	0.00171	0.00171	0.00142	0.00142	0.00108	0.00108
	MAPE	0.54	0.54	0.44	0.44	0.33	0.33
	%MAE > 0.005	9.2	9.2	7.0	7.0	4.8	4.8
	%MAPE > 2	18.8	18.8	15.4	15.4	13.0	13.0
	Time (s)	0.077	0.016	0.060	0.007	0.065	0.011
$K = 7$	MAE	0.00120	0.00120	0.00104	0.00104	0.00093	0.00093
	MAPE	0.45	0.45	0.38	0.38	0.33	0.33
	%MAE > 0.005	5.9	5.9	4.7	4.7	3.6	3.6
	%MAPE > 2	21.5	21.5	20.2	20.2	18.3	18.3
	Time (s)	0.141	0.031	0.097	0.014	0.105	0.021
$K = 10$	MAE	0.00064	0.00064	0.00057	0.00057	0.00054	0.00054
	MAPE	0.32	0.32	0.29	0.29	0.29	0.29
	%MAE > 0.005	1.6	1.6	1.2	1.2	1.1	1.1
	%MAPE > 2	25.0	25.0	23.1	23.1	22.6	22.6
	Time (s)	0.258	0.061	0.156	0.023	0.182	0.033
$K = 15$	MAE	0.00041	0.00041	0.00040	0.00040	0.00037	0.00037
	MAPE	0.29	0.29	0.28	0.28	0.25	0.25
	%MAE > 0.005	0.2	0.2	0.2	0.2	0.1	0.1
	%MAPE > 2	34.1	34.1	32.6	32.6	32.0	32.0
	Time (s)	0.558	0.146	0.317	0.058	0.352	0.076

Furthermore, there are some findings by comparing the calculation error and calculation time of the different analytic approximation methods (i.e., ME, BME, and TVBS) in Gauss and TensorFlow. In terms of calculation accuracy, it can be seen in Table 2 that "MAE," "MAPE," "% MAE > 0.005," and "%MAPE > 2" of the three analytic approximation methods are basically the same. Therefore, although the new method improves accuracy levels of BVN and TVN probability in TensorFlow, its contribution to accuracy improvement is limited for high-dimensional analytic approximation; in terms of calculation time, it can be seen in Table 2 and Figure 4(b) that under the same dimension and algorithm, the calculation speed of TensorFlow platform is much higher than that of Gauss platform, and the average running time of Gauss is about 5-6 times that of TensorFlow. As shown in Table 1, there is no significant difference in the calculation time of BVN and TVN values on the two platforms. Thus, the declines in calculation time are mainly due to the obvious time advantage of TensorFlow with tensor data structure and graph computation technique.

4.2. Simulation Results for ICLV Model. Following Section 3, the two ICLV models are designed to evaluate the ability for recovering the model parameters. For the three-alternative
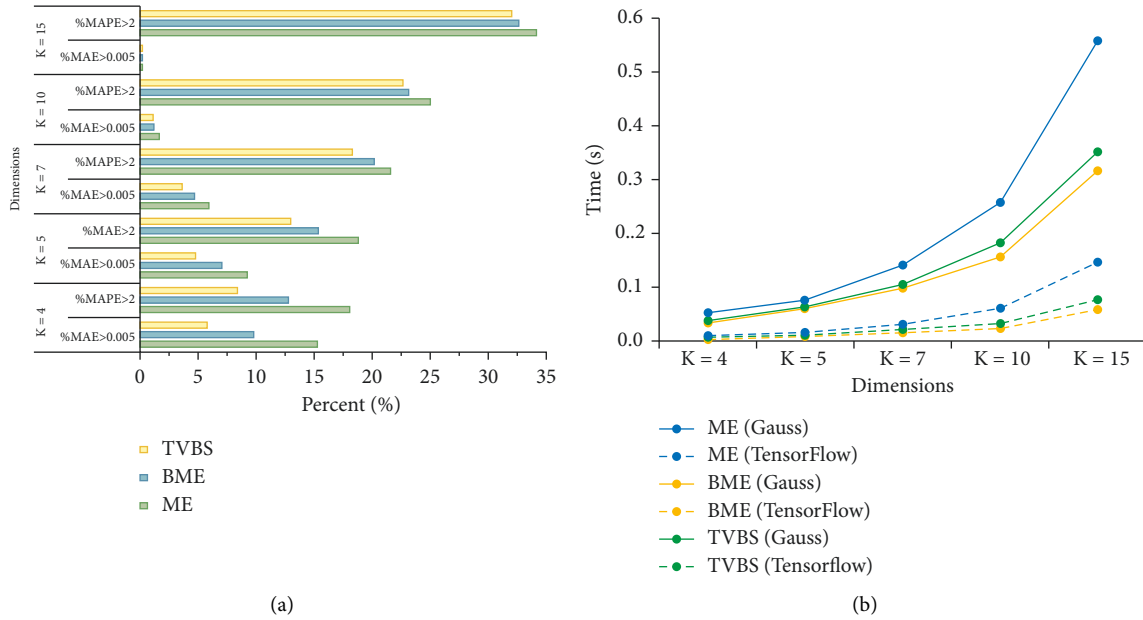
(a)

(b)

Figure 4: Comparison of accuracy and time with different dimensions. (a) "%MAE > 0.005" and "%MAPE > 2" (b) time.

ICLV model with the maximum integral dimension being 3, the performance of the improved trivariate normal integral approach in estimating the parameters and the standard errors of the ICLV model is evaluated in Table 3. The evaluation system can be adopted as follows:

(1) For the parameter estimates, the mean estimate for each model parameter across the 50 datasets can be obtained. Then, the absolute bias of the estimator is computed as Abs.Bias = |(estimate value− true value)| and the absolute percentage bias as APB = |(estimate value − true value)/ true value| × 100.

(2) For the standard error estimates, the standard deviation of each estimated parameter across the 50 datasets can be computed, and this is labeled as the finite sample standard deviation or FSSD (essentially, this is the empirical standard error); also, we can compute the mean standard deviation for each model parameter across the 50 datasets and denote this as the asymptotic standard error or ASE (the standard error here is computed using the Godambe (sandwich) matrix at convergence for a theoretical approximation to the FSSD). Next, to evaluate the accuracy of the ASE formula for the finite sample size used, the relative efficiency of the estimator is computed as relative efficiency = ASE/FSSD. As per Bhat [10], the closeness between ASE and FSEE may be examined as captured by the 0.75–1.25 range for the relative efficiency value.

The following aspects can be observed in Table 3 (the experiment is repeated 50 times with a sample size of 10,000). Firstly, the overall mean value across parameters of APB is 3.7940%, with the individual parameter APB ranging from 0.1% to 19%, which are close to those in Bhat. Besides, the simulation errors are also similar and therefore

considered reasonable and acceptable; secondly, the mean APB values for the $\mathbf{d}$ vector (9.0%), $\boldsymbol{\lambda}$ vector (4.3%), and r in $\boldsymbol{\Gamma}$ matrix (9.0%) are relatively higher than the overall mean APB. This indicates that the coefficients on the latent variable vector $\mathbf{z}^*$ in the measurement equation and the choice model are somewhat more difficult to recover than other parameters. It is expected, since these elements enter into the covariance matrix $\boldsymbol{\Omega}$ in a nonlinear fashion as shown in equation (6), and $\boldsymbol{\Omega}$ itself enters into the composite likelihood function (equation (8)) in a complex manner. Lastly, as for simulation error estimates, the overall mean value of relative efficiency (RE) across the parameters is 1.1190 (and the RE of almost all parameters is between 0.75 and 1.25), which suggests that the FSSE and ASE values are close to each other and the asymptotic formula is performing well in estimating the finite sample standard errors.

The simulation results (with a sample size of 100,000) of the four-alternative ICLV model estimated by the three analytical approximation methods, namely the ME, BME, and TVBS, are presented in Tables 4 –6, respectively. In these tables, Abs.Bias = |estimate value − true value| and APB = |(estimate value − true value)/truevalue| × 100. As shown in the experimental results, it is ensured that all the identifiable parameters can be significantly estimated and seemingly consistent with their true values. We can see that the values of absolute percentage bias (APB) fall in the ranges of 0–15% in the three tables, of which the APB of four parameters exceeds 10% (i.e., α_3, α_4, d_2, and σ_{34}). It is a similar finding showing that the parameter estimators related to $\mathbf{z}^*$ are less consistent. Additionally, in these three tables (also shown in Figure 5), it can be observed that among the three methods (i.e., ME, BME, and TVBS), the overall mean value of absolute bias is 0.0266, 0.0262, and 0.0259, and the average value of APB is 4.269%, 4.240%, and 4.204%, respectively. It is noticed that the values of these two

TABLE 3: Simulation results for the three-alternative ICLV model.

Parm.	True value	Parameter estimates			Standard error estimates		
		Mean est.	Abs. bias	APB (%)	ASE	FSSD	R. eff
α_1	0.5	0.5300	0.0300	5.9924	0.0961	0.0807	1.1908
α_2	0.6	0.6411	0.0411	6.8506	0.1177	0.1003	1.1730
α_3	0.5	0.5022	0.0022	0.4432	0.2037	0.1493	1.3648
α_4	0.6	0.6017	0.0017	0.2775	0.2451	0.1845	1.3288
α_5	0.3	0.3125	0.0125	4.1630	0.0408	0.0466	0.8740
α_6	0.3	0.3027	0.0027	0.9119	0.1336	0.1118	1.1950
α_7	−0.4	−0.4061	0.0061	1.5186	0.1773	0.1455	1.2182
s_1	1	—	—	—	—	—	—
r_{12}	0	—	—	—	—	—	—
r_{13}	0.6	0.6855	0.0855	14.2485	0.2567	0.2516	1.0206
r_{14}	0	—	—	—	—	—	—
s_2	1	—	—	—	—	—	—
r_{23}	0	—	—	—	—	—	—
r_{24}	0.6	0.6230	0.0230	3.8411	0.4800	0.3951	1.2149
s_3	1	—	—	—	—	—	—
r_{34}	0	—	—	—	—	—	—
s_4	1	—	—	—	—	—	—
δ_1	−1	−1.0013	0.0013	0.1312	0.0366	0.0334	1.0974
δ_2	−1	−1.0244	0.0244	2.4418	0.0680	0.0638	1.0664
δ_3	−1	−1.0031	0.0031	0.3117	0.0441	0.0453	0.9720
δ_4	−1	−1.0665	0.0665	6.6513	0.1629	0.1609	1.0124
d_1	0.3	0.2880	0.0120	3.9911	0.0508	0.0441	1.1524
d_2	0.4	0.4459	0.0459	11.4847	0.1550	0.1355	1.1441
d_3	0.5	0.4924	0.0076	1.5296	0.0754	0.0832	0.9064
d_4	0.6	0.7157	0.1157	19.2758	0.3046	0.2619	1.1633
ψ_1	1.5	1.4971	0.0030	0.1967	0.0275	0.0265	1.0375
ψ_2	1.5	1.5354	0.0354	2.3578	0.0873	0.0823	1.0601
ψ_3	1.5	1.4949	0.0051	0.3431	0.0504	0.0549	0.9191
ψ_4	1.5	1.6141	0.1141	7.6054	0.2346	0.2114	1.1101
β_2	0.5	0.4863	0.0137	2.7409	0.0954	0.0888	1.0741
β_3	−1	−0.9972	0.0028	0.2784	0.1360	0.1530	0.8893
β_{tt}	−1	−0.9980	0.0020	0.1971	0.0419	0.0434	0.9667
β_{tc}	−0.8	−0.7978	0.0022	0.2772	0.0338	0.0343	0.9852
λ_1	0.5	0.4807	0.0193	3.8694	0.0885	0.0811	1.0907
λ_2	0.2	0.2129	0.0129	6.4477	0.0661	0.0517	1.2771
λ_3	0.5	0.4929	0.0071	1.4146	0.0718	0.0723	0.9929
λ_4	0.2	0.2112	0.0112	5.6136	0.0622	0.0543	1.1441
σ_{11}	1	—	—	—	—	—	—
σ_{12}	0	—	—	—	—	—	—
σ_{13}	0	—	—	—	—	—	—
σ_{22}	1	—	—	—	—	—	—
σ_{23}	0.6	0.6065	0.0065	1.0794	0.1374	0.1196	1.1483
σ_{33}	1.36	1.3753	0.0153	1.1286	0.2742	0.2571	1.0668
Overall mean value			0.0236	**3.7940**	0.1308	0.1169	**1.1190**

Note: "--" indicates that the corresponding parameter is fixed at the true value.

indicators slightly decrease. On the other hand, the standard errors of all methods are about 0.037. It shows that the standard errors of the four-alternative ICLV model seem reasonable and acceptable. Overall, we have validated the ability for recovering parameters of these three LDLT-based analytic approximation methods for the ICLV model, and we can conclude that TVBS is superior to BME and BME is superior to ME. These findings are consistent with the evaluation results with respect to individual MVNCD functions. Although model parameters can be estimated significantly with a sample of large size, there are issues of empirical identification, namely whether the data can

support the model specification. If the empirical data cannot allow for significantly estimating all the theoretically identifiable parameters, some parameters may be fixed and only parameters of particular interest need to be estimated.

Furthermore, more attempts were made to discuss the operational efficiency of models using the automatic differential technique. The convergence time of the three-alternative ICLV model in Table 3 has a median value of about 8 minutes for the case of 10,000 observations, which is much shorter than that of the literature of Bhat in 2014 (about one hour for the case of 1000 observations). It can also be inferred that the computational time of the improved four-

TABLE 4: Simulation results for the four-alternative ICLV model with the ME method.

Parm.	True value	Estimate	Abs. bias	APB (%)	Std. err.	t-test	P value
α_1	0.5	0.4821	0.0179	3.5806	0.0175	27.4871	≤ 0.001
α_2	0.6	0.5734	0.0266	4.4345	0.0210	27.2569	≤ 0.001
α_3	0.5	0.5623	0.0623	12.4582	0.0545	10.3235	≤ 0.001
α_4	0.6	0.6659	0.0659	10.9858	0.0645	10.3299	≤ 0.001
α_5	0.3	0.3039	0.0039	1.2833	0.0103	29.4528	≤ 0.001
α_6	0.3	0.3059	0.0059	1.9693	0.0306	10.0031	≤ 0.001
α_7	−0.4	−0.4099	0.0099	2.4817	0.0408	−10.0422	≤ 0.001
s_1	1	—	—	—	—	—	—
r_{12}	0	—	—	—	—	—	—
r_{13}	0.6	0.5982	0.0018	0.3082	0.0460	13.0071	≤ 0.001
r_{14}	0	—	—	—	—	—	—
s_2	1	—	—	—	—	—	—
r_{23}	0	—	—	—	—	—	—
r_{24}	0.6	0.6475	0.0475	7.9242	0.1007	6.4324	≤ 0.001
s_3	1	—	—	—	—	—	—
r_{34}	0	—	—	—	—	—	—
s_4	1	—	—	—	—	—	—
δ_1	−1	−1.0068	0.0068	0.6818	0.0115	−87.6878	≤ 0.001
δ_2	−1	−0.9671	0.0329	3.2933	0.0161	−59.9248	≤ 0.001
δ_3	−1	−1.0081	0.0081	0.8070	0.0131	−76.7456	≤ 0.001
δ_4	−1	−0.9875	0.0125	1.2531	0.0373	−26.4410	≤ 0.001
d_1	0.3	0.3147	0.0147	4.9160	0.0125	25.2686	≤ 0.001
d_2	0.4	0.3508	0.0492	12.3020	0.0381	9.2070	≤ 0.001
d_3	0.5	0.4996	0.0004	0.0828	0.0210	23.7820	≤ 0.001
d_4	0.6	0.5910	0.0090	1.5023	0.0793	7.4520	≤ 0.001
ψ_1	1.5	1.5120	0.0120	0.8004	0.0081	186.2783	≤ 0.001
ψ_2	1.5	1.4792	0.0208	1.3835	0.0188	78.6552	≤ 0.001
ψ_3	1.5	1.4999	0.0001	0.0034	0.0145	103.2049	≤ 0.001
ψ_4	1.5	1.5158	0.0158	1.0553	0.0538	28.1516	≤ 0.001
β_2	0.5	0.4547	0.0453	9.0560	0.0304	14.9540	≤ 0.001
β_3	−1	−1.0416	0.0416	4.1616	0.0476	−21.8600	≤ 0.001
β_4	−0.8	−0.8728	0.0728	9.1037	0.0445	−19.5951	≤ 0.001
β_{tt}	−1	−1.0115	0.0115	1.1490	0.0158	−63.8326	≤ 0.001
β_{tc}	−0.8	−0.8107	0.0107	1.3424	0.0128	−63.5246	≤ 0.001
λ_1	0.5	0.5380	0.0380	7.6064	0.0234	22.9537	≤ 0.001
λ_2	0.2	0.1858	0.0142	7.0770	0.0186	10.0106	≤ 0.001
λ_3	0.2	0.1877	0.0123	6.1475	0.0187	10.0234	≤ 0.001
λ_4	0.5	0.4891	0.0109	2.1756	0.0232	21.0530	≤ 0.001
λ_5	0.5	0.5103	0.0103	2.0632	0.0212	24.0984	≤ 0.001
λ_6	0.2	0.1997	0.0003	0.1565	0.0204	9.7718	≤ 0.001
σ_{11}	1	—	—	—	—	—	—
σ_{12}	0	—	—	—	—	—	—
σ_{13}	0	—	—	—	—	—	—
σ_{14}	0	—	—	—	—	—	—
σ_{22}	1	—	—	—	—	—	—
σ_{23}	0.6	0.5956	0.0044	0.7365	0.0526	11.3276	≤ 0.001
σ_{24}	0.6	0.6407	0.0407	6.7772	0.0460	13.9313	≤ 0.001
σ_{33}	1.36	1.4008	0.0408	2.9999	0.1074	13.0425	≤ 0.001
σ_{34}	0.6	0.6939	0.0939	15.6422	0.0779	8.9014	≤ 0.001
σ_{44}	1.36	1.4723	0.1123	8.2542	0.1007	14.6142	≤ 0.001
Overall mean value			**0.0266**	**4.2691**	**0.0371**		

Note: "—" indicates that the corresponding parameter is fixed at the true value.

alternative ICLV model based on TensorFlow is reduced. This can be attributed to two factors: one is that the LDLT-based analytic approximation method (instead of the MACML approach) can shorten the calculation time of the ICLV model, as mentioned in Bhat. The improvement on bivariate normal CDF and trivariate normal CDF also facilitates this process; the other is based on the time advantage of TensorFlow in calculating tensor and multidimensional arrays. In addition, in the absence of coding the analytic gradient artificially, the convergence speed of the model can be improved using automatic differentiation to obtain the gradient of the likelihood function. For the four-alternative ICLV model with a sample size of 100,000, the convergence times of three analytical approximation methods are

TABLE 5: Simulation results for the four-alternative ICLV model with the BME method.

Parm.	True value	Estimate	Abs. bias	APB (%)	Std. err.	t-test	P value
α_1	0.5	0.4821	0.0179	3.5760	0.0175	27.4890	≤ 0.001
α_2	0.6	0.5734	0.0266	4.4307	0.0210	27.2577	≤ 0.001
α_3	0.5	0.5622	0.0622	12.4302	0.0540	10.4099	≤ 0.001
α_4	0.6	0.6658	0.0658	10.9583	0.0639	10.4166	≤ 0.001
α_5	0.3	0.3044	0.0044	1.4537	0.0103	29.4308	≤ 0.001
α_6	0.3	0.3060	0.0060	2.0160	0.0306	9.9954	≤ 0.001
α_7	−0.4	−0.4101	0.0101	2.5280	0.0409	−10.0344	≤ 0.001
s_1	1	—	—	—	—	—	—
r_{12}	0	—	—	—	—	—	—
r_{13}	0.6	0.5990	0.0010	0.1662	0.0461	12.9991	≤ 0.001
r_{14}	0	—	—	—	—	—	—
s_2	1	—	—	—	—	—	—
r_{23}	0	—	—	—	—	—	—
r_{24}	0.6	0.6477	0.0477	7.9500	0.1004	6.4540	≤ 0.001
s_3	1	—	—	—	—	—	—
r_{34}	0	—	—	—	—	—	—
s_4	1	—	—	—	—	—	—
δ_1	−1	−1.0068	0.0068	0.6816	0.0115	−87.6890	≤ 0.001
δ_2	−1	−0.9671	0.0329	3.2903	0.0161	−60.1612	≤ 0.001
δ_3	−1	−1.0076	0.0076	0.7647	0.0131	−76.8664	≤ 0.001
δ_4	−1	−0.9873	0.0127	1.2682	0.0373	−26.4510	≤ 0.001
d_1	0.3	0.3147	0.0147	4.9110	0.0125	25.2700	≤ 0.001
d_2	0.4	0.3509	0.0491	12.2775	0.0378	9.2838	≤ 0.001
d_3	0.5	0.4985	0.0015	0.2924	0.0210	23.7831	≤ 0.001
d_4	0.6	0.5906	0.0094	1.5625	0.0793	7.4485	≤ 0.001
ψ_1	1.5	1.5120	0.0120	0.8000	0.0081	186.2896	≤ 0.001
ψ_2	1.5	1.4793	0.0207	1.3804	0.0187	79.1925	≤ 0.001
ψ_3	1.5	1.4993	0.0007	0.0456	0.0145	103.4676	≤ 0.001
ψ_4	1.5	1.5156	0.0156	1.0393	0.0538	28.1629	≤ 0.001
β_2	0.5	0.4546	0.0454	9.0790	0.0304	14.9559	≤ 0.001
β_3	−1	−1.0417	0.0417	4.1748	0.0476	−21.8780	≤ 0.001
β_4	−0.8	−0.8719	0.0719	8.9935	0.0445	−19.5883	≤ 0.001
β_{tt}	−1	−1.0112	0.0112	1.1215	0.0158	−63.9360	≤ 0.001
β_{tc}	−0.8	−0.8106	0.0105	1.3187	0.0127	−63.6256	≤ 0.001
λ_1	0.5	0.5379	0.0379	7.5720	0.0234	22.9583	≤ 0.001
λ_2	0.2	0.1859	0.0141	7.0455	0.0184	10.0864	≤ 0.001
λ_3	0.2	0.1877	0.0123	6.1600	0.0186	10.1024	≤ 0.001
λ_4	0.5	0.4882	0.0118	2.3654	0.0232	21.0692	≤ 0.001
λ_5	0.5	0.5094	0.0094	1.8768	0.0211	24.1120	≤ 0.001
λ_6	0.2	0.1995	0.0005	0.2665	0.0204	9.7651	≤ 0.001
σ_{11}	1	—	—	—	—	—	—
σ_{12}	0	—	—	—	—	—	—
σ_{13}	0	—	—	—	—	—	—
σ_{14}	0	—	—	—	—	—	—
σ_{22}	1	—	—	—	—	—	—
σ_{23}	0.6	0.5952	0.0048	0.7927	0.0525	11.3353	≤ 0.001
σ_{24}	0.6	0.6377	0.0377	6.2887	0.0457	13.9456	≤ 0.001
σ_{33}	1.36	1.4002	0.0402	2.9534	0.1073	13.0519	≤ 0.001
σ_{34}	0.6	0.6930	0.0930	15.4950	0.0776	8.9351	≤ 0.001
σ_{44}	1.36	1.4628	0.1028	7.5551	0.1001	14.6143	≤ 0.001
Overall mean value			**0.0262**	**4.2400**	**0.0370**		

Note: "—" indicates that the corresponding parameter is fixed at the true value.

3 h 7 min (for ME), 1 h 14 min (for BME), and 2 h 6 min (for TVBS), respectively. Among them, the BME algorithm has obvious advantages in terms of time consumption. Hence, if the accuracy level can be lowered under some situations (e.g., in a variable screening process), the BME algorithm can also be chosen to save computational time. In this section, it is not necessary to evaluate the ICLV model of Bhat for comparison, because Bhat has explicitly shown that the LDLT matrix-based analytic approximation method runs much faster than other methods (such as the SSJ method) in evaluating MVNCD functions and estimating MNP model. It is worth noting that all simulation experiments are

TABLE 6: Simulation results for the four-alternative ICLV model with the TVBS method.

Parm.	True value	Estimate	Abs. bias	APB (%)	Std. err.	t-test	P value
α_1	0.5	0.4820	0.0180	3.6006	0.0175	27.4956	≤ 0.001
α_2	0.6	0.5733	0.0267	4.4550	0.0210	27.2643	≤ 0.001
α_3	0.5	0.5621	0.0621	12.4248	0.0541	10.3956	≤ 0.001
α_4	0.6	0.6657	0.0657	10.9532	0.0640	10.4023	≤ 0.001
α_5	0.3	0.3042	0.0042	1.3913	0.0104	29.3686	≤ 0.001
α_6	0.3	0.3058	0.0058	1.9333	0.0305	10.0157	≤ 0.001
α_7	−0.4	−0.4098	0.0098	2.4453	0.0408	−10.0548	≤ 0.001
s_1	1	—	—	—	—	—	—
r_{12}	0	—	—	—	—	—	—
r_{13}	0.6	0.5986	0.0014	0.2263	0.0460	13.0038	≤ 0.001
r_{14}	0	—	—	—	—	—	—
s_2	1	—	—	—	—	—	—
r_{23}	0	—	—	—	—	—	—
r_{24}	0.6	0.6472	0.0472	7.8628	0.1002	6.4570	≤ 0.001
s_3	1	—	—	—	—	—	—
r_{34}	0	—	—	—	—	—	—
s_4	1	—	—	—	—	—	—
δ_1	−1	−1.0068	0.0068	0.6844	0.0115	−87.6833	≤ 0.001
δ_2	−1	−0.9671	0.0329	3.2898	0.0161	−60.1169	≤ 0.001
δ_3	−1	−1.0078	0.0078	0.7807	0.0131	−76.7399	≤ 0.001
δ_4	−1	−0.9876	0.0124	1.2408	0.0373	−26.4477	≤ 0.001
d_1	0.3	0.3148	0.0148	4.9407	0.0125	25.2747	≤ 0.001
d_2	0.4	0.3509	0.0491	12.2728	0.0379	9.2709	≤ 0.001
d_3	0.5	0.4989	0.0011	0.2154	0.0210	23.7248	≤ 0.001
d_4	0.6	0.5913	0.0087	1.4555	0.0793	7.4595	≤ 0.001
ψ_1	1.5	1.5120	0.0120	0.8026	0.0081	186.2650	≤ 0.001
ψ_2	1.5	1.4793	0.0207	1.3799	0.0187	79.0895	≤ 0.001
ψ_3	1.5	1.4996	0.0004	0.0299	0.0145	103.1776	≤ 0.001
ψ_4	1.5	1.5160	0.0160	1.0677	0.0538	28.1597	≤ 0.001
β_2	0.5	0.4544	0.0456	9.1144	0.0304	14.9438	≤ 0.001
β_3	−1	−1.0425	0.0425	4.2479	0.0476	−21.8930	≤ 0.001
β_4	−0.8	−0.8728	0.0728	9.1025	0.0445	−19.6065	≤ 0.001
β_{tt}	−1	−1.0116	0.0116	1.1607	0.0158	−63.9212	≤ 0.001
β_{tc}	−0.8	−0.8109	0.0109	1.3567	0.0127	−63.6051	≤ 0.001
λ_1	0.5	0.5382	0.0382	7.6404	0.0234	22.9612	≤ 0.001
λ_2	0.2	0.1860	0.0140	6.9835	0.0185	10.0749	≤ 0.001
λ_3	0.2	0.1878	0.0122	6.1180	0.0186	10.0900	≤ 0.001
λ_4	0.5	0.4887	0.0113	2.2598	0.0232	21.0338	≤ 0.001
λ_5	0.5	0.5099	0.0099	1.9814	0.0212	24.0589	≤ 0.001
λ_6	0.2	0.1997	0.0003	0.1510	0.0204	9.7806	≤ 0.001
σ_{11}	1	—	—	—	—	—	—
σ_{12}	0	—	—	—	—	—	—
σ_{13}	0	—	—	—	—	—	—
σ_{14}	0	—	—	—	—	—	—
σ_{22}	1	—	—	—	—	—	—
σ_{23}	0.6	0.5923	0.0077	1.2832	0.0524	11.3058	≤ 0.001
σ_{24}	0.6	0.6382	0.0382	6.3702	0.0458	13.9236	≤ 0.001
σ_{33}	1.36	1.3895	0.0295	2.1659	0.1074	12.9430	≤ 0.001
σ_{34}	0.6	0.6875	0.0875	14.5838	0.0775	8.8672	≤ 0.001
σ_{44}	1.36	1.4629	0.1029	7.5643	0.1001	14.6142	≤ 0.001
Overall mean value			**0.0259**	**4.2037**	**0.0370**		

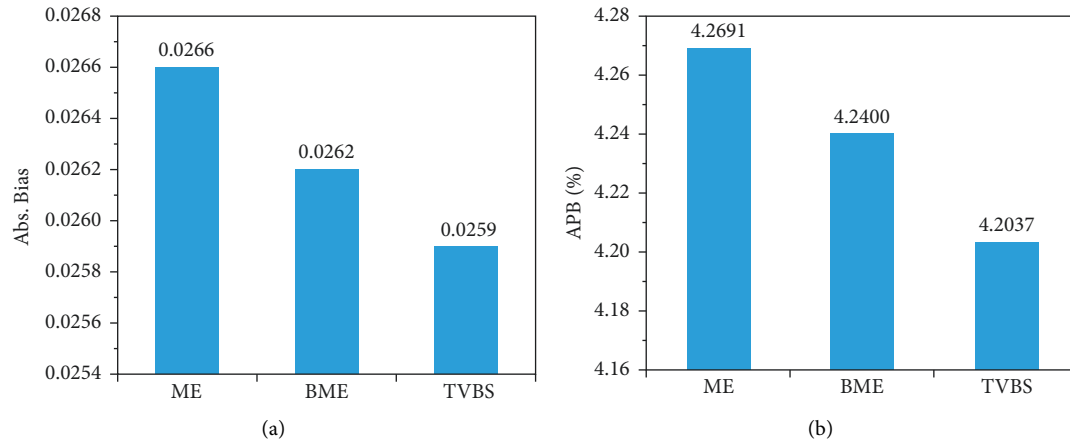*Note.* "—" indicates that the corresponding parameter is fixed at the true value.

Figure 5: Comparison of three methods for the four-alternative ICLV model: (a) overall mean value of Abs. bias and (b) overall mean value of APB (%).

conducted on a desktop computer with an Intel Core™ i5-4460@3.20 GHz Processor and 16 GB of RAM.

5. Conclusions and Discussion

On the basis of the research advances of Bhat in 2018 and the automatic differentiation (AD) technique on the TensorFlow platform, this study integrates the LDLT matrix-based analytic approximation method into the probit kernel-based ICLV formulation, which can further improve the performance of the ICLV model in terms of accuracy level and computational speed. The main achievements of this study can be summarized as follows.

Firstly, the individual MVNCD function was evaluated with different dimensions. By improving the numerical integration algorithm of CDF of BVN and TVN, the accuracy level of BVN and TVN functions was enhanced in this study (compared with algorithms in GAUSS). Based on the TensorFlow platform, we also tested the errors and operation time of three types of matrix-based analytic approximation methods in various dimensions. It is found that the operation time of BME is the shortest and the accuracy level of TVBS is the highest in the same dimension. By comparing the calculation of MVNCD functions on Gauss and TensorFlow platforms, it shows that the calculation speed of TensorFlow is much higher than that of Gauss.

Subsequently, we designed the simulation experiments of the three-alternative ICLV model and four-alternative ICLV model. The estimates and standard errors of both models are within a reasonable range, of which the parameters related to latent variables may be more difficult to recover. For the three-alternative ICLV model, the numerical integration method of calculating TVN value was employed and the simulation experiments were repeated 50 times with a sample size of 10,000. The simulation results show that the overall mean value of APB is 3.79% and the relative efficiency is 1.1190, exhibiting a desirable capability of parameter recovery. After that, this study applies three analytic approximation methods (i.e., ME, BME, and TVBS)

in the MVNCD function calculation of the ICLV model and conducts simulation experiments with a sample size of 100,000 regarding the four-alternative ICLV model, and significant results of consistent estimation value can be obtained. The average values of absolute bias and APB decreased in the order of ME, BME, and TVBS, indicating the increased accuracy. To sum up, when we estimate the ICLV model with high-dimensional integration, the TVBS method can be chosen with the highest accuracy level. Also, the BME method is an alternative to save computational time.

Meanwhile, with the assistance of the AD technique, the researchers are freed from coding analytical gradients. Compared with the probit-based ICLV model proposed by Bhat, the computational time of the proposed model is reduced remarkably, which is mainly attributed to two reasons: one is adopting the LDLT matrix-based analytic approximation methods to evaluate MVNCD function; the other is based on the advantages of TensorFlow in calculating tensor (represented as multidimensional array). The operation of a three-alternative ICLV model with a sample size of 10,000 only takes 8 min with the time advantage of LDLT-based analytic approximation methods, which can facilitate the applications of the proposed method in practice.

In the future work, we will further explore the practical application of the improved ICLV model with multiple alternatives and incorporate more psychological attitude variables into the model so as to explore the potential influencing mechanism of travel behaviors by using real data. If possible, further comparisons in terms of parameter estimation results between the ICLV model and previous models will be discussed.

Data Availability

The data used to support the findings of this study are simulation data and can be obtained from the corresponding author upon request.

Conflicts of Interest

Acknowledgments

References

[1] A. De Witte, J. Hollevoet, F. Dobruszkes, M. Hubert, and C. Macharis, "Linking modal choice to motility: a comprehensive review," *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 329–341, 2013.

[2] M. Ben-Akiva, J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou, "Integration of choice and latent variable models," *In Perpetual Motion*, pp. 431–470, 2002.

[3] D. Bolduc and M. Ben-Akiva, *Hybrid Choice Models with Logit Kernel: Applicability to Large Scale Models*, Emerald Group Publishing Limited, Bingley, UK, 2005.

[4] M. Vredin Johansson, T. Heldt, and P. Johansson, "The effects of attitudes and personality traits on mode choice," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp. 507–525, 2006.

[5] D. Temme, M. Paulssen, and T. Dannewald, "Incorporating Latent Variables into Discrete Choice Models - A Simultaneous Estimation Approach Using SEM Software," *Business Research*, vol. 1, no. 2, pp. 220–237, 2008.

[6] A. Daly, S. Hess, B. Patruni, D. Potoglou, and C. Rohr, "Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail Travel behaviour," *Transportation*, vol. 39, no. 2, pp. 267–297, 2012.

[7] R. A. Daziano and D. Bolduc, "Incorporating pro-environmental preferences towards green automobile technologies through a bayesian hybrid choice model," *Transportmetrica A: Transport Science*, vol. 9, no. 1, pp. 74–106, 2013.

[8] C. R. Bhat, "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model," *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 677–693, 2001.

[9] X. Fu and Z. Juan, "Estimation of multinomial probit-kernel integrated choice and latent variable model: comparison on one sequential and two simultaneous approaches," *Transportation*, vol. 44, no. 1, pp. 91–116, 2017.

[10] C. R. Bhat and S. K. Dubey, "A new estimation approach to integrate latent psychological constructs in choice modeling," *Transportation Research Part B: Methodological*, vol. 67, pp. 68–85, 2014.

[11] C. R. Bhat, S. K. Dubey, and K. Nagel, "Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice," *Transportation Research Part B: Methodological*, vol. 78, pp. 341–363, 2015.

[12] M. Kamargianni, S. Dubey, A. Polydoropoulou, and C. Bhat, "Investigating the subjective and objective factors influencing teenagers' school travel mode choice - An integrated choice and latent variable model," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 473–488, 2015.

[13] C. R. Bhat, C. Varin, and N. Ferdous, "A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model," *Maximum Simulated Likelihood Methods and Applications*, vol. 26, pp. 65–106, 2010.

[14] N. R. Mendell and R. C. Elston, "Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks," *Biometrics*, vol. 30, no. 1, pp. 41–57, 1974.

[15] C. R. Bhat, "The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 923–939, 2011.

[16] 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.

[17] S. Linnainmaa, "Taylor expansion of the accumulated rounding error," *BIT*, vol. 16, no. 2, pp. 146–160, 1976.

[18] H. J. Skaug, "Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models," *Journal of Computational and Graphical Statistics*, vol. 11, no. 2, pp. 458–470, 2002.

[19] D. A. Fournier, H. J. Skaug, J. Ancheta et al., "AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models," *Optimization Methods and Software*, vol. 27, no. 2, pp. 233–249, 2012.

[20] V. P. Godambe, "An optimum property of regular maximum likelihood estimation," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1208–1211, 1960.

[21] Y. Zhao and H. Joe, "Composite likelihood estimation in multivariate data analysis," *Canadian Journal of Statistics*, vol. 33, no. 3, pp. 335–356, 2005.

[22] R. D. Connors, S. Hess, and A. Daly, "Analytic approximations for computing probit choice probabilities," *Transportmetrica A: Transport Science*, vol. 10, no. 2, pp. 119–139, 2014.

[23] P. N. Patil, S. K. Dubey, A. R. Pinjari, E. Cherchi, R. Daziano, and C. R. Bhat, "Simulation evaluation of emerging estimation techniques for multinomial probit models," *Journal of Choice Modelling*, vol. 23, pp. 9–20, 2017.

[24] D. Kim, J. Ko, and Y. Park, "Factors affecting electric vehicle sharing program participants' attitudes about car ownership and program participation," *Transportation Research Part D: Transport and Environment*, vol. 36, pp. 96–106, 2015.

[25] A. Genz, "Numerical computation of rectangular bivariate and trivariate normal and t probabilities," *Statistics and Computing*, vol. 14, no. 3, pp. 251–260, 2004.

[26] Z. Drezner and G. O. Wesolowsky, "On the computation of the bivariate normal integral," *Journal of Statistical Computation and Simulation*, vol. 35, no. 1-2, pp. 101–107, 1990.

[27] H. Gassmann, "Rectangle probabilities of trivariate normal distributions. Unpublished preprint available from website," 2000, https://link.springer.com.

[28] A. Verma, "An introduction to automatic differentiation," *Current Science*, vol. 78, no. 7, p. 4, 2000.