

Article

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

Maksat Jumamyradov *, Murat Munkin , William H. Greene and Benjamin M. Craig

Department of Economics, University of South Florida, Tampa, FL 33620, USA; mmunkin@usf.edu (M.M.); greenew@usf.edu (W.H.G.); bcraig@usf.edu (B.M.C.)

* Correspondence: maksat@usf.edu

Abstract: In a recent study, it was demonstrated that the maximum simulated likelihood (MSL) estimator produces significant biases when applied to the bivariate normal and bivariate Poisson-lognormal models. The study's conclusion suggests that similar biases could be present in other models generated by correlated bivariate normal structures, which include several commonly used specifications of the mixed logit (MIXL) models. This paper conducts a simulation study analyzing the MSL estimation of the error components (EC) MIXL. 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. In some cases, the biases in the estimated marginal effects are as large as 12% of the true values. These biases are largely invariant to increases in the number of Halton draws.

Keywords: maximum simulated likelihood; mixed logit; discrete choice models

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, which allows for 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 could be present in other models generated by correlated bivariate normal structures, which include the 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 a type 1 extreme value (EV1) across the 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 for the unobserved variation in individual tastes (i.e., taste heterogeneity) in the population, meaning that the coefficients, associated with alternative-specific variables and observable alternative attributes that vary among 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](#)).



Citation: Jumamyradov, Maksat, Murat Munkin, William H. Greene, and Benjamin M. Craig. 2024. Biases in the Maximum Simulated Likelihood Estimation of the Mixed Logit Model. *Econometrics* 12: 8. <https://doi.org/10.3390/econometrics12020008>

Academic Editor: Willi Semmler

Received: 10 February 2024

Revised: 11 March 2024

Accepted: 23 March 2024

Published: 27 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The first contribution to 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 the consumer taste coefficients associated with the attributes of the alternatives to vary among 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 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 the panel data MIXL model. [Bhat \(2000\)](#) studied urban work travel mode choices by incorporating observed and unobserved individual characteristics into the panel data MIXL model. Similar to 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 are well-known in choice modeling (see [Hensher et al. 2005](#)).

Next, the i.i.d. assumption of the MNL model was relaxed, allowing for 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 an EV1 distribution and additional components varying among the alternatives and individuals. This specification induces various correlation structures (i.e., taste and substitution patterns) as well as heteroskedasticity through the nests or cross nests created among the alternatives as a result of shared error components. [Brownstone and Train \(1998\)](#) used this approach to forecast new product penetration rates by allowing for flexible substitution patterns among the alternative sources of fuel for vehicles.

Recent studies related to discrete choice modeling have recognized the necessity of the heterogeneity of the scale parameter (see [Louviere et al. 1999, 2002, 2008](#)), which led to another specification of the MIXL model that relaxed the i.i.d. assumption. The scale parameter is directly related to the variance of the EV1 error terms, and is usually restricted to one 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 for individual variation in the variance of the EV1 error terms (i.e., scale heterogeneity) along with the unobserved individual heterogeneity of the slope coefficients. Although it has been shown that the GMIXL model performed 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 as to 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, this leads to the intractability of the choice probabilities, which cannot be evaluated analytically since they do not have a closed form. Therefore, the estimation of the MIXL model relies on a numerical approximation of the choice probabilities through simulation. The MSL estimator was introduced by [Lerman and Manski \(1980\)](#) to replace the 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 limited, as is always the case in applications (see [Gourieroux and Monfort 1996; Lee 1995; Hajivassiliou et al. 1996; Train 2009](#)). Nevertheless, the estimation of the MIXL model in the literature is based on the MSL estimator, including in studies by [Ben-Akiva et al. \(1993\)](#), [Revelt and Train \(1998\)](#), [Bhat \(1998\)](#), [Brownstone and Train \(1998\)](#), [McFadden and Train \(2000\)](#), and [Hess et al. \(2005\)](#). The usual practice is to use the 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 used less

than 1000 Halton draws in their estimations of the mixed logit model. Furthermore, 72% and 40% of these papers used less than 500 and 250 Halton draws, respectively. [Czajkowski and Budziński \(2019\)](#) found 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](#)).

[Jumamyradov and Munkin \(2021\)](#) primarily focused their analysis on the estimation of the correlation parameter in the bivariate normal and bivariate Poisson-lognormal models. In this paper, we closely follow their strategy and allow for correlation across the utilities 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 that varies among the 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. In this paper, we simulate the data according to the MIXL model and assess the MSL performance based on the difference between the true and estimated parameters. Our findings confirm simulation biases even in cases with zero correlation.

There have been several studies that have compared the MIXL results produced by estimators and software packages. [Huber and Train \(2001\)](#), [Regier et al. \(2009\)](#), [Haan et al. \(2015\)](#), [Bastin and Cirillo \(2010\)](#), and [Elshiewy et al. \(2017\)](#) compared the MSL and Bayesian estimation of the MIXL model. The first three of these studies were based on a single panel dataset. [Bastin and Cirillo \(2010\)](#) estimated the simulation biases in the MSL with respect to the number of draws and sample sizes, without comparing the MSL with the Bayesian estimation. [Elshiewy et al. \(2017\)](#) used cross-sectional and panel data with three empirical and four simulated datasets. Although [Elshiewy et al. \(2017\)](#) found MSL biases in the correlation parameter of the cross-sectional MIXL model, they only tested 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 the best of our 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. Section 3 presents different logit model specifications. Section 4 presents numerical examples using MIXL data simulation and produces MSL estimation results. Section 5 discusses the results.

2. Maximum Simulated Likelihood Estimator

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

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

where (y_i, x_i) is a set of independent observations for $i = 1, \dots, N$. However, the ML is not feasible when $f(y_i|x_i, \theta)$ does not have a tractable closed form. This could be because the density is specified only conditionally on latent variables, which cannot be integrated out. Then, the MSL estimator is a possible alternative, which we define following [Gourieroux and Monfort \(1990, 1996\)](#). 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

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

where u_i^s ($s = 1, \dots, S$) are drawn independently for each individual i from the distribution 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 $\tilde{f}(y_i, x_i, u, \theta)$. However, although $\tilde{f}(y_i, x_i, u, \theta)$ is an unbiased simulator of $f(y_i|x_i, \theta)$, its log transformation $\log \tilde{f}(y_i, x_i, u, \theta)$ is not an unbiased simulator of $\log f(y_i|x_i, \theta)$, which results in simulation biases in the MSL estimator.

The 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} ($\sqrt{N}/S \rightarrow 0$), 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 or 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. Materials and Methods

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 the utility of individual $i = 1, \dots, N$ from the chosen alternative $j = 1, \dots, J$ 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:

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

Since the utilities are latent, the choice probabilities are evaluated at relative measures, where the utility of one of the alternatives is taken as the reference. In order to calculate the choice probabilities, the 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 an extreme value type 1 (EV1) distribution. As a result, the difference between two i.i.d. EV1 error terms ($\varepsilon_{ik} - \varepsilon_{ij}$) has a logistic distribution with the cumulative distribution function

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

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

3.2. Multinomial Logit (MNL) Model

The MNL model is derived under the assumption that all the coefficients are fixed, implying that all the 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 the referent category. Therefore, we work with two utility differences, $Z_{i1} = U_{i1} - U_{i3}$ and $Z_{i2} = U_{i2} - U_{i3}$, defined as

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

where $\varepsilon_{i1} \sim \text{Logistic}(0, 1)$ and $\varepsilon_{i2} \sim \text{Logistic}(0, 1)$ are i.i.d. 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 coefficients of the 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 $x_{i1} \sim N(0, 1)$ and $x_{i2} \sim N(0, 1)$. The observability conditions for the outcome variables y_{i1} , y_{i2} and y_{i3} are defined as

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

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 for various taste and substitution patterns through a correlation among the utilities of the 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

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

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

$$\begin{aligned} Z_{i1} &= (\alpha_1 + u_{i1}) + \beta x_{i1} + \varepsilon_{i1} \\ Z_{i2} &= (\alpha_2 + u_{i2}) + \beta x_{i2} + \varepsilon_{i2} \end{aligned} \quad (6)$$

where, once again, $(u_{i1}, u_{i2}) \sim N((0, 0), \Sigma)$. 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

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

We define the lower triangular matrix

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

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

$$\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} = A \begin{bmatrix} v_{i1} \\ v_{i2} \end{bmatrix} \quad (9)$$

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

Both the EC1 and EC2 specifications induce correlation in the utilities of the different alternatives. The 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. The EC2 specification allows for correlations through the ASCs, similar to the classic red bus–blue bus example. This is also known as the 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

The 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

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

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

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

and in the EC2 specification,

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

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

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

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

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

where the simulated choice probabilities are

$$\begin{aligned}\hat{P}(y_{i1} = 1) &= \frac{1}{S} \sum_{s=1}^S \left[\frac{\exp(V_{i1}^s)}{1 + \exp(V_{i1}^s) + \exp(V_{i2}^s)} \right] \\ \hat{P}(y_{i2} = 1) &= \frac{1}{S} \sum_{s=1}^S \left[\frac{\exp(V_{i2}^s)}{1 + \exp(V_{i1}^s) + \exp(V_{i2}^s)} \right] \\ \hat{P}(y_{i3} = 1) &= \frac{1}{S} \sum_{s=1}^S \left[\frac{1}{1 + \exp(V_{i1}^s) + \exp(V_{i2}^s)} \right]\end{aligned}\quad (15)$$

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

4. Results

We generate data according to the EC1 and EC2 models. For each data generation process, the following specifications are used: $\alpha_1 = \alpha_2 = -0.25$, $\beta = \beta_1 = \beta_2 = 1$, $x_{i1} \sim N(0, 1)$, $x_{i2} \sim N(0, 1)$, $(u_{i1}, u_{i2}) \sim N((0, 0), (1, \rho\sigma_2, \sigma_2^2))$, $i = 1, \dots, N$, and $N = 1000$. We test three different values for $\sigma_2 = (0.25, 0.5, 1)$ and generate data sets for all values of the correlation parameter 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 do not impose any restrictions on the covariance matrix and estimate all the parameters. Under M1, we restrict the correlation parameter to zero, $\rho = 0$, and estimate the remaining parameters. Finally, under M2, we restrict the correlation parameter to its true value ($\rho = TV$) and estimate the remaining parameters. In each example, the number of Halton draws are chosen to be $H = (250, 500, 1000)$, which is consistent with the levels used in leading MSL applications.

In summary, new MIXL data sets are generated for all 117 values of the covariance matrices for both the EC1 and EC2 specifications. For each data set, three specifications (M0, M1, and M2) are estimated, each with three different numbers of Halton draws (250, 500, and 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 (i.e., Wald test).

4.1. Taste Patterns: EC1 Simulation Evidence

Table 1 presents the MSL results (M0, M1, M2) for the EC1 simulations with a high and negative correlation value of $\rho = -0.95$. This extreme case produces a few results that deserve attention. First, there are biases in all the coefficient estimates under the M0 specification. For instance, when the true values are $\sigma_2 = 0.25$, $\rho = -0.95$, 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 hypotheses that $H_0 : \alpha_1 = \alpha_2 = -0.25$ and $H_0 : \beta_1 = \beta_2 = 1$ are overwhelmingly rejected. Notice also that there are no apparent reductions in the biases for α_1 , α_2 , β_1 , and β_2 , regardless of whether we increase the true variance σ_2 or the number of Halton draws.

Second, the MSL produces biased results for σ_2 with 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 $H_0 : \sigma_2 = 0.25$ is rejected. However, the estimated σ_2 comes 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 case when $\sigma_2 = 1$ produces a similar result. However, the biased results for small variances do not change with the number 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

$H_0 : \sigma_2 = 0.25$ is rejected. It is also interesting to notice that when the true standard deviation is $\sigma_2 = 1$, the estimated σ_2 is much smaller for M0 than for M1 and M2.

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

Third, for almost all the parameter sets presented in Table 1, the estimated ρ is within three standard errors from its true value. The only case where the null hypothesis, that $H_0 : \rho = -0.95$, may be rejected is when $H = 250$ and $\sigma_2 = 1$. The EC1 simulation results for all the other 38 correlation values are provided in the Supplementary Materials.

Figure 1 plots the estimated $\hat{\rho}$ against its true values, which ranges 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 $H = 1000$.

Finally, when the researcher erroneously assumes that the true correlation is zero (M1), there is no substantial worsening in the performance of the 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 the MSL estimation

of taste patterns are mostly caused by difficulties in estimating the correlation parameter, with the efficiency of the ρ estimates declining for smaller values of σ_2 .

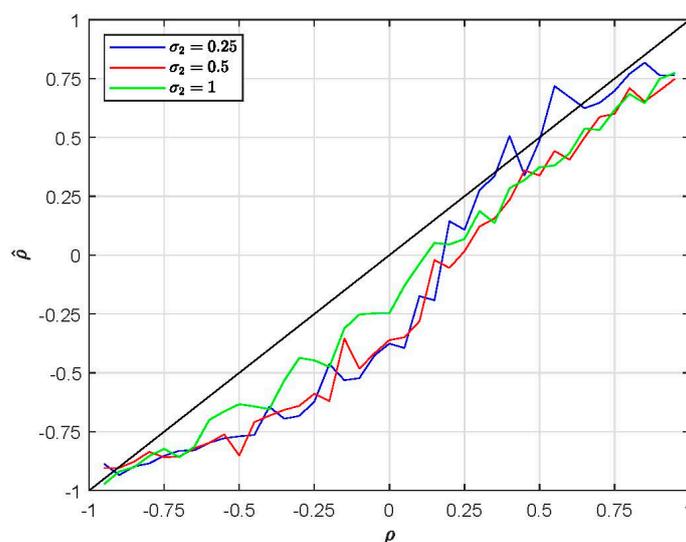


Figure 1. Plots of $\hat{\rho}$ for EC1 (taste patterns) using M0 ($H = 1000$).

4.2. Substitution Patterns: EC2 Simulation Evidence

Table 2 presents the 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 the 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 hypotheses that $H_0 : \alpha_2 = -0.25$ and $H_0 : \beta = 1$. However, when we increase σ_2 to one, the estimated α_1 , α_2 , and β are within 3 standard errors of their respective true values. This conclusion is irrespective of the number of Halton draws.

Second, the standard errors for the estimated σ_2 are substantially larger than those for α_1 , α_2 , and β . As a result, the estimated σ_2 is within 3 standard errors of its true value across almost all the 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 the 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 the correlation is restricted in the specifications M1 and M2.

Third, the correlation parameter ρ is estimated with substantial biases in all the M0 specifications. The 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 $H_0 : \rho = -0.95$ is rejected in all the cases. The EC2 results for the other 38 correlation values are provided in the Supplementary Materials.

Figure 2 plots the 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 that of [Jumamyradov and Munkin \(2021\)](#) for the bivariate normal and bivariate Poisson-lognormal models. They report larger biases for smaller standard deviations. Overall, the M0 results show biases for all five parameters α_1 , α_2 , β , σ_2 , and ρ .

It is also interesting to notice that the 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), and are 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 numbers of Halton draws. Thus, misspecifying the model setting correlation to $\rho = 0$ results in very large

biases in α_1 and α_2 . Moreover, M1 produces larger positive biases for 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), while 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 the M2 estimation, although there is an improvement in the estimation of σ_2 .

Table 2. MSL estimates for 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 (0.012)	-0.271 (0.012)	-0.241 (0.012)	-0.278 (0.011)	-0.246 (0.012)	-0.243 (0.012)	-0.285 (0.012)	-0.266 (0.011)	-0.238 (0.012)
	$\hat{\alpha}_2$	-0.341 (0.011)	-0.299 (0.013)	-0.227 (0.012)	-0.321 (0.014)	-0.289 (0.015)	-0.233 (0.014)	-0.349 (0.014)	-0.301 (0.013)	-0.216 (0.012)
	$\hat{\beta}$	1.160 (0.009)	1.127 (0.009)	1.078 (0.010)	1.165 (0.009)	1.126 (0.010)	1.071 (0.009)	1.146 (0.008)	1.127 (0.008)	1.069 (0.009)
	$\hat{\sigma}_2$	0.258 (0.055)	0.469 (0.050)	0.892 (0.044)	0.284 (0.058)	0.484 (0.060)	0.874 (0.044)	0.252 (0.053)	0.442 (0.050)	0.853 (0.043)
	$\hat{\rho}$	-0.781 (0.040)	-0.810 (0.034)	-0.799 (0.036)	-0.826 (0.037)	-0.740 (0.039)	-0.819 (0.033)	-0.805 (0.034)	-0.768 (0.035)	-0.803 (0.036)
M1	$\hat{\alpha}_1$	-0.243 (0.009)	-0.200 (0.009)	-0.095 (0.009)	-0.247 (0.009)	-0.188 (0.009)	-0.093 (0.010)	-0.255 (0.010)	-0.199 (0.010)	-0.095 (0.009)
	$\hat{\alpha}_2$	-0.285 (0.010)	-0.241 (0.011)	-0.161 (0.013)	-0.292 (0.013)	-0.244 (0.013)	-0.171 (0.016)	-0.312 (0.011)	-0.247 (0.013)	-0.161 (0.013)
	$\hat{\beta}$	1.153 (0.010)	1.111 (0.009)	1.085 (0.011)	1.157 (0.009)	1.117 (0.011)	1.080 (0.012)	1.132 (0.008)	1.113 (0.010)	1.084 (0.011)
	$\hat{\sigma}_2$	0.335 (0.052)	0.534 (0.053)	1.199 (0.058)	0.452 (0.049)	0.660 (0.053)	1.199 (0.060)	0.365 (0.041)	0.558 (0.050)	1.211 (0.055)
M2	$\hat{\alpha}_1$	-0.283 (0.012)	-0.280 (0.012)	-0.278 (0.013)	-0.283 (0.012)	-0.265 (0.013)	-0.281 (0.010)	-0.291 (0.013)	-0.278 (0.011)	-0.278 (0.013)
	$\hat{\alpha}_2$	-0.319 (0.013)	-0.312 (0.015)	-0.308 (0.016)	-0.319 (0.016)	-0.303 (0.017)	-0.319 (0.019)	-0.347 (0.015)	-0.316 (0.016)	-0.308 (0.016)
	$\hat{\beta}$	1.160 (0.010)	1.134 (0.010)	1.123 (0.011)	1.160 (0.010)	1.129 (0.011)	1.117 (0.011)	1.144 (0.009)	1.135 (0.010)	1.123 (0.011)
	$\hat{\sigma}_2$	0.235 (0.048)	0.452 (0.047)	1.027 (0.055)	0.223 (0.052)	0.449 (0.053)	1.058 (0.044)	0.222 (0.045)	0.451 (0.045)	1.028 (0.055)

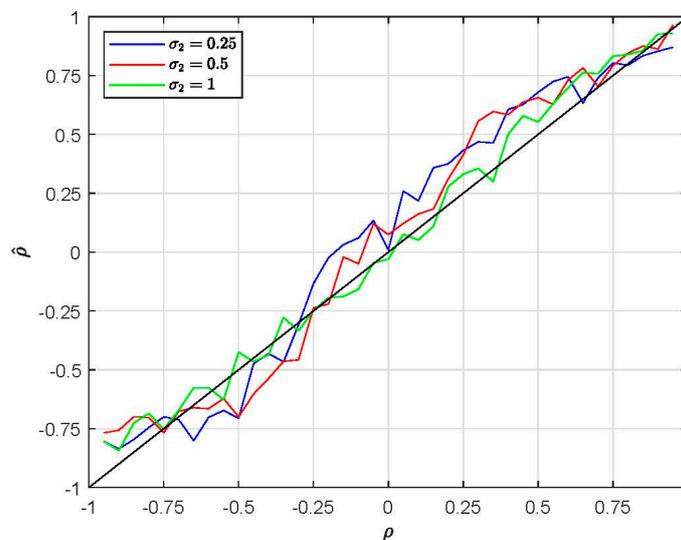


Figure 2. Plots of $\hat{\rho}$ for EC2 (substitution patterns) using M0 ($H = 1000$).

4.3. Choice Probabilities and Marginal Effects

Next, we examine how these reported biases affect the estimated choice probabilities and marginal effects (see Appendix A for the formulas of the marginal effects). Figure 3 plots the true and estimated $P(y = 1)$ calculated based on the M0 estimates of the EC1 specification with 500 Halton draws. The true probability means are calculated for the true values of all the parameters. The straight lines represent the true choice probabilities and the dashed lines represent the estimated choice probabilities. Figure 4 plots the true and estimated $P(y = 1)$ based on the M0 estimates of the 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 the EC1 and EC2 specifications (i.e., taste and substitution patterns).

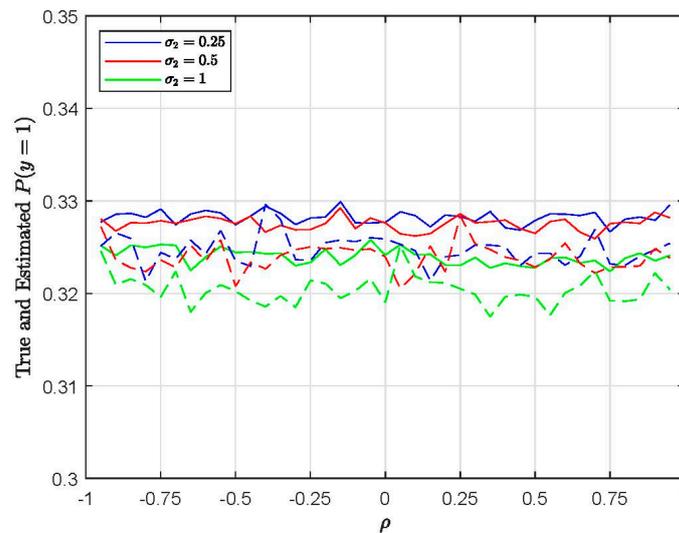


Figure 3. Plots of true and estimated $P(y = 1)$ for EC1 (taste patterns) using M0 ($H = 500$).

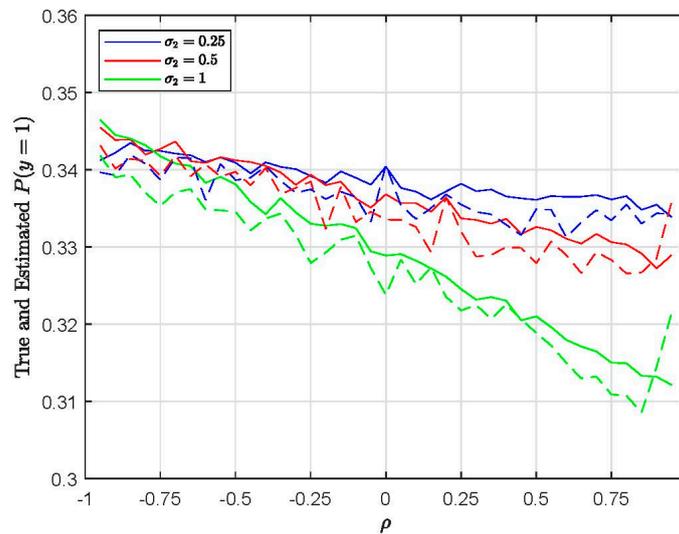


Figure 4. Plots of true and estimated $P(y = 1)$ for EC2 (substitution patterns) using M0 ($H = 500$).

However, when comparing the true and estimated marginal effects, the differences are considerable. Figure 5 plots the 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 the 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%.

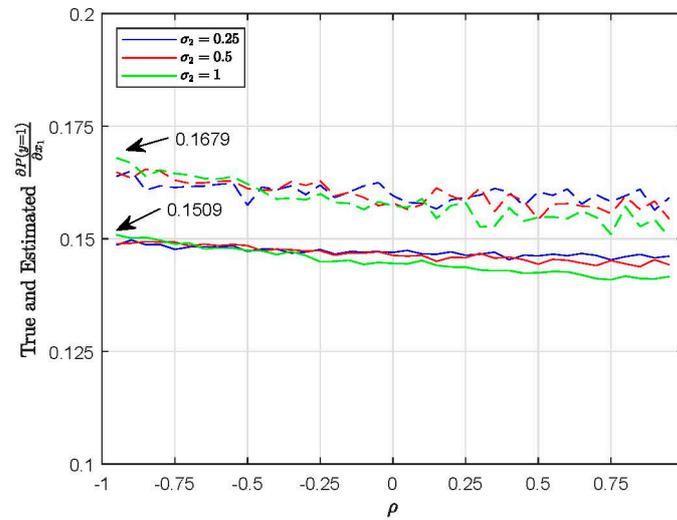


Figure 5. Plots of true and estimated $\frac{\partial P(y=1)}{\partial x_1}$ for EC1 (taste patterns) using M0 (H = 500).

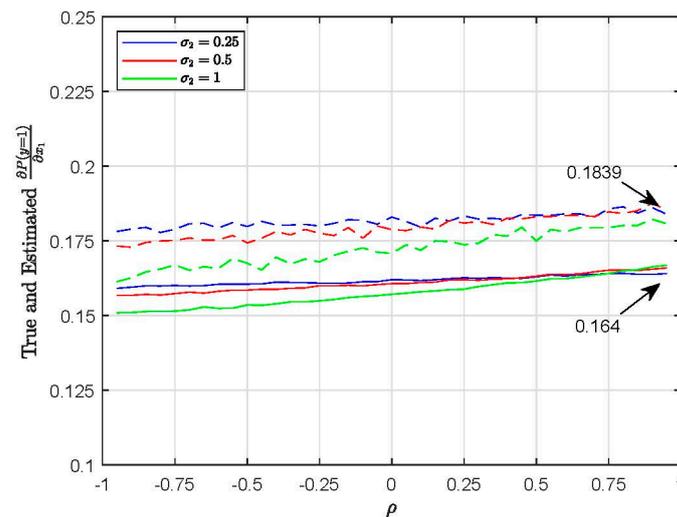


Figure 6. Plots of true and estimated $\frac{\partial P(y=1)}{\partial x_1}$ for EC2 (substitution patterns) using M0 (H = 500).

5. Discussion

In this paper, we examine the 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 biases in the estimated marginal effects can be 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 the performance of the MSL depends on other factors, such as the model specification (i.e., EC1 or EC2), distributional assumptions, exogenous variation, as well as the 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.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/econometrics12020008/s1>. Table S1: Results for all parameter combinations.

Author Contributions: Conceptualization, M.J., M.M. and B.M.C.; methodology, M.J., M.M., W.H.G. and B.M.C.; software, M.J.; validation, M.J., M.M., W.H.G. and B.M.C.; formal analysis, M.J.; investigation, M.J.; data curation, M.J.; writing—original draft preparation, M.J.; writing—review and editing, M.J., M.M., W.H.G. and B.M.C.; visualization, M.J.; supervision, M.M., W.H.G. and B.M.C.; project administration, M.M., W.H.G. and B.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The marginal effects taken with respect to x_1 for EC1 are presented below. The 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 methods

$$\begin{aligned}\hat{\beta}_{i1} &= E\left(\hat{\beta}_{i1}^q \mid y, \mathbf{X}\right) = \frac{\frac{1}{Q} \sum_{q=1}^Q \hat{\beta}_{i1}^q \left(\frac{\exp(\hat{V}_{i1}^q)}{1 + \exp(\hat{V}_{i1}^q) + \exp(\hat{V}_{i2}^q)} \right)}{\frac{1}{Q} \sum_{q=1}^Q \left(\frac{\exp(\hat{V}_{i1}^q)}{1 + \exp(\hat{V}_{i1}^q) + \exp(\hat{V}_{i2}^q)} \right)} \\ \hat{V}_{i1}^q &= \hat{\alpha}_1 + \hat{\beta}_{i1}^q x_{i1} \\ \hat{V}_{i2}^q &= \hat{\alpha}_2 + \hat{\beta}_{i2}^q x_{i2} \\ \hat{\beta}_{i1}^q &= \hat{\beta}_1 + v_{i1}^q \\ \hat{\beta}_{i2}^q &= \hat{\beta}_2 + \hat{\sigma}_2 \hat{\rho} v_{i1}^q + \hat{\sigma}_2 v_{i2}^q \sqrt{1 - \hat{\rho}^2}\end{aligned}$$

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

References

- Bastin, Fabian, and Cinzia Cirillo. 2010. Reducing simulation bias in mixed logit model estimation. *Journal of Choice Modelling* 3: 71–88. [\[CrossRef\]](#)
- Ben-Akiva, Moshe, Denis Bolduc, and Mark Bradley. 1993. Estimation of travel choice models with randomly distributed values of time. *Transportation Research Record* 1413: 88–97.
- Bhat, Chandra. 1998. Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel mode and departure time choice. *Transportation Research Part B: Methodological* 32: 455–66. [\[CrossRef\]](#)
- Bhat, Chandra. 2000. Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science* 34: 228–38. [\[CrossRef\]](#)
- Bhat, Chandra. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological* 35: 677–93. [\[CrossRef\]](#)
- Boyd, Hayden, and Robert Mellman. 1980. The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis. *Transportation Research Part A: General* 14: 367–78. [\[CrossRef\]](#)
- Brownstone, David, and Kenneth Train. 1998. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89: 109–29. [\[CrossRef\]](#)
- Cardell, Scott, and Frederick Dunbar. 1980. Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General* 14: 423–34. [\[CrossRef\]](#)
- Czajkowski, Mikołaj, and Wiktor Budziński. 2019. Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling* 31: 73–85. [\[CrossRef\]](#)
- Debreu, Gerard. 1960. Review of Duncan Luce, Individual Choice Behavior: A theoretical analysis. *American Economic Review* 50: 186–88.
- Elshiewy, Ossama, German Zenetti, and Yasemin Boztug. 2017. Differences between classical and Bayesian estimates for mixed logit models: A replication study. *Journal of Applied Econometrics* 32: 470–76. [\[CrossRef\]](#)
- Fiebig, Denzil, Michael Keane, Jordan Louviere, and Nada Wasi. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29: 393–421. [\[CrossRef\]](#)
- Gourieroux, Christian, and Alain Monfort. 1990. Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique* 20/21: 69–107. [\[CrossRef\]](#)
- Gourieroux, Christian, and Alain Monfort. 1996. *Simulation-Based Econometric Methods*. New York: Oxford University Press.
- Greene, William, and David Hensher. 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* 37: 413–28. [\[CrossRef\]](#)
- Haan, Peter, Daniel Kempton, and Arne Uhlendorff. 2015. Bayesian procedures as a numerical tool for the estimation of an intertemporal discrete choice model. *Empirical Economics* 49: 1123–41. [\[CrossRef\]](#)
- Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud. 1996. Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics* 72: 85–134. [\[CrossRef\]](#)
- Hensher, David, and William Greene. 2003. The mixed logit model: The state of practice. *Transportation* 30: 133–76. [\[CrossRef\]](#)
- Hensher, David, John Rose, and William Greene. 2005. *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press.
- Hess, Stephane, and John Rose. 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation* 39: 1225–39. [\[CrossRef\]](#)
- Hess, Stephane, and Kenneth Train. 2017. Correlation and scale in mixed logit models. *Journal of Choice Modelling* 23: 1–8. [\[CrossRef\]](#)
- Hess, Stephane, Michel Bierlaire, and John Polak. 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice* 39: 221–36. [\[CrossRef\]](#)
- Huber, Joel, and Kenneth Train. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters* 12: 259–69. [\[CrossRef\]](#)
- Jumamyradov, Maksat, and Murat Munkin. 2021. Biases in maximum simulated likelihood estimation of bivariate models. *Journal of Econometric Methods* 11: 55–70. [\[CrossRef\]](#)
- Keane, Michael, and Nada Wasi. 2013. Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics* 28: 1018–45. [\[CrossRef\]](#)
- Lee, Lung-Fei. 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory* 11: 437–83. [\[CrossRef\]](#)
- Lerman, Steven, and Charles Manski. 1980. On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of Discrete Data with Econometric Applications*. Edited by Charles Manski and Daniel McFadden. Cambridge: MIT Press, pp. 305–19.
- Louviere, Jordan, Deborah Street, Leonie Burgess, Nada Wasi, Towhidul Islam, and Anthony Marley. 2008. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1: 128–64. [\[CrossRef\]](#)
- Louviere, Jordan, Deborah Street, Richard Carson, Andrew Ainslie, Jay DeShazo, Trudy Cameron, David Hensher, Robert Kohn, and Tony Marley. 2002. Dissecting the random component of utility. *Marketing Letters* 13: 177–93. [\[CrossRef\]](#)
- Louviere, Jordan, Robert Meyer, David Bunch, Richard Carson, Benedict Dellaert, Michael Hanemann, David Hensher, and Julie Irwin. 1999. Combining sources of preference data for modeling complex decision processes. *Marketing Letters* 10: 205–17. [\[CrossRef\]](#)

- McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*. Edited by Paul Zarembka. New York: Academic Press, pp. 105–42.
- McFadden, Daniel. 2001. Economic choices. *American Economic Review* 91: 351–78. [[CrossRef](#)]
- McFadden, Daniel, and Kenneth Train. 2000. Mixed mnl models for discrete response. *Journal of Applied Econometrics* 15: 447–70. [[CrossRef](#)]
- Palma, Marco, Dmitry Vedenov, and David Bessler. 2020. The order of variables, simulation noise, and accuracy of mixed logit estimates. *Empirical Economics* 58: 2049–83. [[CrossRef](#)]
- Quandt, Richard. 1970. *The Demand for Travel: Theory and Measurement*. Lexington: Heath and Company.
- Regier, Dean A., Mandy Ryan, Euan Phimister, and Carlo A. Marra. 2009. Bayesian and classical estimation of mixed logit: An application to genetic testing. *Journal of Health Economics* 28: 598–610. [[CrossRef](#)] [[PubMed](#)]
- Revelt, David, and Kenneth Train. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics* 80: 647–57. [[CrossRef](#)]
- Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.