Biases in the Maximum Simulated Likelihood Estimation of the Mixed Logit Model

Abstract

Jumamyradov and Munkin (2021) showed that the maximum simulated likelihood (MSL) estimator produces significant biases when applied to the bivariate normal and bivariate Poisson-lognormal models. Their conclusion is that similar biases can be present in other models generated by correlated bivariate normal structures, which include most commonly used specifications of the mixed logit (MIXL) models. This paper conducts a simulation study analyzing MSL estimation of the error-components (EC) MIXL. We find that the MSL estimator produces significant biases in the estimated parameters, leading to up to 12% difference in the true and estimated marginal effects.

Keywords: Maximum Simulated Likelihood, Mixed Logit.

1 Introduction

This paper examines the maximum simulated likelihood (MSL) estimator of the error components (EC) mixed logit (MIXL) model. The MIXL has been preferred by applied economists due to its flexible latent structure allowing various specifications of behavioral patterns. Since the model does not have a closed-form, its estimation relies on simulation-based methods, specifically the MSL estimator, which has been the dominant estimation strategy for more than 20 years. However, Jumamyradov and Munkin (2021) showed that the MSL estimator produces significant biases when applied to the bivariate normal and bivariate Poisson-lognormal models. Their conclusion is that similar biases can be present in other models generated by correlated bivariate normal structures, which include most commonly used specifications of the MIXL models. Therefore, further analysis of the MSL estimator in the context of the MIXL model is necessary.

The multinomial logit (MNL) model was introduced by McFadden (1974). It has a closed-form solution due to two convenient, however, restrictive assumptions. First, the MNL model assumes that the error terms are independently and identically distributed (i.i.d.) as type 1 extreme value (EV1) across individuals and alternatives. As a result, the MNL model suffers from the independence from irrelevant alternatives (IIA) property (Debreu, 1960), which in the literature has been illustrated by the ”red-bus, blue-bus” example (Quandt, 1970). Second, the MNL model does not allow unobserved variation of individual tastes (i.e., taste heterogeneity) in the population, meaning that the coefficients associated with alternative specific variables and observable alternative attributes that vary over individuals are fixed. Although the MNL model has become the ”workhorse” in discrete choice analysis (Hensher and Greene, 2003), its inconsistencies with realistic behavioral patterns have led researchers to look for more flexible alternative models. The MIXL model was derived by relaxing these restrictive assumptions (see McFadden, 2001).

The first contribution in the development of the MIXL model came with relaxing the assumption of homogeneous parameters. Specifically, Boyd and Mellman (1980) and Cardell and Dunbar (1980) analyzed market demand for automobiles by allowing consumer taste coefficients associated with attributes of the alternatives to vary over individuals in the form of random variables representing random taste heterogeneity (i.e., taste patterns). This specification of the MIXL model is also known as the random coefficients, with application examples including Revelt and Train (1998) and Bhat (2000). Revelt and Train (1998) analyzed households' choices of efficiency levels for refrigerators based on rebates and loans using panel data MIXL model. Bhat (2000) studied urban work travel mode choices by incorporating observed and unobserved individual characteristics into panel data

MIXL model. Like random coefficients, alternative-specific constants (ASCs) may not be homogeneous within a sample, potentially leading to substitution patterns (e.g., red-bus-blue-bus)(Quandt, 1970). The challenges of accommodating such heterogeneity is well-known in choice modelling (see Hensher et al., 2005).

Next, the i.i.d. assumption of the MNL model was relaxed allowing non-independent and non-identical errors, leading to the EC MIXL and generalized mixed logit (GMIXL). The EC specification of the MIXL model assumes that the stochastic portion of the utility consists of two parts, the i.i.d. errors with EV1 distribution and additional components varying over alternatives and individuals. This specification induces various correlation structures (i.e., taste and substitution patterns) as well as heteroskedasticity through nests or cross-nests created among alternatives as a result of shared error components. Brownstone and Train (1998) used this approach to forecast new product penetration rates by allowing flexible substitution patterns among alternative sources of fuel for vehicles.

Recent studies related to discrete choice modelling recognized the necessity for heterogeneity of the scale parameter (see Louviere et al., 1999; Louviere et al., 2002; Louviere et al., 2008), which lead to another specification of the MIXL model that relaxes the i.i.d. assumption. The scale parameter is directly related to the variance of the EV1 error terms, and usually is restricted to 1 because it cannot be identified separately from the slope coefficients. However, Fiebig et al. (2010) as well as Greene and Hensher (2010) proposed the generalized mixed logit (GMIXL) model that allows individual variation in the variance of the EV1 error terms (i.e., scale heterogeneity) along with unobserved individual heterogeneity in the slope coefficients. Although it has been shown that the GMIXL model performs better than the standard MIXL model (Fiebig et al., 2010; Keane and Wasi, 2013), Hess and Train (2017), as well as Hess and Rose (2012) raised concerns about the identifiability of the GMIXL model. It is an open research question, what additional assumptions need to be imposed to make the GMIXL model estimable.

The flexibility of the MIXL model is achieved by introducing latent variables into the model. However, it leads to intractability of the choice probabilities, which cannot be evaluated analytically since they do not have a closed-form. Therefore, estimation of the MIXL model relies on numerical approximation of the choice probabilities through simulation. The MSL estimator was introduced by Lerman and Manski (1980) to replace intractable choice probabilities of the multinomial probit (MNP) model with simulated probabilities.

A well-known limitation of the MSL estimator is that it is biased when the number of simulations is fixed (see Gourieroux and Monfort, 1996; Lee, 1995; and Hajivassiliou et al., 1996; Train, 2009). Nevertheless, estimation of the MIXL model in the literature is based on the MSL estimator, including studies by Ben-Akiva et al. (1993), Revelt and Train (1998), Bhat (1998), Brownstone and Train

(1998), McFadden and Train (2000), Hess et al. (2005). The usual practice is to use MSL in combination with Halton draws to reduce the simulation bias. Bhat (2001) showed that 100 Halton draws provide better approximation results than 1000 pseudo-random draws for the mixed logit model. According to Palma et al. (2020), around 93% of over 150 papers indexed in the Research Papers in Economics (RePEc) produced during 2008-2018 use less than 1000 Halton draws in their estimation of the mixed logit model. Furthermore, 72% and 40% of these papers use less than 500, and 250 Halton draws, respectively. Czajkowski and Budziński (2019) find that more than 3000 Halton draws are necessary to achieve a Minimum Tolerance Level of 5%. However, in the RePEc database, only 5.6% of papers used more than 2000 Halton draws (Palma et al., 2020). In this paper, we simulate MIXL data and assess the MSL performance based the difference between the true and estimated parameters.

Jumamyradov and Munkin (2021) primarily focused their analysis on estimation of the correlation parameter in the bivariate normal and bivariate Poisson-lognormal models. In this paper, we closely follow their strategy and allow correlation across utility of different alternatives. We also utilize Halton draws and analyze two error components specifications of the MIXL model. The first specification is the MIXL model with correlated slope coefficients and fixed alternative specific coefficients (ASCs). The second example is the MIXL model with correlated ASCs and fixed slope coefficients. Moreover, for simplicity, we assume that there is only one attribute varying over alternatives and individuals. It should be noted that in most specifications of the MIXL model used by practitioners, the correlation parameter is assumed to be zero for simplicity, compromising robustness to the IIA property. However, practitioners are mostly interested in the estimated mean and variance of the random parameters. Nevertheless, our findings illustrate simulation biases with zero correlation based on the differences between the true and estimated parameters.

There have been several studies that compared MIXL results by estimator and software package. Huber and Train (2001), Regier et al. (2009), Haan et al. (2015) and Elshiewy et al. (2017) compare MSL and Bayesian estimation of the MIXL model. The first three of these studies are based on a single panel dataset. The last one uses cross-sectional and panel data with three empirical and four simulated datasets. Although Elshiewy et al. (2017) find MSL biases of the correlation parameter in the cross-sectional MIXL model, they only test two values (0.75 and 0.25). We analyze the MSL estimator with respect to an extensive range of values of the correlation parameter and standard deviation, as well as different numbers of Halton draws. To our best knowledge, an extensive Monte-Carlo simulation study like this has not been conducted before.

The rest of the paper is organized as follows. Section 2 introduces the MSL estimator generally. Section 3 presents the logit model specifications. Section 4 presents numerical examples with MIXL data simulation and MSL estimation results. Section 5 concludes.

2 Maximum Simulated Likelihood Estimator

The maximum likelihood (ML) estimator of parameter vector θ can be utilized when $f(y_i|x_i,\theta)$, the density of dependent variable y_i conditional on the vector of independent variables x_i, has a closed-form such that

$$\hat{\theta}_N = \arg\max_{\theta} \sum_{i=1}^{N} \log f(y_i|x_i,\theta),$$

where (y_i, x_i) is a set of independent observations for $i = 1,...,N$. However, ML is not feasible when $f(y_i|x_i,\theta)$ does not have a tractable closed-form. This can be because the density is specified only conditional on latent variables which cannot be integrated out. Then the MSL estimator is a possible alternative, which we define following Gourieroux and Monfort (1990) and Gourieroux and Monfort (1996). Suppose $\tilde{f}(y_i,x_i,u,\theta)$ is an unbiased simulator of the conditional density $f(y_i|x_i,\theta)$ such that

$$f(y_i|x_i,\theta) = E_u[\tilde{f}(y_i,x_i,u,\theta)|y_i,x_i]$$

where the distribution of u is known and independent of y_i and x_i. Then the MSL estimator of θ is defined as

$$\hat{\theta}_{SN} = \arg\max_{\theta} \sum_{i=1}^{N} \log \left[\frac{1}{S} \sum_{s=1}^{S} \tilde{f}(y_i,x_i,u_i^s,\theta) \right],$$

where u_i^s $(s = 1,...,S)$ are drawn independently for each individual i from the distribution of u_i. The MSL estimator is obtained by replacing the intractable conditional p.d.f. $f(y_i|x_i,\theta)$ with its unbiased approximation based on the simulator $\tilde{f}(y_i,x_i,u_i,\theta)$. However, although $\tilde{f}(y_i,x_i,u_i,\theta)$ is an unbiased simulator of $f(y_i|x_i,\theta)$, its log transformation $\log\tilde{f}(y_i,x_i,u_i,\theta)$ is not an unbiased simulator of $\log f(y_i|x_i,\theta)$, which results in simulation biases in the MSL estimator.

Asymptotic properties of the MSL estimator are determined by the relationship between S and N. For instance, the MSL estimator is biased when S is fixed and N tends to infinity (Property 1 in Gourieroux and Monfort, 1990). If S increases with N, then the MSL estimator is consistent (Property 2 in Gourieroux and Monfort, 1990). If S increases faster than $\sqrt{N}$ ($\sqrt{N}/S \to 0$), then the MSL estimator is also efficient, and therefore asymptotically equivalent to the ML estimator (Property 7 in Gourieroux and Monfort, 1990). In practice neither N nor S might be close enough to infinity. However, the expectation is that there are achievable levels large enough for the biases to become acceptably small.

3 Model Specifications

In this section we define the MNL model and two specifications of the EC MIXL model. We also provide detailed information on how to simulate the corresponding likelihood functions.

3.1 Random Utility Maximization

Discrete choice models are usually introduced based on the random utility maximization (RUM) theory (see McFadden, 1974), which states that utility of individual $i = 1,...,N$ from choosing alternative $j = 1,...,J$ can be presented as $U_{ij} = V_{ij} + \varepsilon_{ij}$, where V_{ij} is the observed part of the utility and ε_{ij} is the stochastic portion, unobserved by the researcher. Individual i will choose alternative j if and only if the level of utility associated with alternative j is higher than the levels associated with the other alternatives

$$\begin{aligned} P_{ij} &= P(U_{ij} > U_{ik}, \forall\, k \neq j) \\ P_{ij} &= P(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}, \forall\, k \neq j) \\ P_{ij} &= P(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}, \forall\, k \neq j) \end{aligned} \tag{1}$$

Since utilities are latent, choice probabilities are evaluated at relative measures where utility of one of the alternatives is taken as a reference. In order to calculate choice probabilities, distributional assumptions of the stochastic utility must be made. In the logit family of models, ε_{ij} is assumed to be independently and identically distributed (i.i.d.) across individuals and alternatives with extreme value type 1 (EV1) distribution. As a result, the difference of two i.i.d. EV1 error terms $\left(\varepsilon_{ik} - \varepsilon_{ij}\right)$ has a logistic distribution with the cumulative distribution function

$$P_{ij} = \frac{1}{1 + \sum_{k=1}^{J} \exp(V_{ik} - V_{ij})}, \forall\, k \neq j. \tag{2}$$

The observed utility V_{ij} is a function of individual characteristics and alternative attributes, and usually assumed to be linear in the parameters.

3.2 Multinomial Logit (MNL) Model

The MNL model is derived under the assumption that all coefficients are fixed, implying that all individuals in the population have homogeneous tastes. In this paper, we consider the case of three alternatives, in which the third alternative is restricted as a referent category. Therefore, we work with two utility differences $Z_{i1} = U_{i1} - U_{i3}$ and $Z_{i2} = U_{i2} - U_{i3}$ defined as

$$\begin{aligned} Z_{i1} &= \alpha_1 + \beta_1 x_{i1} + \varepsilon_{i1} \\ Z_{i2} &= \alpha_2 + \beta_2 x_{i2} + \varepsilon_{i2} \end{aligned} \tag{3}$$

where $\varepsilon_{i1} \overset{i.i.d.}{\sim} Logistic(0,1)$ and $\varepsilon_{i2} \overset{i.i.d.}{\sim} Logistic(0,1)$ are logistically distributed, x_{i1} and x_{i2} are alternative attributes, α_1 and α_2 are alternative-specific coefficients (ASC), and β_1 and β_2 are coefficient of alternative attributes. In some specifications, these coefficients are restricted to be equal $\beta_1 = \beta_2 = \beta$. In the numerical examples, we choose the distribution of the covariates to be standard normal such that $x_{i1} \overset{i.i.d.}{\sim} N(0,1)$ and $x_{i2} \overset{i.i.d.}{\sim} N(0,1)$. The observability conditions for the outcome variables y_{i1}, y_{i2} and y_{i3} are defined as

$$\begin{aligned} y_{i1} &= 1 \text{ if and only if } Z_{i1} \geq Z_{i2}, Z_{i1} \geq 0 \\ y_{i2} &= 1 \text{ if and only if } Z_{i2} > Z_{i1}, Z_{i2} \geq 0 \\ y_{i3} &= 1 \text{ if and only if } Z_{i1} < 0, Z_{i2} < 0. \end{aligned} \tag{4}$$

In other words, individual i chooses the alternative with the highest utility.

3.3 Mixed Logit (MIXL) Model

The assumption of homogeneous preferences leads to computationally convenient functional forms for the choice probabilities. However, preference homogeneity is not consistent with realistic behavioral patterns. Next, we present two specifications of the EC MIXL model that allow various taste and substitution patterns through correlation between the utilities of different alternatives. The first specification is the MIXL model with correlated slope coefficients and fixed ASCs. The second example is the MIXL model with correlated ASCs and fixed slope coefficients. We refer to these two examples as EC1 and EC2, respectively. Under the EC1 specification taste patterns, we assume that

$$\begin{aligned} Z_{i1} &= \alpha_1 + (\beta_1 + u_{i1}) x_{i1} + \varepsilon_{i1} \\ Z_{i2} &= \alpha_2 + (\beta_2 + u_{i2}) x_{i2} + \varepsilon_{i2} \end{aligned} \tag{5}$$

where u_{i1} and u_{i2} are jointly normally distributed $(u_{i1}, u_{i2}) \overset{i.i.d.}{\sim} N((0,0), \Sigma)$ with covariance matrix Σ. Similarly, under the EC2 specification substitution patterns, we assume

$$Z_{i1} = (\alpha_1 + u_{i1}) + \beta x_{i1} + \varepsilon_{i1} \qquad (6)$$
$$Z_{i2} = (\alpha_2 + u_{i2}) + \beta x_{i2} + \varepsilon_{i2}$$

where once again $(u_{i1}, u_{i2}) \overset{i.i.d.}{\sim} N((0,0), \Sigma)$. The covariance matrix in both cases is parametrized as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where restriction $\sigma_1 = 1$ is imposed for identification such that

$$\Sigma = \begin{bmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^2 \end{bmatrix} \qquad (7)$$

Define lower triangular matrix

$$L = \begin{bmatrix} 1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{bmatrix} \qquad (8)$$

to be Choleski decomposition of the covariance matrix such as $\Sigma = LL'$. Then bivariate normal u_{i1} and u_{i2} can be written as

$$\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} = L \begin{bmatrix} v_{i1} \\ v_{i2} \end{bmatrix} \qquad (9)$$

where $v_{i1} \overset{i.i.d.}{\sim} N(0,1)$ and $v_{i2} \overset{i.i.d.}{\sim} N(0,1)$ which helps us to approximate the simulated likelihood function drawing from the known densities.

Both EC1 and EC2 specifications induce correlation in utilities of different alternatives. EC1 specification allows for correlation through the coefficients associated with alternative attributes x_{i1} and x_{i2}. This correlation is known as taste patterns because the weights for an attribute are associated with the weights of another attribute. EC2 specification allows for correlations through the ASCs, similar to the classic red-bus-blue-bus example. This is also known as substitution patterns because the weights of an alternative are associated with those of another (e.g., red and blue bus). Each MIXL specification relaxes the preference homogeneity assumption in a slightly different way and may be warranted depending on the decision context.

3.4 Simulated Likelihood Function of MIXL

MIXL choice probabilities unconditional of the unobserved latent variables v_{i1} and v_{i2} can be written as integrals over the density $f(v_{i1}, v_{i2})$ such that

$$
\begin{aligned}
P(y_{i1}=1) &= \int_{v_{i1}} \int_{v_{i2}} \left[\frac{exp(V_{i1})}{1+exp(V_{i1})+exp(V_{i2})} \right] f(v_{i1}, v_{i2}) dv_{i2} dv_{i1} \\
P(y_{i2}=1) &= \int_{v_{i1}} \int_{v_{i2}} \left[\frac{exp(V_{i2})}{1+exp(V_{i1})+exp(V_{i2})} \right] f(v_{i1}, v_{i2}) dv_{i2} dv_{i1} \\
P(y_{i3}=1) &= \int_{v_{i1}} \int_{v_{i2}} \left[\frac{1}{1+exp(V_{i1})+exp(V_{i2})} \right] f(v_{i1}, v_{i2}) dv_{i2} dv_{i1}
\end{aligned} \tag{10}
$$

where the form of V_{i1} and V_{i2} depends on the EC model. In the EC1 specification

$$
\begin{aligned}
V_{i1} &= \alpha_1 + (\beta_1 + v_{i1}) x_{i1} \\
V_{i2} &= \alpha_2 + (\beta_2 + \sigma_2 \rho v_{i1} + \sigma_2 v_{i2} \sqrt{1-\rho^2}) x_{i2},
\end{aligned} \tag{11}
$$

and in the EC2 specification

$$
\begin{aligned}
V_{i1} &= (\alpha_1 + v_{i1}) + \beta x_{i1} \\
V_{i2} &= (\alpha_2 + \sigma_2 \rho v_{i1} + \sigma_2 v_{i2} \sqrt{1-\rho^2}) + \beta x_{i2}.
\end{aligned} \tag{12}
$$

The log-likelihood function to be maximized can be written as

$$
LL = \sum_{i=1}^{N} \left(\sum_{j=1}^{3} I\left\{ y_{ij} = 1 \right\} \ln P(y_{ij} = 1) \right). \tag{13}
$$

However, the choice probabilities in equation (10) do not have a closed form, and the log-likelihood function cannot be calculated analytically. Therefore, we approximate the choice probabilities through simulation and maximize the simulated log-likelihood function

$$
SLL = \sum_{i=1}^{N} \left(\sum_{j=1}^{3} I\left\{ y_{ij} = 1 \right\} \ln \widehat{P}(y_{ij} = 1) \right), \tag{14}
$$

where simulated choice probabilities are

$$\widehat{P}(y_{i1}=1)=\frac{1}{S}\sum_{s=1}^{S}\left[\frac{exp(V_{i1}^{s})}{1+exp(V_{i1}^{s})+exp(V_{i2}^{s})}\right] \tag{15}$$

$$\widehat{P}(y_{i2}=1)=\frac{1}{S}\sum_{s=1}^{S}\left[\frac{exp(V_{i2}^{s})}{1+exp(V_{i1}^{s})+exp(V_{i2}^{s})}\right]$$

$$\widehat{P}(y_{i3}=1)=\frac{1}{S}\sum_{s=1}^{S}\left[\frac{1}{1+exp(V_{i1}^{s})+exp(V_{i2}^{s})}\right]$$

and $s=1,\ldots,S$ represents the draw for v_{i1}^{s} and v_{i2}^{s} used to evaluate V_{i1}^{s} and V_{i2}^{s}.

4 MIXL Numerical Examples

We generate data according to the EC1 and EC2 models. For each data generation process, the following specifications are used: $\alpha_1=\alpha_2=-0.25$, $\beta=\beta_1=\beta_2=1$, $x_{i1}\sim N(0,1)$, $x_{i2}\sim N(0,1)$, $(u_{i1},u_{i2})\sim N((0,0),(1,\rho\sigma_2,\sigma_2^2))$, $i=1,\ldots,N$, and $N=1000$. We test three different values for $\sigma_2=(0.25,0.5,1)$ and generate data set for all values of the correlation parameters ρ ranging from -0.95 to 0.95 with increments of 0.05, $\rho=\{-0.95:0.05:0.95\}$. Hence, a total of $3\times 39=117$ covariance matrices for each specification is analyzed.

To examine the performance of the MSL estimator, we estimate the MIXL model under three different sets of restrictions imposed on the covariance matrix (M0, M1 and M2). Under M0, we did not impose any restrictions on the covariance matrix and estimated all parameters. Under M1, we restricted the correlation parameter to zero $\rho=0$ and estimated the remaining parameters. Finally, under M2, we restricted the correlation parameter to its true value ($\rho=TV$) and estimated the remaining parameters. In each example, the numbers of Halton draws are chosen to be $H=(250,500,1000)$, which is consistent with the levels used in leading MSL applications.

Table 1: MSL Estimates for the EC1 (Taste Patterns)

		$\rho = -0.95$								
		250 draws			500 draws			1000 draws		
		$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$	$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$	$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$
M0	$\hat{\alpha}_1$	-0.307	-0.299	-0.299	-0.299	-0.301	-0.303	-0.308	-0.298	-0.308
		(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
	$\hat{\alpha}_2$	-0.349	-0.342	-0.273	-0.349	-0.342	-0.296	-0.348	-0.344	-0.289
		(0.010)	(0.009)	(0.011)	(0.009)	(0.010)	(0.012)	(0.010)	(0.010)	(0.011)
	$\hat{\beta}_1$	1.131	1.122	1.163	1.133	1.141	1.154	1.132	1.137	1.159
		(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.012)
	$\hat{\beta}_2$	1.248	1.197	1.070	1.265	1.241	1.136	1.245	1.220	1.120
		(0.013)	(0.013)	(0.017)	(0.013)	(0.019)	(0.027)	(0.013)	(0.014)	(0.025)
	$\hat{\sigma}_2$	0.413	0.569	0.885	0.439	0.662	0.930	0.405	0.566	0.897
		(0.026)	(0.023)	(0.038)	(0.030)	(0.033)	(0.054)	(0.025)	(0.022)	(0.050)
	$\hat{\rho}$	-0.877	-0.938	-0.975	-0.875	-0.899	-0.970	-0.887	-0.904	-0.973
		(0.026)	(0.017)	(0.008)	(0.025)	(0.023)	(0.009)	(0.025)	(0.021)	(0.009)
M1	$\hat{\alpha}_1$	-0.308	-0.297	-0.294	-0.300	-0.302	-0.302	-0.308	-0.298	-0.302
		(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
	$\hat{\alpha}_2$	-0.342	-0.342	-0.295	-0.342	-0.337	-0.323	-0.342	-0.341	-0.323
		(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.012)	(0.010)	(0.010)	(0.012)
	$\hat{\beta}_1$	1.151	1.147	1.196	1.151	1.167	1.190	1.151	1.162	1.190
		(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.010)	(0.012)
	$\hat{\beta}_2$	1.250	1.227	1.179	1.261	1.254	1.264	1.249	1.239	1.263
		(0.014)	(0.015)	(0.020)	(0.016)	(0.019)	(0.025)	(0.014)	(0.016)	(0.025)
	$\hat{\sigma}_2$	0.305	0.522	1.013	0.322	0.563	1.107	0.320	0.508	1.108
		(0.033)	(0.036)	(0.042)	(0.035)	(0.039)	(0.049)	(0.031)	(0.032)	(0.048)
M2	$\hat{\alpha}_1$	-0.307	-0.295	-0.291	-0.299	-0.299	-0.299	-0.307	-0.297	-0.299
		(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
	$\hat{\alpha}_2$	-0.348	-0.354	-0.315	-0.348	-0.349	-0.341	-0.348	-0.352	-0.340
		(0.010)	(0.010)	(0.011)	(0.010)	(0.009)	(0.011)	(0.010)	(0.010)	(0.011)
	$\hat{\beta}_1$	1.127	1.114	1.148	1.126	1.132	1.140	1.127	1.129	1.140
		(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.010)	(0.012)
	$\hat{\beta}_2$	1.249	1.239	1.193	1.262	1.266	1.272	1.249	1.250	1.272
		(0.012)	(0.014)	(0.017)	(0.013)	(0.016)	(0.024)	(0.012)	(0.014)	(0.024)
	$\hat{\sigma}_2$	0.430	0.677	1.200	0.446	0.722	1.271	0.430	0.649	1.271
		(0.025)	(0.030)	(0.034)	(0.028)	(0.030)	(0.045)	(0.025)	(0.025)	(0.045)

In summary, new MIXL data sets are generated for all 117 values of the covariance matrices for both EC1 and EC2 specifications. For each data set, three specifications (M0, M1, M2) are estimated, each with three different numbers of Halton draws (250, 500, 1000). We repeat each simulation 100 times, $R = 100$, generating a new data set and collecting the MSL estimates. The reported results are based on the means and standard errors calculated for these 100 simulations.

4.1 Taste Patterns: EC1 Simulation Evidence

Table 1 presents MSL results (M0, M1, M2) for EC1 simulations with high and negative correlation value $\rho = -0.95$. This extreme case produces a few results that deserve attention. First, there are biases in all coefficient estimates under M0 specification. For instance, when the true values $\sigma_2 = 0.25$, $\rho = -0.95$, and $H = 250$ the estimated values for α_1, α_2, β_1 and β_2 are -0.307 (0.008), -0.349 (0.010), 1.131 (0.011) and 1.248 (0.013), respectively. In other words, $\hat{\alpha}_1$, $\hat{\alpha}_2$ $\hat{\beta}_1$ and $\hat{\beta}_2$ are separated from their true values by 7, 10, 12 and 19 standard errors, respectively and, therefore, the null hypothesis that $H_o : \alpha_1 = \alpha_2 = -0.25$ and $H_o : \beta_1 = \beta_2 = 1$ are overwhelmingly rejected. Notice also that there is no apparent reductions in the biases for α_1, α_2, β_1 and β_2 regardless whether we increase the true variance σ_2 or the number of Halton draws.

Second, MSL produce biased results for σ_2 in small true values regardless of the chosen number of Halton draws. For example, when $\sigma_2 = 0.25$ and $H = 250$, the estimated value of σ_2 is 0.413 (0.026), which is separated from its true value by 6.27 standard errors, therefore the null hypothesis $H_o : \sigma_2 = 0.25$ is rejected. However, the estimated σ_2 gets closer to its true value when we increase the variance. For example, when $\sigma_2 = 0.5$, the estimated σ_2 is 0.569 (0.023). The null hypothesis $H_o : \sigma_2 = 0.5$ is not rejected. The case when $\sigma_2 = 1$ produces a similar result. However, the biased results for small variances do not change with the numbers of Halton draws. For instance, when $H = 500$ and $H = 1000$ the estimated values for the true $\sigma_2 = 0.25$ are 0.439 (0.030) and 0.405 (0.025), respectively. In both of these cases the null hypothesis $H_o : \sigma_2 = 0.25$ is rejected. It is also interesting to notice that when the true standard deviation is $\sigma_2 = 1$, the estimated σ_2 are much smaller in M0 than in M1 and M2.

Third, for almost all parameter sets presented in Table 1, the estimated ρ is within three standard errors from its true values. The only case where the null hypothesis that $H_o : \rho = -0.95$ may be rejected is when $H = 250$ and $\sigma_2 = 1$. EC1 simulation results for all other 38 correlation values are provided in Supplementary Materials.

Figure 1 plots estimated $\hat{\rho}$ against their true values that range from -0.95 to 0.95 with increments of 0.05, where $\hat{\rho}$ is calculated as the averages of ρ MSL estimates under M0 specification, obtained based on 100 samples ($R = 100$) generated for the same set of true values and estimated with 1000 Halton draws ($H = 1000$). The diagonal black line represents the true value of ρ. The blue, red and green lines correspond to $\sigma_2 = 0.25$, $\sigma_2 = 0.5$ and $\sigma_2 = 1$, respectively. Figure 1 shows that $\hat{\rho}$ is mostly biased downward for $H = 1000$.

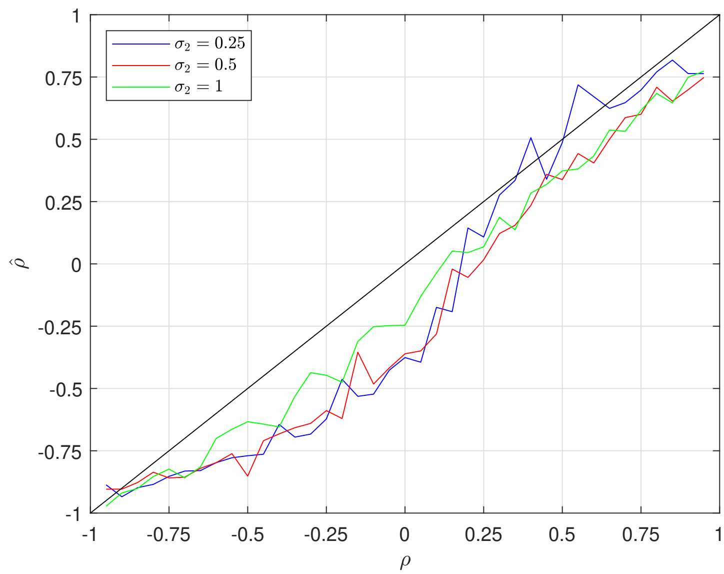

Figure 1: Plots of $\hat{\rho}$ for EC1 (Taste Patterns) using M0 (H=1000)

Finally, when the researcher erroneously assumes that the true correlation is zero (M1), there is no substantial worsening in the performance of MSL estimates. Similarly, when ρ is restricted at the true values (M2), there is no substantial improvement in the estimation of the parameters. A potential explanation for this is that the biases in MSL estimation of taste patterns are mostly caused by difficulties in estimating the correlation parameter with the efficiency of ρ estimates declining for smaller values of σ_2.

4.2 Substitution Patterns: EC2 Simulation Evidence

Table 2 presents M0, M1 and M2 results for the EC2 simulations when the true correlation is $\rho = -0.95$. First, notice that increasing the true value of variance σ_2 in M0 reduces the bias in α_1, α_2 and β. For example, given estimated α_1, α_2 and β as -0.274 (0.012), -0.314 (0.011) and 1.16 (0.009), respectively, we may reject the null hypothesis that $H_o : \alpha_2 = -0.25$ and $H_o : \beta = 1$. However, when we increase σ_2 to 1, the estimated α_1, α_2 and β are withing 3 standard errors or their respective true values. This conclusion is irrespective of the number of the Halton draws.

Second, the standard errors of estimated σ_2 are substantially larger than those of α_1, α_2 and β. As a result, the estimated σ_2 is within 3 standard errors from its true value across almost all parameter sets presented in Table 2. Therefore, the null hypothesis that σ_2 is equal to the true value cannot be rejected for almost all cases. The only exception is the case when $\sigma_2 = 1$ and $H = 1000$, and $\hat{\sigma}_2$ is 0.853 (0.043) and it is separated from the true value by 3.4 standard errors. The standard errors decrease slightly when correlation is restricted in specifications M1 and M2.

Table 2: MSL estimates for the EC2 (Substitution Patterns)

		$\rho = -0.95$								
		250 draws			500 draws			1000 draws		
		$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$	$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$	$\sigma_2 = 0.25$	$\sigma_2 = 0.5$	$\sigma_2 = 1$
M0	$\hat{\alpha}_1$	-0.274	-0.271	-0.241	-0.278	-0.246	-0.243	-0.285	-0.266	-0.238
		(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
	$\hat{\alpha}_2$	-0.314	-0.299	-0.227	-0.321	-0.289	-0.233	-0.349	-0.301	-0.216
		(0.011)	(0.013)	(0.012)	(0.014)	(0.015)	(0.014)	(0.014)	(0.013)	(0.012)
	$\hat{\beta}$	1.160	1.127	1.078	1.165	1.126	1.071	1.146	1.127	1.069
		(0.009)	(0.009)	(0.010)	(0.009)	(0.010)	(0.009)	(0.008)	(0.008)	(0.009)
	$\hat{\sigma}_2$	0.258	0.469	0.892	0.284	0.484	0.874	0.252	0.442	0.853
		(0.055)	(0.050)	(0.044)	(0.058)	(0.060)	(0.044)	(0.053)	(0.050)	(0.043)
	$\hat{\rho}$	-0.781	-0.810	-0.799	-0.726	-0.740	-0.819	-0.805	-0.768	-0.803
		(0.040)	(0.034)	(0.036)	(0.037)	(0.039)	(0.033)	(0.034)	(0.035)	(0.036)
M1	$\hat{\alpha}_1$	-0.243	-0.200	-0.095	-0.247	-0.188	-0.093	-0.255	-0.199	-0.095
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)
	$\hat{\alpha}_2$	-0.285	-0.241	-0.161	-0.292	-0.244	-0.171	-0.312	-0.247	-0.161
		(0.010)	(0.011)	(0.013)	(0.013)	(0.013)	(0.016)	(0.011)	(0.013)	(0.013)
	$\hat{\beta}$	1.152	1.111	1.085	1.157	1.117	1.080	1.132	1.113	1.084
		(0.010)	(0.009)	(0.011)	(0.009)	(0.011)	(0.012)	(0.008)	(0.010)	(0.011)
	$\hat{\sigma}_2$	0.335	0.534	1.199	0.452	0.660	1.199	0.365	0.558	1.211
		(0.052)	(0.053)	(0.058)	(0.049)	(0.053)	(0.060)	(0.041)	(0.050)	(0.055)
M2	$\hat{\alpha}_1$	-0.283	-0.280	-0.278	-0.283	-0.265	-0.281	-0.291	-0.278	-0.278
		(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.010)	(0.013)	(0.011)	(0.013)
	$\hat{\alpha}_2$	-0.319	-0.312	-0.308	-0.319	-0.303	-0.319	-0.347	-0.316	-0.308
		(0.013)	(0.015)	(0.016)	(0.016)	(0.017)	(0.019)	(0.015)	(0.016)	(0.016)
	$\hat{\beta}$	1.160	1.134	1.123	1.160	1.129	1.117	1.144	1.135	1.123
		(0.010)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.009)	(0.010)	(0.011)
	$\hat{\sigma}_2$	0.235	0.452	1.027	0.223	0.449	1.058	0.222	0.451	1.028
		(0.048)	(0.047)	(0.055)	(0.052)	(0.053)	(0.044)	(0.045)	(0.045)	(0.055)

Third, correlation parameter ρ is estimated with substantial biases in all M0 specifications. Estimated $\hat{\rho}$ is separated from the true value by 4 ($H = 1000$, $\sigma_2 = 1$) to 6 standard errors ($H = 500$, $\sigma_2 = 0.25$), and the null hypothesis $H_o : \rho = -0.95$ is rejected in all cases. EC2 results for the other 38 correlation values is provided in Supplementary Materials.

Figure 2 plots estimated $\hat{\rho}$ against the true values once again ranging from -0.95 to 0.95 with increments of 0.05. Although $\hat{\rho}$ is close to ρ for some values, the estimated correlation parameter mostly displays biases. The biases are smaller for $\sigma_2 = 1$ relative to when $\sigma_2 = 0.25$ or $\sigma_2 = 0.5$. This finding is consistent with Jumamyradov and Munkin (2021) in the bivariate normal and bivariate Poisson-lognormal models. They report larger biases for smaller standard deviations. Over-all M0 results show biases for all five parameters α_1, α_2, β, σ_2, ρ.

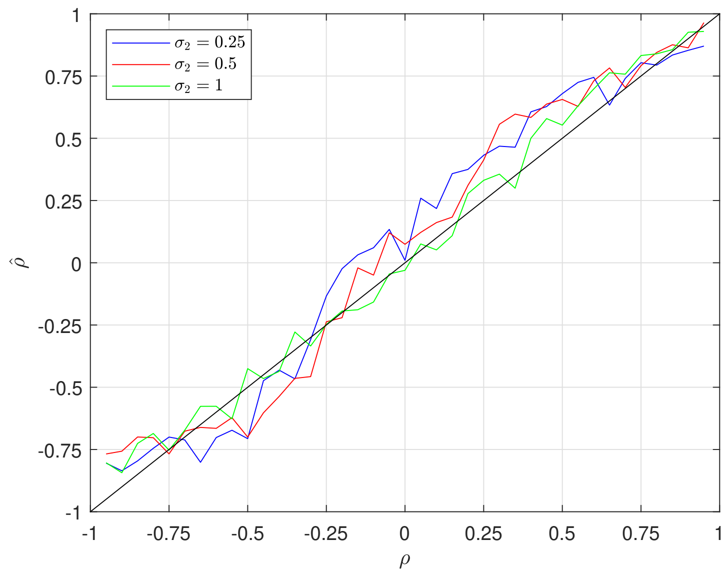

Figure 2: Plots of $\hat{\rho}$ for EC2 (Substitution Patterns) using M0 (H=1000)

It is also interesting to notice that MSL estimates of α_1 and α_2 in M1 have larger biases than in M0 for larger variances, and this is regardless of the number of Halton draws. For example, when $H = 250$ and $\sigma_2 = 1$, the estimated α_1 and α_2 are -0.095 (0.009) and -0.161 (0.013), separated from their true value $\alpha_1 = \alpha_2 = -0.25$ by 17 and 7 standard errors, respectively. This does not change much for larger values of Halton draws. Thus, misspecifying the model setting correlation $\rho = 0$ results in very large biases in α_1 and α_2. Moreover, M1 produces larger positive biases of σ_2 compared to M0. For example, when $H = 1000$, the estimated σ_2 are 0.365 (0.041), 0.558 (0.050) and 1.211 (0.055) for the true $\sigma_2 = 0.25$, $\sigma_2 = 0.5$ and $\sigma_2 = 1$, respectively. Moreover, the estimates of α_1, α_2 and β improve with larger variances in M0, however, we do not observe similar patterns in M2 estimation, although there is an improvement in estimation of σ_2.

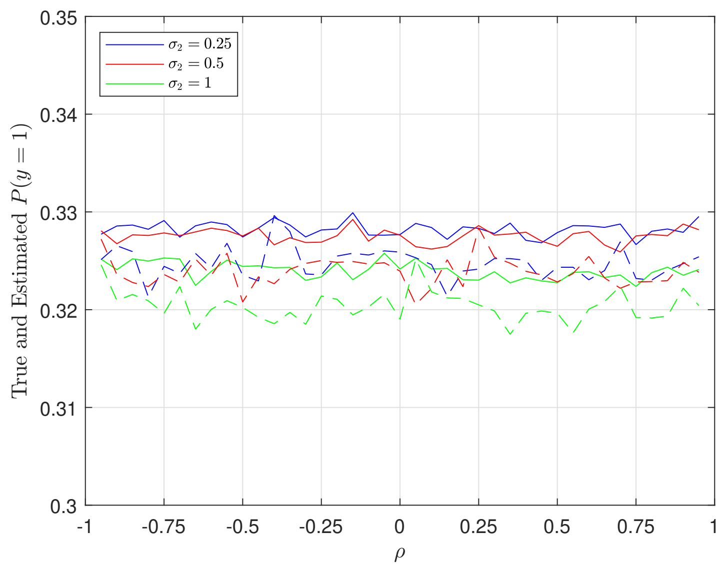

Figure 3: Plots of True and Estimated $P(y=1)$ for EC1 (Taste Patterns) using M0 (H=500)

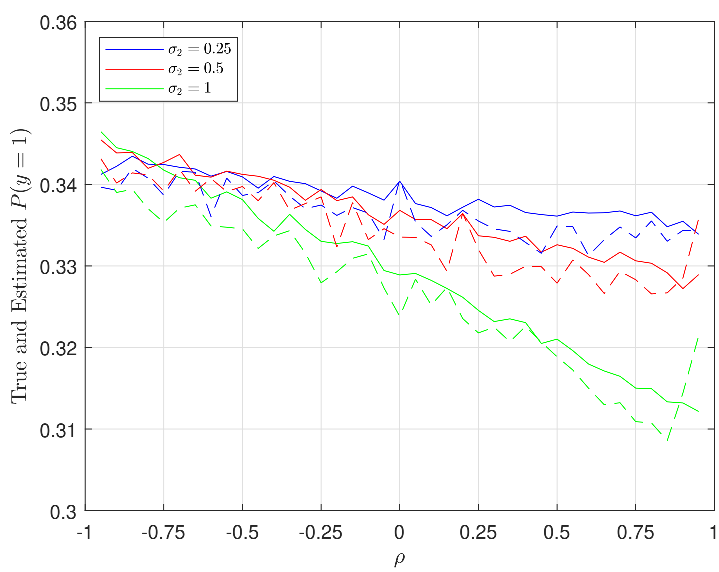

Figure 4: Plots of True and Estimated $P(y=1)$ for EC2 (Substitution Patterns) using M0 (H=500)

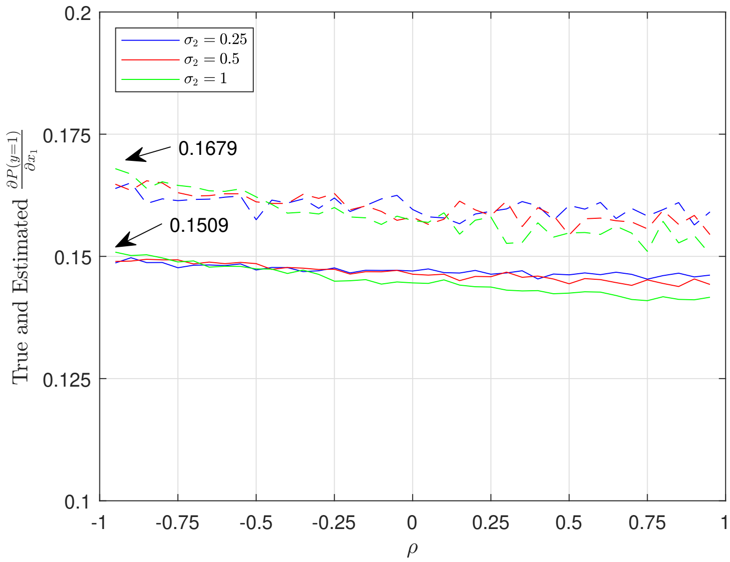

Figure 5: Plots of True and Estimated $\frac{\partial P(y=1)}{\partial x_1}$ for EC1 (Taste Patterns) using M0 (H=500)

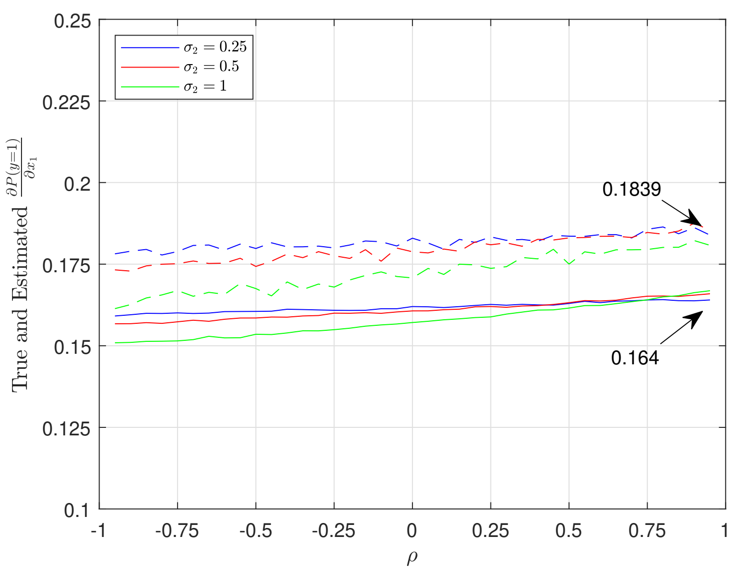

Figure 6: Plots of True and Estimated $\frac{\partial P(y=1)}{\partial x_1}$ for EC2 (Substitution Patterns) using M0 (H=500)

4.3 Choice Probabilities and Marginal Effects

Next we examine how these reported biases affect the estimated choice probabilities and marginal effects. Figure 3 plots true and estimated $P(y=1)$ calculated based on M0 estimates of EC1 specification with 500 Halton draws. True probability means are calculated at the true values of all parameters. Straight lines represent true choice probabilities and dashed lines represent estimated choice probabilities. Figure 4 plots true and estimated $P(y=1)$ based on M0 estimates of EC2 specification with 500 Halton draws. Even though there are significant biases in the estimated parameters, as expected the choice probabilities are close to their true values for both EC1 and EC2 specification (i.e., taste and substitution patterns).

However, when comparing true and estimated marginal effects, the differences are considerable. Figure 5 plots true and estimated $\partial P(y=1)/\partial x_1$ for EC1 (M0, 500 Halton draws). For example, when $\sigma_2 = 1$ and $\rho = 0.95$, the true $\partial P(y=1)/\partial x_1$ is 0.1509 and the estimated $\partial P(y=1)/\partial x_1$ is 0.1679. Thus, the marginal effect in this case is overestimated by 11%. Figure 6 plots true and estimated $\partial P(y=1)/\partial x_1$ for EC2 (M0, 500 Halton draws). For example, when $\sigma_2 = 1$ and $\rho = -0.95$, the true $\partial P(y=1)/\partial x_1$ is 0.164 and the estimated $\partial P(y=1)/\partial x_1$ is 0.1839, which is overestimated by 12%.

5 Conclusion

In this paper we examine properties of the MSL estimator in the context of two MIXL model specifications, EC1 and EC2 (i.e., taste and substitution patterns), where random parameters are generated by a correlated bivariate normal structure. We find that the MSL estimator produces significant biases in the estimated parameters. The problem becomes worse when the true value of the variance parameter is small and the correlation parameter is large in magnitude. Furthermore, we find that the marginal effects are biased as large as 12% of the true values. These biases are largely invariant to increases in the number of Halton draws. Since the existing literature has relied heavily on the MSL estimator in the analysis of the MIXL model our findings should be an important additional warning to researchers about potential sizable biases in the results.

We also discover that performance of MSL depends on other factors such as model specification (i.e. EC1 or EC2), distributional assumptions, exogenous variation, as well as true values of variance and correlation parameters. Therefore, we believe that biases in empirical applications (e.g., discrete choice experiments in health preference research) are likely to be worse due to real-world complexity, however, more research is needed to address such questions. Future simulation studies may examine biases in more complex specifications such as the generalized MIXL or EC MIXL with more than two random parameters.

Appendix

Marginal effects taken with respect to x_1 for EC1 are presented below. Marginal effects with respect to x_2 can be derived in the same way.

$$
\begin{aligned}
\frac{\partial \hat{P}(y=1)}{\partial x_1} &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial\left(\frac{exp(\hat{V}_{i1})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)}{\partial x_{i1}} \\
&= \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i1}\left(\frac{exp(\hat{V}_{i1})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)\left(1-\frac{exp(\hat{V}_{i1})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right) \\
\frac{\partial \hat{P}(y=2)}{\partial x_1} &= -\frac{1}{N}\sum_{i=1}^{N} \frac{\partial\left(\frac{exp(\hat{V}_{i2})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)}{\partial x_{i1}} \\
&= -\frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i1}\left(\frac{exp(\hat{V}_{i1})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)\left(\frac{exp(\hat{V}_{i2})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right) \\
\frac{\partial \hat{P}(y=3)}{\partial x_1} &= -\frac{1}{N}\sum_{i=1}^{N} \frac{\partial\left(\frac{1}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)}{\partial x_{i1}} \\
&= -\frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i1}\left(\frac{exp(\hat{V}_{i1})}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)\left(\frac{1}{1+exp(\hat{V}_{i1})+exp(\hat{V}_{i2})}\right)
\end{aligned}
$$

where the conditional mean $\hat{\beta}_{i1}$ is estimated by simulation as

$$
\hat{\beta}_{i1} = E(\hat{\beta}_{i1}^{q}|\mathbf{y},\mathbf{X}) = \frac{\frac{1}{Q}\sum_{q=1}^{Q}\hat{\beta}_{i1}^{q}\left(\frac{exp(\hat{V}_{i1}^{q})}{1+exp(\hat{V}_{i1}^{q})+exp(\hat{V}_{i2}^{q})}\right)}{\frac{1}{Q}\sum_{q=1}^{Q}\left(\frac{exp(\hat{V}_{i1}^{q})}{1+exp(\hat{V}_{i1}^{q})+exp(\hat{V}_{i2}^{q})}\right)}
$$

$$
\begin{aligned}
\hat{V}_{i1}^{q} &= \hat{\alpha}_1 + \hat{\beta}_{i1}^{q}x_{i1} \\
\hat{V}_{i2}^{q} &= \hat{\alpha}_2 + \hat{\beta}_{i2}^{q}x_{i2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_{i1}^{q} &= \hat{\beta}_1 + v_{i1}^{q} \\
\hat{\beta}_{i2}^{q} &= \hat{\beta}_2 + \hat{\sigma}_2\hat{\rho}v_{i1}^{q} + \hat{\sigma}_2 v_{i2}^{q}\sqrt{1-\hat{\rho}^2}
\end{aligned}
$$

where v_{i1}^{q} and v_{i2}^{q} are independent standard normal random variables for individual $i = 1,\ldots,N$ and random draws $q = 1,\ldots,Q$.

References

Ben-Akiva, M., D. Bolduc, and M. Bradley (1993): "Estimation of travel choice models with randomly distributed values of time," *Transportation Research Record*, 1413, 88–97.

Bhat, C. R. (1998): "Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice," *Transportation Research Part B: Methodological*, 32, 455–466.

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

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

Boyd, H. J. and R. E. Mellman (1980): "The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis," *Transportation Research Part A: General*, 14, 367–378.

Brownstone, D. and K. Train (1998): "Forecasting new product penetration with flexible substitution patterns," *Journal of Econometrics*, 89, 109–129.

Cardell, N. S. and F. C. Dunbar (1980): "Measuring the societal impacts of automobile downsizing," *Transportation Research Part A: General*, 14, 423–434.

Czajkowski, M. and W. Budziński (2019): "Simulation error in maximum likelihood estimation of discrete choice models," *Journal of Choice Modelling*, 31, 73–85.

Debreu, G. (1960): "Review of R.D. Luce, Individual Choice Behavior: A theoretical analysis," *American Economic Review*, 50, 186–188.

Elshiewy, O., G. Zenetti, and Y. Boztug (2017): "Differences between classical and Bayesian estimates for mixed logit models: A replication study," *Journal of Applied Econometrics*, 32, 470–476.

Fiebig, D., M. Keane, J. Louviere, and N. Wasi (2010): "The generalized multinomial logit model: accounting for scale and coefficient heterogeneity," *Marketing Science*, 29, 393–421.

Gourieroux, C. and A. Monfort (1990): "Simulation based inference in models with heterogeneity," *Annales d'Economie et de Statistique*, 69–107.

Gourieroux, C. and A. Monfort (1996): *Simulation-based econometric methods*, New York: Oxford University Press.

Greene, W. H. and D. A. Hensher (2010): "Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models," *Transportation*, 37, 413–428.

Haan, P., D. Kemptner, and A. Uhlendorff (2015): "Bayesian procedures as a numerical tool for the estimation of an intertemporal discrete choice model," *Empirical Economics*, 49, 1123–1141.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996): "Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results," *Journal of Econometrics*, 72, 85–134.

Hensher, D. A. and W. H. Greene (2003): "The mixed logit model: the state of practice," *Transportation*, 30, 133–176.

Hensher, D. A., J. M. Rose, and W. H. Greene (2005): *Applied choice analysis: a primer*, Cambridge: Cambridge University Press.

Hess, S., M. Bierlaire, and J. W. Polak (2005): "Estimation of value of travel-time savings using mixed logit models," *Transportation Research Part A: Policy and Practice*, 39, 221–236.

Hess, S. and J. M. Rose (2012): "Can scale and coefficient heterogeneity be separated in random coefficients models?" *Transportation*, 39, 1225–1239.

Hess, S. and K. Train (2017): "Correlation and scale in mixed logit models," *Journal of Choice Modelling*, 23, 1–8.

Huber, J. and K. Train (2001): "On the similarity of classical and Bayesian estimates of individual mean partworths," *Marketing Letters*, 12, 259–269.

Jumamyradov, M. and M. K. Munkin (2021): "Biases in maximum simulated likelihood estimation of bivariate models," *Journal of Econometric Methods*, 11, 55–70.

Keane, M. and N. Wasi (2013): "Comparing alternative models of heterogeneity in consumer choice behavior," *Journal of Applied Econometrics*, 28, 1018–1045.

Lee, L. (1995): "Asymptotic bias in simulated maximum likelihood estimation of discrete choice models," *Econometric Theory*, 11, 437–483.

Lerman, S. and C. Manski (1980): "On the use of simulated frequencies to approximate choice probabilities," in C. Manski and D. McFadden, eds., *Structural analysis of discrete data with econometric applications*, Cambridge, MA: MIT Press, 305–319.

Louviere, J. J., R. J. Meyer, D. S. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher, and J. Irwin (1999): "Combining sources of preference data for modeling complex decision processes," *Marketing Letters*, 10, 205–217.

Louviere, J. J., D. Street, L. Burgess, N. Wasi, T. Islam, and A. A. Marley (2008): "Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information," *Journal of Choice Modelling*, 1, 128–164.

Louviere, J. J., D. Street, R. Carson, A. Ainslie, J. DeShazo, T. Cameron, D. Hensher, R. Kohn, and T. Marley (2002): "Dissecting the random component of utility," *Marketing Letters*, 13, 177–193.

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in P. Zarambka, ed., *Frontiers in Econometrics*, New York: Academic Press, 105–142.

McFadden, D. (2001): "Economic choices," *American Economic Review*, 91, 351–378.

McFadden, D. and K. Train (2000): "Mixed mnl models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

Palma, M. A., D. V. Vedenov, and D. Bessler (2020): "The order of variables, simulation noise, and accuracy of mixed logit estimates," *Empirical Economics*, 58, 2049–2083.

Quandt, R. (1970): *The Demand for Travel: Theory and Measurement*, Lexington, MA: D.C. Heath and Company.

Regier, D. A., M. Ryan, E. Phimister, and C. A. Marra (2009): "Bayesian and classical estimation of mixed logit: an application to genetic testing," *Journal of Health Economics*, 28, 598–610.

Revelt, D. and K. Train (1998): "Mixed logit with repeated choices: households' choices of appliance efficiency level," *The Review of Economics and Statistics*, 80, 647–657.

Train, K. E. (2009): *Discrete choice methods with simulation*, New York: Cambridge university press.