The Thesis Committee for Priyadarshan Nandkumar Patil
Certifies that this is the approved version of the following thesis:

Simulation Evaluation of Emerging Estimation Techniques for Multinomial Probit Models

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ____________________________________
Chandra R. Bhat

____
Abdul Rawoof Pinjari

Simulation Evaluation of Emerging Estimation Techniques for Multinomial Probit Models

by

Priyadarshan Nandkumar Patil, B.Tech.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

December 2016

Acknowledgements

I am most grateful to Dr. Chandra Bhat for his tireless efforts, expertise and insights, which have greatly assisted this research study and furthered my understanding of choice modeling. I would also like to express my gratitude to Dr. Abdul Pinjari for helping polish this manuscript and providing constructive comments, helping improve this thesis.

I would like to thank Lisa Macias, for her support and help throughout this study, and her willingness to make life easier for us.

A special thanks to Subodh, Tarun and Sebastian for keeping me motivated and helping me throughout this research study. Their help and knowledge made this journey more enriching. I would also like to thank Udit, Devendra, Rachel, Madhura, Venktesh, Jennifer and all my other friends for making life enjoyable and fun.

Lastly, I would like to thank my parents and family for always being there for me and always offering a helping hand. Their support has been invaluable for me.

Abstract

Simulation Evaluation of Emerging Estimation Techniques for Multinomial Probit Models

Priyadarshan Nandkumar Patil, M.S.E.
The University of Texas at Austin, 2016

Supervisor: Chandra R. Bhat

A simulation evaluation is presented to compare alternative estimation techniques for a five-alternative multinomial probit (MNP) model with random parameters, including cross-sectional and panel datasets and for scenarios with and without correlation among random parameters. The different estimation techniques assessed are: (1) The maximum approximate composite marginal likelihood (MACML) approach; (2) The Geweke-Hajivassiliou-Keane (GHK) simulator with Halton sequences, implemented in conjunction with the composite marginal likelihood (CML) estimation approach; (3) The GHK approach with sparse grid nodes and weights, implemented in conjunction with the composite marginal likelihood (CML) estimation approach; and (4) a Bayesian Markov Chain Monte Carlo (MCMC) approach. In addition, for comparison purposes, the GHK simulator with Halton sequences was implemented in conjunction with the traditional, full information maximum likelihood approach as well. The results indicate that the MACML approach provided the best performance in terms of the accuracy and precision of parameter recovery and estimation time for all data generation settings considered in this

study. For panel data settings, the GHK approach with Halton sequences, when combined with the CML approach, provided better performance than when implemented with the full information maximum likelihood approach, albeit not better than the MACML approach. The sparse grid approach did not perform well in recovering the parameters as the dimension of integration increased, particularly so with the panel datasets. The Bayesian MCMC approach performed well in datasets without correlations among random parameters, but exhibited limitations in datasets with correlated parameters.

Table of Contents

List of Tables

1 INTRODUCTION

Two general econometric model structures have been commonly used in the literature for random utility maximization (RUM)-based discrete choice analysis. These are: (1) the mixed multinomial logit (MMNL) model (McFadden and Train, 2000) and (2) the multinomial probit (MNP) model (Daganzo, 1979).

The MMNL model is typically estimated using the MSL approach, whose desirable asymptotic properties are obtained at the expense of computational cost (because the number of simulation draws has to rise faster than the square root of the number of observations used for estimation). Unfortunately, in situations where the dimensionality of integration is high, such as when spatial/social dependencies are of interest or when considering multi-level (e.g., intra- and inter-individual) unobserved variations in parameters, the computational cost to ensure good asymptotic estimator properties can be prohibitive or, sometimes, simply impractical. Moreover, the MSL estimation and inference can be affected by simulation noise, which might cause problems ranging from non-convergence to inaccuracy and/or non-inversion of the Hessian of the log-likelihood function. Yet, the MSL continues to be the inference approach of choice for MMNL model estimation.

In contrast to the MMNL model, the MNP model has seen relatively little use in the past couple of decades, mainly because its likelihood function involves a truncated multivariate integral (i.e., the cumulative multivariate normal (MVN) function) that is generally more difficult to evaluate using simulation methods compared to the untruncated multivariate integration in the MMNL model. Many studies in the 1990s and earlier on estimating MNP focused on simulation-based estimation, leading up to important advances, including the well known Geweke-Hajivassiliou-Keane (GHK) approach. These studies (for example, Hajivassiliou et al., 1996) demonstrated that the GHK outperformed many other simulation based approaches at that time. As a result, the GHK approach is by far the most commonly used to estimate MNP models. It is worth noting, however, that the GHK is an MSL inference procedure and faces the same problems discussed above in the context of the MSL estimation of the MMNL model. The dimensionality of integration in the MNP model choice probability expressions depends on the number of choice alternatives (and the number of choice occasions per individual in the case of panel data with a general error structure specification). Therefore, the computational cost increases significantly as the number of choice alternatives (or the number of choice occasions per individual) increases. Besides, the GHK simulator is perceived to be relatively more difficult to understand and implement than the MSL simulator for the MMNL (see Train, 2009).

Despite the considerations discussed above, there has been continued interest in MNP for a variety of reasons. The MNP model can indeed be more parsimonious (computationally) than the MMNL in many situations, such as when the number of random coefficients is much more than the number of alternatives (and when the random coefficients are normally distributed). This is because the MNP likelihood function can be expressed as an integral whose dimensionality does not depend on the number of random coefficients in the specification. Besides, in some contexts, the MVN distributional assumption of the MNP may carry better appeal than the extreme value (or multivariate extreme value) distribution used in logit-based models. For example, in social or spatial interaction models, it is much easier to specify parsimonious correlation structures using the MNP kernel than the logit kernel, primarily because of the conjugate nature of the multivariate normal distribution under affine transformations. This is reflected in the almost exclusive use of the MNP kernel for discrete choice models with spatial/social dependence (see a review in Bhat, 2015). Moreover, more recently, there has been a renewed excitement in revisiting the estimation of MNP models using a variety of different methods to approximate or simulate the MVN integrals. Most of these methods can be classified into one of the following three broad categories, each of which is discussed in the following section: (1) Improvements to numerical quadrature methods such as the sparse grid integration (SGI)-based quadrature methods advanced by Heiss and Winschel (2008) and Heiss (2010), (2) Bhat's (2011) maximum approximate composite marginal likelihood (MACML) method, which combines the use of analytic approximations to the MVN integral, as opposed to simulation or numerical evaluation, with a composite marginal likelihood (CML) framework and (3) Advances in Bayesian Markov Chain Monte Carlo (MCMC) methods, particularly those using data augmentation techniques (McCulloch et al., 2000; Imai and van Dyk, 2005).

1.1 The Current Research

Given the increasing interest in MNP models and the emergence of new methods to estimate these models, it is timely to evaluate and compare the performance of different estimation methods available in the literature. Most techniques mentioned above have been compared with traditional frequentist simulation-based approaches, particularly the simulation-based GHK approach (Heiss, 2010; Abay, 2015) or the mixed probit MSL approach (Bhat and Sidharthan, 2011). Some efforts have solely focused on the accuracy of evaluating MVN integrals without examining parameter estimation (Sándor and András, 2004; Connors et al., 2014). To our knowledge, little exists on a comprehensive comparison of the recently emerging methods for MNP estimation in terms of different metrics of importance – the accuracy of parameter recovery, precision of parameter recovery, and the estimation time. The objective of this thesis is to fill this gap. Specifically, the following approaches to estimate MNP models are compared:

(a) The **MACML** approach (as in Bhat, 2011);

(b) The SGI-based quadrature method embedded into the GHK approach, labeled the **GHK-SGI** method, and used in conjunction with the CML estimation approach;

(c) The GHK-simulator using quasi Monte Carlo draws from Halton sequences (Bhat, 2001; Bhat, 2003), labeled the **GHK-Halton** method in the rest of the thesis. This method was used in conjunction with the traditional, full information maximum likelihood (FIML) approach as well as the CML approach; and

(d) The **Bayesian MCMC** approach with data augmentation (as in McCulloch et al., 2000).

To undertake the comparison among these approaches, simulation experiments were conducted with synthetic datasets for a five-alternative MNP model with five random coefficients (in the rest of this thesis, the number of choice alternatives is denoted by I and the number of random coefficients is denoted by K; so I=5 and K=5), for scenarios with and without correlation among random parameters, and for both cross-sectional and panel data settings in an aspatial context. For panel (or repeated choice) data, five choice occasions per individual were simulated. Subsequently, the relative merits of the different estimation techniques for MNP model estimation were evaluated.

In addition to the above discussed methods, Heiss's (2010) modified version of the GHK-SGI method was explored, where he implements the SGI in conjunction with an efficient importance sampling (EIS) technique in the GHK approach. However, our experiments with this approach were not successful, with most estimation attempts encountering convergence problems. After a brief description of this method in Section 2.1.3, the reasons why the method may not have worked in the context of this study are discussed.

A few important caveats here in terms of our study. In our simulation design, independence in the utility kernel error terms across alternatives at each choice occasion is assumed. Technically, in the cross-sectional case, one can then view the model as a mixed multinomial probit (MMNP) model for estimation in which the likelihood function was written as the product of univariate cumulative normal functions integrated over an untruncated $(K+1)$-dimensional (i.e, 6-dimensional) integral space (see Equation (2) in Bhat and Sidharthan, 2011). However, the estimation is easier done using the traditional MNP model basis that involves only a truncated (I-1)-dimensional (i.e, 4-dimensional) integral space (see next section). So, the MNP model basis was used. In the panel case, with a general error structure with individual-specific heterogeneity, choice occasion heterogeneity (or intra-individual variations in taste sensitivity across choice occasions), as well as a general covariance structure across the utilities of alternatives at each choice occasion, the result is an (I-1)*5-dimensional (i.e., 20-dimensional) integral for evaluating MNP probabilities in the likelihood function. In spirit, this general structure as the model basis for evaluating different estimation procedures has been assumed, even though simpler versions of this structure are used (that is, only assume individual-specific heterogeneity in the random coefficients) in the simulation design itself. Technically, in the panel case, assuming only individual-specific heterogeneity simplifies the likelihood function when viewed as an MMNP model for estimation.[1]

[1] This is because the individual likelihood function can be written as the product of univariate cumulative normals integrated over an inside untruncated one-dimensional integral (to obtain the choice occasion-specific probability of the individual), followed by the product of all the choice occasion-specific probabilities across the choice occasions of the individual integrated over an outside untruncated K-dimensional integral space (see Equation (4) in Bhat and Sidharthan, 2011). Obviously, this way of integral evaluation in our simulation setting using the MMNP model basis is much easier to estimate than

For all the frequentist approaches tested on panel data in this thesis (except one exception as discussed at the end of this paragraph), the CML estimation approach within the generic MNP model basis is considered, which reduces the dimensionality of integration by compounding (within each individual) all pairs (or couplets) of choice occasion probabilities. Doing so reduces the dimensionality of the MVNCD function to be evaluated in the CML function to $[2 \times (K-1)]$ dimensions (that is, to an 8-dimensional MVNCD function in the simulation case). That is, all the frequentist approaches (the GHK-Halton, the GHK-SGI, and the MACML) are applied in the panel case using the CML estimation approach rather than the full maximum likelihood estimation approach (for the cross-sectional case, the CML and the full maximum likelihood estimation approaches collapse to being exactly the same).[2] However, to examine the benefit of the CML-based approach for the GHK-Halton simulator (i.e., the GHK-Halton-CML approach), the performance of the traditional GHK-Halton simulator embedded within the full information maximum likelihood (i.e., the GHK-Halton-FIML approach) was evaluated.

---

the 20-dimensional integral in the generic MNP model basis. However, the generic MNP model basis will be used here too as this is the conceptual (and general) basis for this study.

[2] In the CML approach for the panel case, all pairings of the couplet probabilities were considered within an individual (that is, all 10 pairings were considered across the five choice occasions of each individual; see Section 2 for details). However, the CML approach does not need all pairings. A subset of the authors is testing the consequence of using fewer pairings within each individual within the CML context (see Bhat, 2014 for additional details). Doing so can lead to substantial reductions in computation time beyond what is presented here for the MACML and other frequentist approaches.

2 THE MULTINOMIAL PROBIT MODEL

The structure of the panel version of the MNP model is presented here. The cross-sectional model corresponds to a panel with one choice instance per individual. Also, for ease in presentation, a balanced panel is assumed with all alternatives available at each choice instance for each individual.

Let t be the index for choice instance $(t = 1,2,...,T)$, q be the index for individual $(q = 1,2,...,Q)$, and i be the index for choice alternatives $(i = 1,2,...,I)$. Next, write the utility that an individual q derives from choosing alternative i at choice instance t as:

$$U_{qti} = \boldsymbol{\beta}_q' \boldsymbol{x}_{qti} + \xi_{qti}, \tag{1}$$

where $\boldsymbol{x}_{qti}$ is a $(K \times 1)$ vector of exogenous attributes and $\boldsymbol{\beta}_q$ is an individual specific $(K \times 1)$ column vector of coefficients. Assume that $\boldsymbol{\beta}_q$ is distributed multivariate normal with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \mathbf{LL}'$, where $\mathbf{L}$ is a lower-triangular Cholesky factor of $\boldsymbol{\Omega}$. That is, $\boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q$, where $\tilde{\boldsymbol{\beta}}_q \sim \text{MVN}_K(\mathbf{0}_K, \boldsymbol{\Omega})$ where $\mathbf{0}_K$ is a $(K \times 1)$ vector of zeros. In this thesis, no correlation across random coefficients of different individuals was assumed $(\text{Cov}(\boldsymbol{\beta}_q, \boldsymbol{\beta}_{q'}) = \mathbf{0}, \forall q \neq q')$, and no variation in $\boldsymbol{\beta}_q$ across different choice occasions of the same individual was assumed as well. In addition, the ξ_{qti} terms are assumed as IID normally distributed across individuals, alternatives, and choice occasions, with mean zero and variance 0.5.[3]

With the notations as above, the utility expression in Equation (1) can be written as:

$$\begin{aligned} U_{qti} &= (\boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q)' \boldsymbol{x}_{qti} + \xi_{qti} \\ &= \boldsymbol{b}' \boldsymbol{x}_{qti} + (\tilde{\boldsymbol{\beta}}_q' \boldsymbol{x}_{qti} + \xi_{qti}) \\ &= \boldsymbol{b}' \boldsymbol{x}_{qti} + \varepsilon_{qti}, \text{ where } \varepsilon_{qti} = (\tilde{\boldsymbol{\beta}}_q' \boldsymbol{x}_{qti} + \xi_{qti}) \end{aligned} \tag{2}$$

[3] Some of these assumptions may be relaxed to generate a variety of spatial/local dependence or time-varying coefficients (see Bhat, 2014).

Next, define the following vectors and matrices (where $\mathbf{IDEN}_T$ stands for the identity matrix of dimension *T*):

$\boldsymbol{U}_{qt} = (U_{qt1}, U_{qt2}, ..., U_{qtI})'$ [$(I \times 1)$ vector], $\boldsymbol{U}_q = (\boldsymbol{U}'_{q1}, \boldsymbol{U}'_{q2}, ..., \boldsymbol{U}'_{qT})'$ [$(TI \times 1)$ vector],

$\boldsymbol{\xi}_{qt} = (\xi_{qt1}, \xi_{qt1}, ..., \xi_{qtI})'$ [$(I \times 1)$ vector], $\boldsymbol{\xi}_{qt} \sim \text{MVN}_I(\mathbf{0}_I, 0.5 * \mathbf{IDEN}_I)$,

$\boldsymbol{\xi}_q = (\boldsymbol{\xi}'_{q1}, \boldsymbol{\xi}'_{q2}, ..., \boldsymbol{\xi}'_{qT})'$ [$(TI \times 1)$ vector],

$\boldsymbol{\xi}_q \sim \text{MVN}_{TI}(\mathbf{0}_{TI}, 0.5 * \mathbf{IDEN}_{TI})$ where $\mathbf{IDEN}_{TI} = \mathbf{IDEN}_T \otimes \mathbf{IDEN}_I$,

$\boldsymbol{x}_{qt} = (\boldsymbol{x}'_{qt1}, \boldsymbol{x}'_{qt2}, ..., \boldsymbol{x}'_{qtI})'$ [$(I \times K)$ matrix], $\boldsymbol{x}_q = (\boldsymbol{x}'_{q1}, \boldsymbol{x}'_{q2}, ..., \boldsymbol{x}'_{qT})'$ [$(TI \times K)$ matrix], and

$\boldsymbol{\varepsilon}_{qt} = (\varepsilon_{qt1}, \varepsilon_{qt1}, ..., \varepsilon_{qtI})'$ [$(I \times 1)$ vector], $\boldsymbol{\varepsilon}_{qt} \sim \text{MVN}_I(\mathbf{0}_I, \boldsymbol{x}_{qt}\boldsymbol{\Omega}\boldsymbol{x}'_{qt} + 0.5 * \mathbf{IDEN}_I)$, and

$\boldsymbol{\varepsilon}_q = (\boldsymbol{\varepsilon}'_{q1}, \boldsymbol{\varepsilon}'_{q2}, ..., \boldsymbol{\varepsilon}'_{qT})'$ [$(TI \times 1)$ vector].

Equation (2) may now be written in a compact form for all individuals and choice occasions as:

$$\boldsymbol{U}_q = \mathbf{V}_q + \boldsymbol{\varepsilon}_q \text{, where } \mathbf{V}_q = \boldsymbol{x}_q \boldsymbol{b} \text{ and } \boldsymbol{\varepsilon}_q = \boldsymbol{x}_q \tilde{\boldsymbol{\beta}}_q + \boldsymbol{\xi}_q. \tag{3}$$

The distribution of $\boldsymbol{\varepsilon}_q$ may be expressed as: $\boldsymbol{\varepsilon}_q \sim \text{MVN}_{TI}(\mathbf{0}_{TI}, \boldsymbol{\Xi}_q)$ with $\boldsymbol{\Xi}_q = \left[\boldsymbol{x}_q \boldsymbol{\Omega} \boldsymbol{x}'_q + 0.5 * \mathbf{IDEN}_{TI}\right]$ and that of $\boldsymbol{U}_q$ may be expressed as: $\boldsymbol{U}_q \sim \text{MVN}_{TI}(\mathbf{V}_q, \boldsymbol{\Xi}_q)$.

Let individual q be assumed to choose alternative m_{qt} at choice occasion *t*. Let $\boldsymbol{m}_q = (m_{q1}, m_{q2}, ..., m_{qT})'$ [$(T \times 1)$ vector]. To estimate the model, the likelihood was estimated that the utility differences (with respect to the chosen alternative) for all choice occasions are less than zero. To do so, define $\mathbf{M}_q$ as a $\left[T \times (I-1)\right] \times TI$ block diagonal matrix, with each block diagonal being of size $(I-1) \times I$ and containing the matrix $\mathbf{M}_{qt}$. $\mathbf{M}_{qt}$ Itself is constructed as an identity matrix of size $(I-1)$ with an extra column of "-1" values added at the $m_{qt}{}^{\text{th}}$ column. Then the construction of the likelihood expression for individual *q* (i.e. the joint probability of the sequence of choices ($\boldsymbol{m}_q$) made by the individual *q*) is given below:

$$
\begin{aligned}
L_q &= \Pr(\boldsymbol{m}_q) \\
&= \Pr(\mathbf{M}_q \boldsymbol{U}_q < \mathbf{0}_{T(I-1)}) \\
&= \Pr(\mathbf{M}_q(\mathbf{V}_q + \boldsymbol{\varepsilon}_q) < \mathbf{0}_{T(I-1)}) \\
&= \Pr(\boldsymbol{\eta}_q < \mathbf{B}_q) \text{ where } \boldsymbol{\eta}_q = \mathbf{M}_q \boldsymbol{\varepsilon}_q \text{ and } \mathbf{B}_q = -\mathbf{M}_q \mathbf{V}_q
\end{aligned} \tag{4}
$$

That is, $L_q = \int_{\boldsymbol{\eta}_q=-\infty}^{\boldsymbol{B}_q} f(\boldsymbol{\eta}_q)\, d\boldsymbol{\eta}_q$ where $f(\boldsymbol{\eta}_q)$ is the multivariate normal density function for a $G = [(I-1)\times T]$ dimensional normal variate with mean $\mathbf{0}_{T(I-1)}$ and covariance matrix $\boldsymbol{\Lambda}_q = \mathbf{M}_q \boldsymbol{\Xi}_q \mathbf{M}'_q$. To rewrite L_q in terms of the standard multivariate normal distribution, define $\boldsymbol{\omega}_{\Lambda_q}$ as the diagonal matrix of standard deviations of $\boldsymbol{\Lambda}_q$. The vector $\boldsymbol{\omega}_{\Lambda_q}^{-1}\boldsymbol{\eta}_q$ is standard multivariate normally distributed with correlation matrix $\boldsymbol{\Lambda}_q^* = \boldsymbol{\omega}_{\Lambda_q}^{-1} \boldsymbol{\Lambda}_q \boldsymbol{\omega}_{\Lambda_q}^{-1}$. Equation (4) may now be written as:

$$
\begin{aligned}
L_q &= \int_{\boldsymbol{\alpha}_q=-\infty}^{\mathbf{B}_q^*} \phi_G(\mathbf{0}_G; \boldsymbol{\Lambda}_q^*)\, d\boldsymbol{\alpha}_q \\
&= \Phi_G[\mathbf{B}_q^*; \boldsymbol{\Lambda}_q^*]
\end{aligned} \tag{5}
$$

where $\phi_G[.;.]$ and $\Phi_G[.;.]$ are the standard MVN probability density and the MVNCD function of dimension *G*, respectively, and $\mathbf{B}_q^* = \boldsymbol{\omega}_{\Lambda_q}^{-1}\mathbf{B}_q$.

The dimensionality of the integration in Equation (5) is $(I-1)\times T$. Therefore, as the number of choice alternatives or the number of choice occasions per individual increases, the likelihood function becomes computationally expensive or in some cases infeasible to evaluate at a level of accuracy and smoothness needed for parameter estimation using traditional techniques.

A potential solution to reduce the dimensionality of integration is to use the composite marginal likelihood (CML) approach, where the overall likelihood function is calculated as the product of low dimensional marginal densities (see Bhat, 2014). In the current context, the CML function for an individual *q* may be written as a product of the pairwise joint probabilities of the individual's choices over all pairs of choice occasions (Bhat, 2011):

$$L_{q,CML} = \prod_{t=1}^{T-1} \prod_{g=t+1}^{T} \Pr\left(m_{qt}, m_{qg}\right) \quad (6)$$

To further develop the CML function above, define $\breve{\mathbf{B}}_{qtg} = \mathbf{\Delta}_{qtg}\mathbf{B}_q$, $\breve{\mathbf{\Lambda}}_{qtg} = \mathbf{\Delta}_{qtg}\mathbf{\Lambda}_q\mathbf{\Delta}'_{qtg}$, $\breve{\mathbf{B}}^*_{qtg} = \boldsymbol{\omega}^{-1}_{\breve{\Lambda}_{qtg}}\breve{\mathbf{B}}_{qtg}$, and $\breve{\mathbf{\Lambda}}^*_{qtg} = \boldsymbol{\omega}^{-1}_{\breve{\Lambda}_{qtg}}\breve{\mathbf{\Lambda}}_{qtg}\,\boldsymbol{\omega}^{-1}_{\breve{\Lambda}_{qtg}}$, where $\mathbf{\Delta}_{qtg}$ is a $2(I-1)\times T(I-1)$-selection matrix with an identity matrix of size ($I-1$) occupying the first ($I-1$) rows and the $\left[(t-1)\times(I-1)+1\right]^{th}$ through $\left[t\times(I-1)\right]^{th}$ columns, and another identity matrix of size ($I-1$) occupying the last ($I-1$) rows and the $\left[(g-1)\times(I-1)+1\right]^{th}$ through $\left[g\times(I-1)\right]^{th}$ columns. All other elements of $\mathbf{\Delta}_{qtg}$ take a value of zero. Then $L_{q,CML}$ in Equation (6) may be written as:

$$L_{q,CML} = \prod_{t=1}^{T-1} \prod_{g=t+1}^{T} \Phi_{2\times(\text{I-1})}\left[\breve{\mathbf{B}}^*_{qtg}, \breve{\mathbf{\Lambda}}^*_{qtg}\right] \quad (7)$$

Working with the above CML function helps reduce the dimensionality of integration from $(I-1)\times T$ (in the likelihood function of Equation (5)) to $(I\text{ -}1)\times 2$, thereby reducing the model estimation time substantially, and alleviating convergence and parameter recovery problems arising due to large dimensional integrals in the original likelihood function. Of course, if there is only one choice occasion, then the CML expression in Equation (7) collapses to the usual full-information likelihood based estimation approach.

2.1 MNP Estimation Techniques

In this section, the different approaches evaluated in this study for estimating MNP models are discussed.

2.1.1 THE MAXIMUM APPROXIMATE COMPOSITE MARGINAL LIKELIHOOD (MACML) APPROACH

Bhat (2011) proposed the MACML approach that utilizes a CML estimation procedure (Varin, 2008; Bhat, 2014) combined with an analytic approximation to evaluate the MVN cumulative distribution (MVNCD) function in MNP models. The analytic approximation he used is based on the decomposition of the multivariate integral into a product of conditional probabilities that are approximated analytically (see Solow, 1990; and Joe, 1995, though these earlier studies focused on the evaluation of a single MVNCD function, while Bhat proposed an approach to incorporate the analytic approximation in an estimation setting with multiple MVNCD function evaluations). There are at least two advantages of the MACML approach. First, using an analytic expression for the MVNCD function obviates the need for simulation. This also renders the approximated likelihood surface smooth and well behaved for optimization purposes. Second, the CML estimation technique helps in reducing large dimensional integrals (due to panel or repeated-choice data, or spatial/social interactions) into products of lower dimensional integrals. Bhat (2011) has additional detail on the mathematical formulation for this method.

2.1.2 THE GHK-HALTON SIMULATOR

The GHK approach starts with transforming the *correlated* error differences in an MNP model into linear functions of *uncorrelated* standard normal deviates using the Cholesky decomposition of the error difference covariance matrix. Doing so helps in recasting the MVNCD as a recursive product of univariate (conditional) cumulative normal distributions. This simulator is not discussed in any more detail, but refer the reader to Train (2009) for a good exposition of this method. The one difference from the discussion in Train (2009) is that the Halton approach is embedded to recursively simulate draws from the truncated regions (as discussed in detail in Bhat et al., 2010) instead of drawing from pseudo-Monte Carlo sequences in the traditional GHK-simulation approach; therefore, the label GHK-Halton simulator. In addition, as indicated earlier, for panel data settings, the CML estimation approach (using Equation (7)) as well as the full information MSL (FIML) approach (using Equation (5)) in conjunction with the GHK-Halton simulator are considered.

2.1.3 THE GHK APPROACH WITH SPARSE GRID INTEGRATION (GHK-SGI)

Heiss and Winschel (2008) proposed a multivariate quadrature method using the concept of sparse grid integration (SGI) that has been gaining popularity for the evaluation of multidimensional integrals. SGI-based multivariate quadrature is similar to traditional quadrature, except that the multivariate node points at which the integrand is evaluated are chosen cleverly and sparsely (based on a tensor product rule from Smolyak, 1963) to avoid the curse of dimensionality from operating on a full grid of all combinations of nodes in all dimensions. Heiss and Winschel (2008) describe this approach in detail and demonstrate the effectiveness of the approach in evaluating multidimensional integrals of up to 20 dimensions for MMNL (not MNP) parameter estimation. In the current chapter on MNP parameter estimation, the GHK-SGI approach is employed, where the SGI nodes and weights are used within the GHK framework, instead of drawing from Psuedo-Monte Carlo or Quasi-Monte Carlo sequences as in traditional GHK-simulation (see Abay, 2015 for a similar setup for multivariate binary probit models). Further, as indicated earlier, the CML estimation in conjunction with the GHK-SGI approach was used.

2.1.4 THE GHK SIMULATOR WITH EFFICIENT IMPORTANCE SAMPLING AND SPARSE GRIDS

The performance of quasi-random (e.g., Halton) sequence-based GHK simulation may be enhanced through the use of Efficient Importance sampling (EIS), a variance-reduction technique based on the idea that a certain set of draws from a given sequence contribute more toward the approximation of the integral than other draws from the same sequence. If one can sample such '*important*' values more frequently, the approximation will be quicker and more accurate. Hence, the key to importance sampling is to choose an auxiliary distribution (also called the importance sampler) which facilities easy sampling of important draws along with reducing the sampling variance (i.e., distance between the importance sampler and the initial sampler). In this context, Heiss (2010) proposes the use of a normally distributed importance sampler inside the GHK simulator along with a weighted least squares technique proposed by Richard and Zhang (2007) to minimize the sampling variance. Heiss (2010) provides Monte Carlo evidence that the resulting GHK-EIS-SGI approach offers better (and less computationally intensive) parameter recovery than the simulation-based GHK procedure in the context of panel binary probit models.

A potential problem with the use of sparse grids in conjunction with importance sampling is that a significant percentage of sparse grid nodes might be associated with negative weights. And using negative weights within weighted least squares technique to minimize the variance (i.e., minimizing an objective function using negative weights) might lead to undue importance to the negative weights causing convergence issues during parameter estimation. This is a reason why our experiments to estimate MNP models with Heiss's (2010) GHK-EIS-SGI approach were not successful, with most estimation attempts encountering convergence issues. Two *adhoc* solutions were attempted to address this problem: (1) neglect all the SGI nodes with negative weight during the minimization of sampling variance, and (2) replace negative SGI weights by their absolute values. Neither of these approaches appears to guarantee a smooth estimation, as found in the current study. Thus, the GHK-EIS-SGI approach was dropped in further investigations in this study.

2.1.5 THE BAYESIAN MCMC APPROACH

Advances in the Bayesian domain have led to efficient MCMC methods for estimating MNP models, particularly using the data augmentation technique. Application of the Bayesian method of estimation to MNP consists of a Markov chain Monte Carlo (MCMC) approximation of the posterior distribution of the model parameters. The basic idea is to augment the parameter space so that simulated realizations of the random utility are generated. Therefore, data augmentation in this case implies that the dependent variable (the utility function) becomes observable, making it possible to use standard Bayesian regression techniques for estimating both the population parameters and the random taste variations. Both McCulloch et al. (2000) and Imai and van Dyk (2005) follow this approach. However, the McCulloch et al. (2000) method incorporates the normalization for utility scale after the estimation is done. Imai and van Dyk's (2005) method improved on this by considering the normalization for utility scale explicitly at the beginning of estimation (similar to the frequentist approach). In the current study, the performance of the McCulloch et al. (2000) method is evaluated and the Imai and van Dyk (2005) approach is left for future research.

3 DESIGN OF THE SIMULATION EXPERIMENT

MNP models with five alternatives and five independent variables (and five random coefficients, one on each of the independent variables) are considered for the following four data generation processes: (a) Cross-sectional data without correlation among random parameters, (b) Cross-sectional data with correlation among random parameters, (c) Panel data without correlation among random parameters, and (d) Panel data with correlation among random parameters.

Consider a "true" or underlying probit model according to the utility function: $U_{qti} = \boldsymbol{\beta}'_q \boldsymbol{x}_{qti} + \xi_{qti}$ as in Equation (1). For all the datasets generated in this study, the values of each of the five independent variables $\boldsymbol{x}_{qti}$ for the alternatives are drawn from a standard normal distribution. Random coefficients are allowed on all the five independent variables in $\boldsymbol{x}_{qti}$. That is, $\boldsymbol{\beta}_q$ is a vector of normally distributed coefficients with mean $\boldsymbol{b} = \{1.5,\text{-}1.0,2.0,1.0,\text{-}2.0\}$ and covariance matrix $\boldsymbol{\Omega}$. For the case of uncorrelated random parameters, a diagonal covariance matrix is assumed with all the diagonal elements set to a value of 1, entailing the estimation of five diagonal elements (albeit all are of value 1). For the case of correlated random parameters, the matrix $\boldsymbol{\Omega}$ has the following positive definite non-diagonal specification with five diagonal elements and five non-zero off-diagonal elements, entailing the estimation of fifteen covariance matrix parameters:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1.00 & -0.50 & 0.25 & 0.75 & 0.00 \\ -0.50 & 1.00 & 0.25 & -0.50 & 0.00 \\ 0.25 & 0.25 & 1.00 & 0.33 & 0.00 \\ 0.75 & -0.50 & 0.33 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

Finally, each kernel error term ξ_{qti} (q = 1, 2, …, Q; t = 1, 2, 3, ..., T; i = 1, 2, …, I) is generated from a univariate normal distribution with a variance of 0.5.

For the cross-sectional datasets, a sample of 2500 realizations is generated of the five independent variables corresponding to 2500 individuals. For the panel datasets, a sample of 2500 realizations was generated of the five independent variables corresponding to a situation where 500 individuals each have five choice occasions for a total of 2500 choice occasions. These are combined with different realizations of $\boldsymbol{\beta}_q$

and ξ_{qti} terms to compute the utility functions as in Equation (1) for all individuals and choice occasions. Next, for each individual and choice occasion, the alternative with the highest utility for each observation is identified as the chosen alternative. This data generation process is undertaken 200 times with different realizations of the vector of coefficients $\boldsymbol{\beta}_q$ and error term ξ_{qti} to generate 200 different datasets for each of the four variants of MNP.

For each of the above 200 datasets, the MNP was estimated using each of the estimation approaches. For the MACML approach, a single random permutation was used as discussed in Section 2.1.1. For estimating models with the GHK-Halton approach, 500 Halton draws were used. For the GHK-SGI approach, to keep the estimation times similar to the GHK-Halton approach, the MVN integral was computed over 351 and 589 supports points for cross-sectional and panel cases, respectively. For the Bayesian approach, 50,000 MCMC draws were used for all four cases with a burn-in of the first 500 elements of the chain.[4] Finally, for all the three frequentist methods, standard errors of parameter estimates (for each dataset) were computed using the Godambe (1960) sandwich estimator ($H^{-1}JH^{-1}$, where H is the Hessian matrix and J is the sandwich matrix). The Hessian and sandwich matrices were computed at the convergent parameters using analytic expressions that a. For the MCMC method, the standard errors of the parameter estimates (for each dataset) were calculated as the standard deviation of the parameter's posterior distribution at convergence.

To measure the performance of each estimation method, performance metrics were computed as described below.

[4] The issue of the number of iterations in the simulation chain prior to convergence to the joint posterior distribution of parameters (that is, the "burn-in") has received quite a bit of attention in the Bayesian estimation literature, with no clear consensus regarding the number of iterations that should be considered as "burn-in". While some studies (see, for example, Johndrow et al., 2013, Burgette and Reiter, 2013, Wang et al., 2014) use a specific number (such as 1,000 or 3,000 or 10,000) for burn-in iterations, others (see, for example, Zhang et al., 2008) use a specific percentage to arrive at this number (such as 1% or 10% of the total number of iterations of the sampler used in the estimation). Some studies (see, for example, Gelman and Shirley, 2011) even question the use of burn-in iterations. While this "burn-in" issue is not addressed in the current study, we varied the burn-in from 500 to 1000 to 10,000 for a select sample of estimation runs across different data generation cases, and found little impact on the metrics used to assess accuracy and precision of parameter recovery.

(a) For each parameter, compute the mean of its estimates across the 200 datasets to obtain a **mean estimate**. Compute the **absolute percentage (finite sample) bias (APB)** of the estimator as:

$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100.$$

If a true parameter value is zero, the APB is computed by taking the difference of the mean estimate from the true value (= 0), dividing this difference by the value of 1 in the denominator, and multiplying by 100.

(b) For each parameter, compute the standard deviation of the parameter estimate across the 200 datasets, and label this as the **finite sample standard error or FSSE** (essentially, this is the empirical standard error or an estimate of the standard deviation in finite samples). For the Bayesian MCMC method, the FSSEs are calculated as the standard deviation of the mean of the posterior estimates across different datasets.

(c) For each parameter, compute the standard error of the estimate using the Godambe sandwich estimator. Then compute the mean of the standard error across the 200 datasets, and label this as **the asymptotic standard error or ASE** (this is the standard error of the distribution of the estimator as the sample size gets large). For the Bayesian MCMC method, the ASEs are computed as the standard deviation of parameter's chain at the end of the convergence and then averaged across the 200 datasets.

(d) For each parameter, compute the **square root of mean squared error (RMSE)** as

$$RMSE = \sqrt{(Mean\ Estimate - True\ Value)^2 + FSSE^2}$$

(e) For each parameter, compute the **coverage probability (CP)** as below:

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

where, CP is the coverage probability, $\hat{\beta}_X^r$ is the estimated value of the parameter in dataset r, β_X is the true value of the parameter, $\text{se}(\hat{\beta}_X^r)$ is the asymptotic standard error (ASE) of the parameter in the dataset r, $I[.]$ is an indicator function which takes a value of 1 if the argument in the bracket is true (otherwise 0), N is the number of datasets (200), and t_α is the t-statistic value for a given confidence level $(1-\alpha)\times 100.$ CP values were computed for 80% nominal coverage probability (i.e., $\alpha = 0.20$). CP is the empirical probability that a confidence interval contains the true parameter (i.e., the proportion of confidence intervals across the 200 datasets that contain the true parameter). CP values smaller than the nominal confidence level (80% in our study) suggest that the confidence intervals do not provide sufficient empirical coverage of the true parameter.

(f) Store the **run time** for estimation, separately for convergence of the parameter estimates and for calculation of the ASE values necessary for inference.

4 PERFORMANCE EVALUATION RESULTS

Table 1 presents an overall summary of the performance of all the estimation approaches considered in this study – MACML, GHK-Halton, GHK-SGI, and MCMC – for all four cases of the MNP data generation process – cross-sectional uncorrelated, cross-sectional correlated, panel uncorrelated, and panel correlated.[5] Note that two different columns are reported for the GHK-Halton method for panel data settings. One of them corresponds to the traditional GHK-Halton-FIML approach and the other corresponds to the GHK-Halton-CML approach.

For each of the estimation methods and data settings, the first block of rows in Table 1 presents the average APB value (across all parameters), as well as the average APB value computed separately for the mean (the $\boldsymbol{b}$ vector) parameters and the covariance matrix (the $\boldsymbol{\Omega}$ matrix) elements. The second and third blocks provide the corresponding information for the FSSE and ASE measures. The fourth block provides the RMSE and CP measures for all model parameters, and the final block provides the average model estimation run times across all 200 datasets, split by the time for convergence to the final set of parameters and the time needed for ASE computation in the frequentist methods. Several key observations from this table are discussed in the next few sections.

[5] The detailed results for all the cases are available in the appendix at the end of the thesis as well as an online appendix at: http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/SimEval/Appendix.pdf.

Table 1: Overall summary of the simulation results

	Cross-sectional data, uncorrelated random coefficients				Cross-sectional data, correlated random coefficients				Panel data, uncorrelated random coefficients					Panel data, correlated random coefficients				
	MACML	**GHK-Halton**	**GHK-SGI**	**MCMC**	**MACML**	**GHK-Halton**	**GHK-SGI**	**MCMC**	**MACML**	**GHK-Halton FIML**	**GHK-Halton CML**	**GHK-SGI CML**	**MCMC**	**MACML**	**GHK–Halton FIML**	**GHK-Halton CML**	**GHK-SGI CML**	**MCMC**
Absolute Percentage Bias (APB)																		
All parameters	2.64	3.89	3.05	3.45	3.16	4.25	3.72	16.43	3.08	7.53	4.49	28.15	8.23	3.62	8.43	6.14	33.09	20.42
Mean parameters	0.65	1.25	2.27	4.23	0.71	0.85	0.8	1.12	1.63	3.42	2.13	22.35	6.48	2.23	3.86	2.89	24.39	3.45
Covariance parameters	3.30	4.77	3.31	3.19	3.98	5.38	4.69	21.53	3.56	8.90	5.25	30.08	8.81	4.08	9.95	7.22	35.99	26.08
Finite Sample Standard Error (FSSE)																		
All parameters	0.33	0.46	0.42	0.24	0.30	0.34	0.28	0.33	0.26	0.25	0.23	0.16	0.18	0.18	0.21	0.21	0.12	0.21
Mean parameters	0.22	0.28	0.26	0.25	0.28	0.32	0.26	0.35	0.19	0.21	0.19	0.14	0.17	0.19	0.22	0.22	0.13	0.23
Covariance parameters	0.37	0.52	0.47	0.24	0.31	0.35	0.29	0.32	0.28	0.26	0.24	0.16	0.18	0.18	0.20	0.20	0.12	0.2
Asymptotic Standard Error (ASE)																		
All parameters	0.33	0.46	0.44	0.23	0.25	0.36	0.27	0.28	0.22	0.25	0.23	0.16	0.17	0.17	0.22	0.20	0.16	0.19
Mean parameters	0.24	0.30	0.28	0.23	0.27	0.31	0.28	0.31	0.16	0.20	0.19	0.15	0.16	0.20	0.19	0.21	0.17	0.18
Covariance parameters	0.36	0.51	0.49	0.23	0.24	0.38	0.26	0.27	0.24	0.27	0.24	0.16	0.17	0.16	0.23	0.19	0.15	0.19
RMSE and Coverage Probability (CP)																		
RMSE	0.345	0.466	0.429	0.360	0.309	0.429	0.373	0.541	0.291	0.384	0.318	0.560	0.386	0.316	0.395	0.343	0.631	0.507
$CP_{80\%}$	92.12	89.95	90.47	90.14	88.28	86.49	86.12	57.25	80.42	63.73	74.01	52.46	59.23	75.86	58.17	71.53	49.55	55.46
Computation Time (minutes)																		
Convergence time	0.72	1.06	0.86	5.31	2.07	2.79	2.66	6.44	5.88	13.42	6.75	14.48	7.02	8.73	24.53	9.76	27.36	8.97
ASE computation time	0.31	0.33	0.31	--	0.41	0.38	0.36	--	12.98	17.23	13.85	18.29	--	14.36	20.45	17.53	22.01	--
Total runtime	1.03	1.39	1.17	5.31	2.58	3.17	3.02	6.44	18.86	30.65	20.60	32.77	7.02	23.09	44.98	27.29	49.37	8.97

4.1 Accuracy of Parameter Recovery

The APB measures in the first block of Table 1 provide several important insights. The MACML approach outperforms other inference approaches for all the four cases of data generation. This underscores the superiority of the MACML approach in accurately recovering model parameters. In all inference approaches, the overall APB increases as we move from the cross-sectional to panel case, and from the uncorrelated to the correlated case. But, even here, the MACML shows the least APB dispersion among the many data generation cases, while the MCMC and GHK-SGI approaches show the highest dispersions among the data generation cases. The most striking observation is the rapid degradation of the MCMC approach between the uncorrelated and correlated random coefficients cases, for both cross-sectional and panel data sets. The MCMC has the worst APB of all inference approaches (and by a substantial margin) in the correlated random coefficients setting in the cross-sectional case, and the second worst APB in the correlated random coefficients in the panel case. In terms of the performance of the GHK-SGI approach, the most striking observation is the substantially poor performance of the GHK-SGI approach (in combination with the CML approach) in the panel cases relative to the performance of the GHK-SGI approach in the cross-sectional cases.[6]

---

[6] This significant drop in the GHK-SGI performance from the cross-sectional to panel case may be attributed to one or more of three different factors: (1) due to an increase in the dimension of integration (recall that the panel models using the CML approach in this study involve 8-dimensional integrals, while the cross-sectional models involve 4-dimensional integrals), (2) due to the change in the nature of the dataset (cross-sectional to panel), and (3) due to any potential difficulty of using SGI approach in conjunction with the CML method. To disentangle these effects, additional simulation experiments were conducted with the GHK-SGI method. Specifically, models were estimated on simulated data for cross-sectional MNP with uncorrelated random parameters for seven choice alternatives (dimension of integration equals 6) and 9 choice alternatives (dimension of integration equals 8), respectively, with the same simulation configuration as discussed earlier. The overall APB values for the 6 and 8 dimensional integration cases (with the new cross-sectional data) were 6.59 and 12.30, respectively (and the overall APB for the 4 dimensional uncorrelated cross-sectional case is 3.05; see Table 1). These results indicate that the ability of the GHK-SGI method to recover true parameters degrades quickly after 4 or 5 dimensions (another recent study by Abay, 2015 confirms this trend). It is worth noting, however, that the panel data model integrals of 8-dimensions (as in Table 1) show a much poorer performance (APB values are around 30%) compared to cross-sectional data models of the same dimension. This could be due to evaluation of a greater number of 8 dimensional integrals in the panel datasets estimated using CML approach. That is, for a cross-sectional dataset with 2500 observations and 9 alternatives, a total of twenty-five hundred 8-dimensional integrals were evaluated, while for a panel dataset

The APB values from the GHK-Halton approach for the cross-sectional cases are higher than (but in the same order of APB) as the other two frequentist (MACML and GHK-SGI) approaches for the cross-sectional cases. For the panel cases, as already discussed, an FIML version (labeled as GHK Halton-FIML in Table 1) as well as a CML version (labeled as GHK Halton-CML) of the GHK-Halton approach is implemented. Both these GHK Halton versions provide an APB that is higher than the MACML approach, but are superior to the GHK-SGI and MCMC approaches in terms of recovering parameters accurately. Between the FIML and CML versions of this GHK-Halton approach, the latter approach recovers the parameters more accurately; the APB for the GHK-Halton-FIML simulator is 30-50% higher than the GHK-Halton CML simulator. This is a manifestation of the degradation of simulation techniques to evaluate the MVNCD function as the number of dimensions of integration increases. The results clearly show the advantage of combining the traditional GHK simulator with the CML inference technique for panel data, although the MACML approach still dominates over the GHK-Halton CML approach.

The split of the APB by the mean and covariance parameters follow the overall APB trends rather closely. Not surprisingly, except for the MCMC approach with uncorrelated cross-sectional data, it is more difficult to recover the covariance parameters accurately relative to the mean parameters. For the frequentist methods, this is a reflection of the appearance of the covariance parameters in a much more complex non-linear fashion than the mean parameters in the likelihood function, leading to a relatively flat log-likelihood function for different covariance parameter values and more difficulty in accurately recovering these parameters. But the most noticeable observation from the mean and covariance APB values is the difference between these for the MCMC method with correlated random coefficients. In fact, it becomes clear now that the substantially higher overall APB for the MCMC approach (relative to the MACML and GHK-Halton

---

with 500 observations with 5 choice occasions, a total of five thousand 8-dimensional integrals were evaluated. Therefore, it appears that the performance of the SGI method degrades quickly with the dimensionality of integration as well as with the number of integrals evaluated (in this case the number of 8-dimensional integrals doubled due to the CML approach). However, further research is required to fully disentangle the impact of the nature of the dataset and dimension of integration on the performance of the SGI method.

approaches) for the correlated random coefficients case is primarily driven by the poor MCMC ability to recover the covariance parameters, suggesting increasing difficulty in drawing efficiently from the joint posterior distribution of parameters (through a sequence of conditioning mechanisms) when there is covariance in the parameters.

In summary, from the perspective of recovering parameters accurately, the MACML outperforms other approaches for all the four data generation cases. The GHK-Halton also does reasonably well across the board, with the GHK-Halton-CML doing better than the GHK-Halton-FIML for the panel cases. The GHK-SGI is marginally better than the GHK-Halton for the cross-sectional cases, but, when combined with the CML approach, is the worst in the panel cases. The MCMC approach's ability to recover parameters is in the same range as the approaches involving the GHK-Halton for the uncorrelated random coefficients cases, but deteriorates substantially in the presence of correlated random coefficients (note also that 50,000 iterations are used in the MCMC approach in the current study, more than the 5,000-15,000 iterations typically used in earlier MCMC estimations of the MNP; see, for example, Chib et al., 1998; Johndrow et al., 2013; Jiao and van Dyk, 2015).

4.2 Precision in Estimation Across Approaches

The standard errors are considered next.. The FSSE values are useful for assessing the empirical (finite-sample) efficiency (or precision) of the different estimators, while the ASE values provide efficiency results as the sample size gets very large. The ASE values essentially provide an approximation to the FSSE values for finite samples. Table 1 indicates that the MCMC estimator has the advantage of good efficiency (lowest FSSE and ASE) for the cross-sectional, uncorrelated random coefficients case, but the GHK-SGI wins the finite-sample efficiency battle (lowest FSSE) for all the remaining three cases.[7] In

[7] While the sampling distribution (whose standard deviation is represented by FSSEs) is not a Bayesian concept, one may invoke the Bernstein–von Mises Theorem (see Train, 2009, pp. 288) that the posterior distribution of each parameter converges to a normal distribution with the same variance as that of the maximum likelihood estimator (frequentist estimator, to be more inclusive) to use the FSSE values for assessing the empirical efficiency of the MCMC estimator.

terms of ASE, the MACML has the lowest value for the cross-sectional correlated case, while the GHK-SGI has the lowest value for the panel cases. Such a high precision in the estimates in the GHK-SGI, however, is not of much use because of the rather high finite sample bias (APB) in the parameter estimates of the GHK-SGI approach. In all cases, the MACML does very well too in terms of closeness to the approach with the lowest FSSE and ASE. Of particular note is that the MACML estimator's efficiency in terms of both FSSE and ASE is better than the traditional frequentist GHK-Halton simulator for all cases, except in the panel data-uncorrelated random coefficients case. Additionally, the MACML estimator's efficiency, while not as good as that of the MCMC in the two uncorrelated coefficients cases, is better than the MCMC for the two correlated coefficients cases.

For all the frequentist methods, the FSSE and ASE values across all parameters are smaller in the presence of correlation among random parameters than without correlation. As can be observed from the third rows of the tables under FSSE and ASE in Table 1, this pattern is driven by the smaller FSSE and ASE values for the covariance parameters in the correlated case relative to the non-correlated case. As discussed in Bhat et al. (2010), it may be easier to retrieve covariance parameters with greater precision at higher values of covariance because, at lower correlation values, the likelihood surface tends to be flat, increasing the variability in parameter estimation. This trend, however, reverses for the MCMC method, with the FSSE and ASE values being higher in data settings with correlated random parameters than those with non-correlated random parameters, presumably for the same reason that the APB values in the MCMC method are very high in the correlated coefficients case relative to the uncorrelated coefficients case. Across all inference approaches, a consistent result is that the FSSE and ASE are smaller for the mean parameters than the covariance parameters. Also, the closeness of the FSSE and ASE values for the frequentist approaches suggest that the inverse of the Godambe sandwich estimator serves as a good approximation to the finite sample efficiency for the sample size considered in this study. The FSSE and ASE are also close for the MCMC approach.

Overall, in terms of estimator efficiency, it appears that all inference approaches do reasonably well. There are also some more general takeaways from the examination of the

FSSE and ASE values. First, while the full-information maximum likelihood approach is theoretically supposed to be more asymptotically efficient than the limited-information composite marginal likelihood approach (see a proof for this in Bhat, 2015), this result does not necessarily extend to the case when there is no clear analytically tractable expression for the probabilities of choice in a discrete choice model. This is illustrated in the FSSE/ASE estimates from the GHK Halton-FIML and GHK Halton-CML approaches for panel data in Table 1, with the latter proving to be a more efficient estimator than the former. At a fundamental level, when any kind of an approximation is needed (either through simulation methods or analytically) for the choice probabilities, the efficiency results will also depend on how accurately the objective function (the log-likelihood function in FIML and the composite log-likelihood in CML) can be evaluated. The CML approach has lower dimensional integrals, which can be evaluated more accurately than the higher dimensional integrals in the FIML approach, and this can lead to a more efficient CML estimator (as is the case in Table 1). Second, the MACML estimator's efficiency is consistently better than that of the GHK-Halton based simulator for the range of data settings considered in this study. In combination with the superior performance of the MACML in terms of parameter recovery, this lends reinforcement to our claim that accuracy of evaluating the objective function (as a function of the parameters to be estimated) does play a role in determining estimator efficiency. Third, while Bayesian estimators are typically invoked on the grounds of good small sample inference properties in terms of higher efficiency relative to frequentist estimators in finite samples, our results indicate that, at least for the sample size considered in this study, this all depends on the context and is certainly not a foregone conclusion empirically. For instance, while the MCMC approach leads to lower FSSE/ASE values than the MACML approach for the uncorrelated coefficients cases, the MACML leads to lower FSSE/ASE values than the MCMC approach for the correlated coefficients cases.

4.3 Root Mean Squared Error (RMSE) and Coverage Probability (CP)

The RMSE measure combines the bias and efficiency considerations into a single metric, as discussed in Section 3. The results indicate that the MACML approach has the lowest RMSE values for all the four data generation cases. The GHK-SGI approach is the next best for the cross-sectional cases, but is the worst (and by a substantial margin) for the panel cases. The MCMC approach and the GHK-Halton approach are comparable to each other in the cross-sectional uncorrelated coefficients (first) case in Table 1, and both of these are also comparable to the performance of the GHK-SGI approach in this first case. For the panel uncorrelated coefficients (third) case, the MCMC has an RMSE value comparable to the FIML version of the GHK-Halton, but fares clearly worse than the CML version of the GHK-Halton. Of course, for both the correlated coefficients cases (second and fourth cases), the MCMC is not a contender at all based on our analysis.

The coverage probability (CP) values help assess how the parameter estimates spread about the true parameter value. As one may observe from Table 1, all approaches provide good empirical coverage of the 80% nominal confidence interval in the cross-sectional uncorrelated case (all the values are above 80%). The MCMC falls short in the cross-sectional correlated random coefficients case. For the panel cases, the MACML and the GHK-Halton CML approaches are the only two that cover or come very close to covering the 80% confidence interval, with the MACML clearly providing better coverage than the GHK-Halton CML. These results are generally in line with the RMSE value trends.

Overall, based on the RMSE and CP values, the MACML approach is the clear winner across all data generation cases. In terms of stability in performance across all cases, the GHK-Halton turns out to be the second best inference approach in our results (when used in combination with the CML approach in the two cases of panel data).

4.4 Computation Time

The last block of Table 1 provides model estimation times (or run times) for different estimation methods explored in this study. The total run time for the frequentist methods include both the time taken for parameter convergence as well as for the computation of asymptotic standard errors (ASEs) using the Godambe sandwich estimator. The run times reported for the MCMC approach does not include ASE computation; instead, it involves a simple standard deviation of the posterior distribution. The computer configuration used to conduct these tests is: Intel Xeon® CPU E5-1620 @3.70GHz, Windows 7 Enterprise (64 bit), 16.0 GB RAM. Also, all the estimations were performed using codes written in the Gauss matrix programming suite to ensure comparability.

The results in Table 1 shows that the convergence times for the MACML approach is the lowest for the cross-sectional datasets, with the time for ASE computation being about half of the convergence time for the uncorrelated random coefficients case and a fifth of the convergence time for the correlated random coefficients case. The other two frequentist approaches take about the same time as the MACML. However, the time for the Bayesian MCMC approach is substantially higher in the cross-sectional cases (about five times the MACML estimation time for the uncorrelated coefficients case and 2.5 times the MACML estimation time for the correlated coefficients case). As also observed by Train (2009), little change was found in the MCMC estimation time between the uncorrelated and correlated coefficients cases.

For the panel cases, the MACML is the fastest approach in terms of convergence time, though the GHK-Halton implemented with the CML approach has a comparable convergence time. The other two frequentist approaches (GHK-Halton with FIML and the GHK-SGI CML) have a much higher convergence time relative to the MACML and GHK-Halton CML approaches. The MCMC convergence times are in the same range as the MACML and GHK-Halton-CML. However, the MCMC has the advantage that the ASE estimates of parameters are obtained directly from the posterior distribution of parameters at convergence. For the frequentist methods, however, the ASE computation involves the computation of the inverse of the Godambe information matrix, which itself involves the computation of the Hessian matrix that is time consuming (the ASE computation time for the MACML approach, for example, is about twice the time needed for parameter convergence in the two panel cases). When the ASE computation time is added in for the frequentist methods, the MCMC has a speed advantage by a factor of about 2.5 relative to the MACML. The problem, though, is that the MCMC fares much more poorly compared to the MACML (and GHK-Halton CML) approaches for panel data in terms of parameter recovery accuracy and precision, as evident in the RMSE and CP measures.

5 SUMMARY AND CONCLUSIONS

Multinomial Probit (MNP) models are gaining increasing interest for choice model estimation in transportation and other fields. This study has presented an extensive simulation experiment to evaluate different estimation techniques for MNP models. While one cannot make conclusive statements applicable for all possible data generation settings in terms of number of choice alternatives, correlation structures, and sample sizes, the simulations undertaken do provide some key insights that could be used to guide MNP estimation.

Overall, taking all the three metrics (accuracy and precision of parameter recovery and estimation time) into consideration, the MACML approach provided the best performance for the data generation settings examined in this study. These results indicate the promise of this approach for estimating MNP models in different settings. The GHK-Halton simulation, when used in conjunction with the CML approach (for panel models), yielded the second best performance in recovering the parameters. On the other hand, the bias in parameter estimation was more than double that of the MACML approach when the GHK-Halton simulator was used in its original FIML form for panel data models. In fact, the GHK-Halton when combined with the FIML estimator for panel data sets was also less efficient than the GHK-Halton in conjunction with the CML estimator, highlighting the fact that the FIML estimator's theoretical efficiency superiority over the CML estimator may not get manifested in empirical samples when the objective function to be maximized is analytically intractable. In such cases, the accuracy of evaluating the objective function is also important. In the current study, the CML involves lower-dimensional integrals than the FIML, and the ability to evaluate the lower dimensional integrals more accurately leads to more precision in the CML estimator relative to the FIML estimator. These results highlight the potential for gainful applicability of the CML approach with the traditional GHK simulator.

The GHK-based sparse grid integration approach performed well in the cross-sectional cases, but very poorly for panel datasets when combined with the CML approach. These results suggest that the approach may not be applicable for settings with higher than

5 dimensional integrals or panel data settings (see Abay, 2015 for a similar finding). The MCMC approach performed very well for the cross-sectional data without correlation in the parameters and appears to be a good alternative approach to use for such a data setting. But, even in this case, the MACML approach dominates in terms of accuracy and precision of parameter recovery, as well as has a speed advantage by a factor of about five relative to the MCMC approach. Our simulations also indicate a notable limitation of the MCMC approach in recovering MNP parameters in cases where the random coefficients are correlated (both in the cross-sectional and panel settings). This finding needs to be further investigated to examine ways to improve the MCMC method in the presence of correlated random coefficients.

The results in this study are encouraging in that the emerging methods – MACML, CML, and MCMC – are making the estimation of MNP models easier and faster than before. But there is a need for continued simulation experimentation with these alternative methods to provide more general guidance under a wider variety of data settings, including different numbers of alternatives, different sample sizes, different numbers of repeated choice occasions in the panel case, a range of correlation structures across coefficients and choice occasions, and different numbers of exogenous variables and types of exogenous variables (including discrete and binary variables). Also, there are a variety of potential ways to improve upon the MACML and MCMC approaches in particular, such as alternative analytic approximations for the MVNCD function (see Trinh and Genz, 2015) in the MACML, and reducing MACML computation time by sampling pairings for an individual rather than using the full set of pairings as done here. Finally, future research needs to investigate ways to improve the MCMC performance in correlated random coefficients cases and consider Imai and van Dyk's (2005) method of scaling utilities at the beginning of the estimation.

Appendix

Table 2: Evaluation of the ability to recover true parameters for the cross-sectional diagonal case

Parameter	True Value	MACML Method					GHK-MSL Method					GHK-SGI Method					MCMC Method				
		Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error
Mean values of the β vector																					
b 1	1.500	1.510	0.68%	0.202	0.183	0.209	1.518	1.23%	0.238	0.221	0.235	1.535	2.32%	0.186	0.174	0.196	1.442	3.89%	0.315	0.354	0.164
b 2	-1.000	-0.993	0.68%	0.274	0.252	0.264	-1.013	1.33%	0.369	0.346	0.347	-0.977	2.32%	0.334	0.299	0.308	-0.957	4.29%	0.175	0.189	0.258
b 3	2.000	1.988	0.60%	0.321	0.299	0.327	1.976	1.21%	0.304	0.290	0.307	1.956	2.18%	0.265	0.250	0.278	2.093	4.65%	0.213	0.229	0.233
b 4	1.000	1.006	0.61%	0.155	0.147	0.160	0.987	1.31%	0.346	0.306	0.309	1.021	2.11%	0.330	0.318	0.326	0.955	4.48%	0.143	0.148	0.274
b 5	-2.000	-2.014	0.71%	0.238	0.225	0.257	-1.977	1.16%	0.262	0.241	0.263	-2.048	2.41%	0.287	0.265	0.293	-1.923	3.85%	0.295	0.306	0.246
Covariance matrix of the β vector																					
Ω_{11}	1.000	1.028	2.81%	0.343	0.337	0.350	0.953	4.67%	0.536	0.578	0.574	0.962	3.77%	0.485	0.450	0.459	1.032	3.22%	0.146	0.147	0.385
Ω_{12}	0.000	-0.030	3.04%	0.457	0.458	0.467	0.053	5.34%	0.579	0.599	0.593	0.034	3.41%	0.524	0.498	0.504	-0.027	2.71%	0.180	0.188	0.423
Ω_{13}	0.000	0.040	3.96%	0.281	0.296	0.302	0.046	4.58%	0.480	0.496	0.491	0.034	3.44%	0.457	0.419	0.424	-0.038	3.83%	0.212	0.222	0.356
Ω_{14}	0.000	-0.039	3.86%	0.405	0.397	0.405	0.046	4.58%	0.530	0.525	0.520	-0.028	2.85%	0.528	0.506	0.512	0.029	2.90%	0.303	0.317	0.430
Ω_{15}	0.000	0.030	3.00%	0.279	0.299	0.305	-0.050	5.01%	0.529	0.563	0.558	0.028	2.85%	0.487	0.452	0.457	0.034	3.38%	0.240	0.252	0.384
Ω_{22}	1.000	0.971	2.94%	0.417	0.427	0.440	1.047	4.72%	0.484	0.494	0.492	1.035	3.48%	0.510	0.504	0.513	0.965	3.53%	0.189	0.198	0.430
Ω_{23}	0.000	-0.039	3.86%	0.448	0.460	0.469	0.050	5.01%	0.460	0.477	0.472	0.028	2.78%	0.544	0.526	0.532	0.026	2.55%	0.300	0.313	0.446
Ω_{24}	0.000	0.031	3.14%	0.318	0.329	0.336	0.054	5.42%	0.494	0.526	0.520	-0.037	3.71%	0.421	0.408	0.412	-0.035	3.48%	0.289	0.306	0.346
Ω_{25}	0.000	0.031	3.10%	0.333	0.327	0.334	0.044	4.39%	0.411	0.435	0.430	0.031	3.08%	0.449	0.442	0.448	0.036	3.63%	0.210	0.212	0.376
Ω_{33}	1.000	0.962	3.80%	0.319	0.341	0.352	1.041	4.10%	0.465	0.491	0.490	1.032	3.21%	0.442	0.400	0.409	1.037	3.67%	0.161	0.158	0.344
Ω_{34}	0.000	0.029	2.90%	0.283	0.281	0.287	0.040	4.01%	0.598	0.600	0.594	-0.034	3.38%	0.572	0.531	0.537	-0.027	2.68%	0.184	0.186	0.451
Ω_{35}	0.000	0.031	3.10%	0.369	0.362	0.370	-0.049	4.87%	0.449	0.456	0.452	0.037	3.67%	0.523	0.515	0.521	0.034	3.45%	0.320	0.323	0.438
Ω_{44}	1.000	1.026	2.64%	0.328	0.334	0.346	0.955	4.53%	0.521	0.529	0.526	0.968	3.24%	0.565	0.558	0.568	0.974	2.58%	0.203	0.208	0.476
Ω_{45}	0.000	-0.033	3.27%	0.455	0.446	0.456	0.053	5.29%	0.563	0.554	0.549	-0.034	3.41%	0.490	0.463	0.469	0.027	2.71%	0.258	0.273	0.393
Ω_{55}	1.000	1.039	3.86%	0.424	0.453	0.467	1.051	5.09%	0.581	0.600	0.597	0.966	3.38%	0.411	0.406	0.414	0.965	3.54%	0.306	0.297	0.348
Overall Mean Value Across Parameters		-	2.64%	0.330	0.333	0.345	-	3.89%	0.458	0.460	0.466	-	3.05%	0.438	0.418	0.429	-	3.45%	0.232	0.238	0.364
Mean Time		1.03					1.39					1.17					5.31				
Std. dev of Time		0.14					0.27					0.19					1.06				
% of Runs Converged		100%					100%					100%					100%				

Table 3: Evaluation of the ability to recover true parameters for the cross-sectional non-diagonal case

Parameter	True Value	MACML Method					GHK-MSL Method					GHK-SGI Method					MCMC Method				
		Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error
Mean values of the β vector																					
b 1	1.500	1.511	0.76%	0.346	0.369	0.378	1.513	0.88%	0.301	0.306	0.390	1.487	0.84%	0.364	0.328	0.439	1.482	1.18%	0.373	0.419	0.696
b 2	-1.000	-1.008	0.77%	0.236	0.234	0.240	-0.991	0.87%	0.311	0.318	0.397	-1.009	0.86%	0.218	0.209	0.280	-0.989	1.10%	0.259	0.291	0.482
b 3	2.000	1.986	0.71%	0.288	0.301	0.321	2.018	0.88%	0.334	0.353	0.457	1.984	0.81%	0.296	0.280	0.390	2.022	1.08%	0.222	0.243	0.438
b 4	1.000	0.993	0.66%	0.190	0.188	0.195	0.992	0.77%	0.302	0.323	0.403	0.992	0.77%	0.229	0.209	0.279	0.989	1.12%	0.391	0.442	0.726
b 5	-2.000	-2.013	0.67%	0.310	0.337	0.357	-1.983	0.86%	0.321	0.333	0.435	-2.015	0.74%	0.306	0.288	0.399	-2.023	1.13%	0.320	0.363	0.620
Covariance matrix of the β vector																					
Ω11	1.000	1.043	4.34%	0.286	0.394	0.398	0.937	6.29%	0.401	0.356	0.443	1.046	4.64%	0.282	0.299	0.394	1.207	20.67%	0.308	0.353	0.585
Ω12	-0.500	-0.479	4.30%	0.192	0.267	0.268	-0.474	5.22%	0.480	0.445	0.549	-0.478	4.49%	0.319	0.360	0.466	-0.612	22.45%	0.263	0.304	0.498
Ω13	0.250	0.259	3.70%	0.153	0.203	0.203	0.262	4.99%	0.356	0.337	0.415	0.236	5.49%	0.209	0.224	0.289	0.308	23.31%	0.194	0.242	0.396
Ω14	0.750	0.780	3.94%	0.220	0.292	0.295	0.791	5.49%	0.318	0.297	0.370	0.778	3.75%	0.181	0.207	0.274	0.587	21.75%	0.267	0.336	0.550
Ω15	0.000	0.041	4.14%	0.305	0.420	0.419	0.059	5.86%	0.379	0.330	0.406	0.055	5.45%	0.299	0.318	0.411	0.224	22.39%	0.180	0.219	0.358
Ω22	1.000	0.957	4.34%	0.247	0.318	0.323	1.047	4.67%	0.364	0.335	0.419	1.040	4.03%	0.348	0.392	0.512	1.197	19.73%	0.231	0.283	0.474
Ω23	0.250	0.241	3.54%	0.318	0.426	0.425	0.264	5.49%	0.314	0.291	0.358	0.238	4.61%	0.186	0.202	0.261	0.311	24.54%	0.277	0.349	0.569
Ω24	-0.500	-0.477	4.66%	0.304	0.389	0.388	-0.522	4.41%	0.368	0.320	0.396	-0.522	4.46%	0.162	0.174	0.229	-0.601	20.24%	0.292	0.342	0.559
Ω25	0.000	0.043	4.26%	0.254	0.355	0.355	0.049	4.90%	0.345	0.308	0.380	0.054	5.39%	0.275	0.294	0.380	0.209	20.88%	0.273	0.330	0.537
Ω33	1.000	1.035	3.46%	0.159	0.208	0.216	1.051	5.11%	0.360	0.341	0.426	1.055	5.53%	0.222	0.236	0.315	1.197	19.73%	0.266	0.316	0.527
Ω34	0.330	0.317	3.82%	0.319	0.437	0.437	0.347	5.00%	0.363	0.327	0.404	0.315	4.55%	0.336	0.377	0.488	0.400	21.31%	0.261	0.307	0.501
Ω35	0.000	0.037	3.66%	0.176	0.228	0.227	0.061	6.08%	0.476	0.418	0.515	0.049	4.92%	0.218	0.245	0.316	0.213	21.31%	0.289	0.358	0.583
Ω44	1.000	1.041	4.14%	0.159	0.201	0.209	1.053	5.27%	0.434	0.385	0.480	0.952	4.78%	0.278	0.319	0.418	0.770	23.04%	0.354	0.422	0.691
Ω45	0.000	-0.037	3.70%	0.260	0.333	0.332	0.058	5.76%	0.466	0.434	0.534	-0.042	4.17%	0.320	0.354	0.457	0.172	17.22%	0.280	0.328	0.534
Ω55	1.000	1.037	3.66%	0.145	0.186	0.195	1.062	6.19%	0.352	0.321	0.403	1.041	4.13%	0.305	0.351	0.460	0.757	24.33%	0.250	0.302	0.497
Overall Mean Value Across Parameters		-	3.16%	0.251	0.303	0.309	-	4.25%	0.358	0.341	0.429	-	3.72%	0.274	0.285	0.373	-	16.43%	0.279	0.328	0.541
Mean Time		2.58					3.17					3.02					6.44				
Std. dev of Time		0.38					0.62					0.45					1.22				
% of Runs Converged		100%					100%					100%					100%				

Table 4: Evaluation of the ability to recover true parameters for the panel diagonal case

Parameter	True Value	MACML Method					GHK-MSL Method					GHK-CML Method					GHK-SGI Method					MCMC Method				
		Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error
Mean values of the β vector																										
b 1	1.500	1.474	1.76%	0.123	0.145	0.182	1.448	3.47%	0.267	0.280	0.443	1.471	2.02%	0.207	0.199	0.294	1.841	22.76%	0.180	0.165	0.643	1.402	6.52%	0.188	0.199	0.441
b 2	-1.000	-1.016	1.60%	0.121	0.146	0.173	-0.962	3.79%	0.210	0.223	0.347	-1.020	2.02%	0.211	0.195	0.278	-0.756	24.35%	0.246	0.229	0.784	-1.063	6.33%	0.175	0.192	0.418
b 3	2.000	1.970	1.52%	0.246	0.297	0.352	2.066	3.29%	0.125	0.136	0.267	1.957	2.22%	0.131	0.123	0.222	2.428	21.39%	0.065	0.059	0.468	2.129	6.45%	0.147	0.155	0.410
b 4	1.000	0.982	1.76%	0.191	0.225	0.256	1.033	3.33%	0.200	0.205	0.324	0.977	2.38%	0.159	0.151	0.219	0.779	22.08%	0.090	0.084	0.318	1.065	6.45%	0.174	0.182	0.396
b 5	-2.000	-2.030	1.52%	0.120	0.137	0.192	-2.064	3.22%	0.197	0.206	0.357	-2.039	2.00%	0.292	0.281	0.414	-2.423	21.17%	0.168	0.163	0.707	-2.133	6.64%	0.115	0.122	0.360
Covariance matrix of the β vector																										
Ω11	1.000	1.030	2.98%	0.249	0.277	0.313	0.902	9.82%	0.279	0.278	0.428	1.045	4.51%	0.263	0.268	0.373	1.318	31.82%	0.116	0.112	0.458	0.925	7.53%	0.216	0.239	0.506
Ω12	0.000	0.041	4.15%	0.141	0.165	0.182	-0.072	7.17%	0.339	0.314	0.474	0.045	4.51%	0.229	0.219	0.298	0.290	28.99%	0.210	0.208	0.701	0.108	10.79%	0.229	0.249	0.516
Ω13	0.000	-0.030	3.05%	0.206	0.251	0.277	0.072	7.17%	0.187	0.178	0.270	0.057	5.78%	0.202	0.197	0.268	-0.306	30.56%	0.100	0.099	0.339	-0.088	8.80%	0.225	0.241	0.498
Ω14	0.000	-0.041	4.11%	0.172	0.196	0.217	-0.074	7.43%	0.227	0.208	0.314	0.052	5.25%	0.143	0.149	0.202	0.299	29.93%	0.249	0.262	0.882	0.093	9.25%	0.239	0.257	0.531
Ω15	0.000	0.034	3.44%	0.312	0.352	0.388	-0.101	10.09%	0.236	0.237	0.359	0.043	4.35%	0.266	0.261	0.354	0.280	28.04%	0.188	0.187	0.633	-0.082	8.16%	0.091	0.096	0.199
Ω22	1.000	1.035	3.55%	0.301	0.336	0.376	1.102	10.17%	0.358	0.352	0.540	1.057	5.73%	0.275	0.270	0.375	1.268	26.78%	0.170	0.167	0.607	1.073	7.35%	0.095	0.096	0.235
Ω23	0.000	0.032	3.19%	0.321	0.389	0.429	-0.092	9.20%	0.209	0.209	0.316	0.053	5.36%	0.289	0.303	0.411	0.258	25.84%	0.161	0.168	0.566	0.095	9.52%	0.276	0.288	0.595
Ω24	0.000	0.030	3.01%	0.216	0.242	0.267	0.092	9.20%	0.281	0.258	0.390	0.059	5.94%	0.190	0.197	0.268	0.353	35.29%	0.113	0.113	0.385	0.091	9.07%	0.121	0.129	0.267
Ω25	0.000	-0.043	4.25%	0.189	0.216	0.238	0.072	7.17%	0.363	0.353	0.533	0.050	5.04%	0.293	0.292	0.396	-0.293	29.30%	0.071	0.070	0.242	-0.082	8.16%	0.112	0.116	0.240
Ω33	1.000	1.036	3.62%	0.298	0.357	0.399	1.093	9.29%	0.272	0.268	0.417	1.047	4.72%	0.277	0.280	0.389	0.745	25.52%	0.257	0.249	0.848	0.907	9.34%	0.106	0.114	0.257
Ω34	0.000	-0.034	3.37%	0.222	0.267	0.295	0.074	7.43%	0.233	0.226	0.341	0.062	6.31%	0.243	0.250	0.339	0.340	34.03%	0.118	0.121	0.414	-0.080	7.98%	0.087	0.090	0.186
Ω35	0.000	0.039	3.90%	0.296	0.350	0.386	0.098	9.82%	0.210	0.195	0.295	0.048	4.83%	0.247	0.240	0.326	0.268	26.78%	0.077	0.080	0.274	0.084	8.44%	0.194	0.199	0.411
Ω44	1.000	1.038	3.79%	0.245	0.295	0.332	1.103	10.26%	0.209	0.201	0.319	0.947	5.36%	0.202	0.195	0.274	0.622	37.81%	0.130	0.127	0.445	1.095	9.52%	0.167	0.172	0.380
Ω45	0.000	-0.028	2.84%	0.186	0.210	0.232	-0.094	9.38%	0.329	0.328	0.495	-0.063	6.36%	0.223	0.230	0.312	-0.325	32.45%	0.205	0.202	0.682	0.085	8.53%	0.197	0.199	0.411
Ω55	1.000	1.041	4.15%	0.246	0.297	0.334	0.901	9.91%	0.316	0.295	0.453	1.047	4.72%	0.258	0.250	0.350	0.720	28.04%	0.236	0.235	0.803	1.097	9.70%	0.196	0.216	0.463
Overall Mean Value Across Parameters		-	3.08%	0.221	0.256	0.291	-	7.53%	0.252	0.253	0.384	-	7.53%	0.229	0.234	0.318	-	28.15%	0.159	0.164	0.560	-	8.23%	0.170	0.181	0.386
Mean Time		18.86					30.65					20.6					32.77					7.02				
Std. dev of Time		2.95					5.54					3.58					5.08					1.13				
% of Runs Converged		100%					100%					100%					100%					100%				

Table 5: Evaluation of the ability to recover true parameters for the panel non-diagonal case

Parameter	True Value	MACML Method					GHK-MSL Method					GHK-CML Method					GHK-SGI Method					MCMC Method				
		Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE	Parameter Estimates		Standard Error Estimates		RMSE
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error	Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Finite Sample Standard Error	Root Mean Square Error
Mean values of the β vector																										
b 1	1.500	1.467	2.21%	0.307	0.287	0.495	1.554	3.63%	0.171	0.196	0.403	1.545	2.85%	0.156	0.160	0.294	1.874	24.96%	0.145	0.111	0.710	1.558	3.86%	0.225	0.284	0.700
b 2	-1.000	-1.022	2.21%	0.168	0.158	0.279	-0.959	4.13%	0.188	0.220	0.421	-0.968	3.02%	0.134	0.148	0.253	-1.245	24.48%	0.157	0.120	0.643	-0.968	3.23%	0.167	0.225	0.544
b 3	2.000	2.042	2.10%	0.225	0.218	0.412	1.925	3.75%	0.113	0.139	0.331	2.061	2.93%	0.233	0.254	0.450	2.433	21.66%	0.174	0.132	0.877	1.932	3.40%	0.087	0.108	0.358
b 4	1.000	0.979	2.12%	0.183	0.178	0.307	0.960	4.01%	0.219	0.248	0.472	0.972	2.69%	0.242	0.249	0.411	0.746	25.43%	0.176	0.137	0.663	1.036	3.61%	0.223	0.281	0.677
b 5	-2.000	-1.950	2.51%	0.116	0.110	0.250	-1.924	3.78%	0.260	0.297	0.591	-2.062	2.96%	0.285	0.290	0.510	-1.491	25.43%	0.197	0.150	0.789	-2.063	3.16%	0.200	0.252	0.661
Covariance matrix of the β vector																										
$\Omega 11$	1.000	0.954	4.59%	0.095	0.112	0.207	1.101	10.05%	0.291	0.255	0.489	0.936	6.87%	0.132	0.139	0.241	1.420	42.02%	0.098	0.081	0.517	1.290	29.03%	0.154	0.166	0.422
$\Omega 12$	-0.500	-0.521	4.18%	0.115	0.130	0.220	-0.448	10.33%	0.163	0.142	0.267	-0.530	6.40%	0.197	0.213	0.348	-0.694	38.84%	0.130	0.103	0.510	-0.627	25.37%	0.226	0.244	0.579
$\Omega 13$	0.250	0.241	3.58%	0.198	0.224	0.371	0.223	10.61%	0.260	0.230	0.428	0.230	8.51%	0.093	0.100	0.164	0.144	42.37%	0.179	0.141	0.654	0.198	20.92%	0.219	0.227	0.534
$\Omega 14$	0.750	0.775	3.30%	0.201	0.235	0.397	0.808	7.80%	0.271	0.231	0.438	0.801	7.42%	0.243	0.251	0.411	0.498	33.54%	0.147	0.116	0.551	0.950	26.67%	0.091	0.097	0.259
$\Omega 15$	0.000	0.043	4.35%	0.186	0.199	0.331	0.112	11.18%	0.249	0.212	0.394	0.071	7.73%	0.238	0.241	0.389	0.300	30.01%	0.218	0.180	0.837	-0.282	28.24%	0.178	0.190	0.448
$\Omega 22$	1.000	1.037	3.70%	0.219	0.253	0.430	1.095	9.49%	0.226	0.197	0.383	0.915	9.21%	0.258	0.281	0.463	1.417	41.66%	0.220	0.178	0.905	1.272	27.20%	0.148	0.158	0.406
$\Omega 23$	0.250	0.262	4.75%	0.156	0.171	0.284	0.276	10.24%	0.301	0.266	0.495	0.230	8.51%	0.173	0.177	0.287	0.179	28.25%	0.138	0.109	0.503	0.304	21.44%	0.235	0.255	0.601
$\Omega 24$	-0.500	-0.522	4.43%	0.222	0.243	0.406	-0.450	9.96%	0.189	0.163	0.306	-0.529	6.32%	0.167	0.172	0.282	-0.673	34.60%	0.074	0.060	0.329	-0.361	27.72%	0.170	0.183	0.432
$\Omega 25$	0.000	0.047	4.75%	0.207	0.226	0.374	-0.092	9.20%	0.303	0.273	0.507	0.069	7.42%	0.222	0.231	0.373	0.342	34.25%	0.253	0.199	0.923	0.209	20.92%	0.132	0.136	0.321
$\Omega 33$	1.000	1.034	3.38%	0.115	0.134	0.239	0.896	10.43%	0.207	0.173	0.335	1.068	7.34%	0.107	0.112	0.206	0.682	31.78%	0.070	0.054	0.302	1.248	24.84%	0.283	0.285	0.690
$\Omega 34$	0.333	0.344	4.35%	0.118	0.129	0.217	0.359	8.92%	0.164	0.143	0.269	0.305	8.12%	0.205	0.209	0.339	0.197	40.25%	0.135	0.108	0.500	0.235	28.76%	0.265	0.264	0.622
$\Omega 35$	0.000	0.041	4.14%	0.203	0.231	0.383	0.094	9.39%	0.160	0.135	0.251	0.080	8.67%	0.175	0.193	0.312	0.339	33.90%	0.113	0.093	0.439	0.267	26.67%	0.148	0.162	0.383
$\Omega 44$	1.000	0.958	4.18%	0.122	0.136	0.243	0.908	9.20%	0.250	0.223	0.424	0.931	7.50%	0.189	0.200	0.335	1.304	30.36%	0.113	0.092	0.541	0.718	28.24%	0.197	0.212	0.506
$\Omega 45$	0.000	-0.036	3.62%	0.091	0.107	0.177	0.113	11.27%	0.280	0.239	0.444	-0.060	6.48%	0.261	0.282	0.456	0.385	38.49%	0.179	0.143	0.669	-0.241	24.06%	0.193	0.196	0.460
$\Omega 55$	1.000	0.961	3.90%	0.153	0.172	0.299	1.112	11.18%	0.136	0.119	0.252	1.086	9.29%	0.191	0.199	0.338	1.395	39.54%	0.183	0.142	0.756	0.689	31.12%	0.211	0.225	0.538
Overall Mean Value Across Parameters		-	3.62%	0.168	0.184	0.316	-	8.34%	0.222	0.214	0.395	-	6.14%	0.199	0.211	0.343	-	33.09%	0.157	0.125	0.631	-	20.42%	0.193	0.206	0.507
Mean Time		23.09					44.98					27.29					49.37					8.97				
Std. dev of Time		3.58					7.82					4.55					8.37					1.79				
% of Runs Converged		100%					100%					100%					100%					100%				

References

Abay, K. A. (2015). Evaluating simulation-based approaches and multivariate quadrature on sparse grids in estimating multivariate binary probit models. *Economics Letters*, 126, 51-56.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35(7), 677-693.

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, 37(9), 837-855.

Bhat, C.R., (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45 (7), 923-939.

Bhat, C.R. (2014). The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundation and Trends in Econometrics*, 7(1), 1-117.

Bhat, C.R. (2015). A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. *Transportation*, 42(5), 879-914.

Bhat, C. R., and Sidharthan, R. (2011). A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, 45(7), 940-953.

Bhat, C. R., Varin, C., and Ferdous, N. (2010). A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. In Greene, W.H. and Hill, R.C. (Eds.) *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Vol. 26, 65-106, Emerald Group Publishing Limited.

Burgette, L. F., and Reiter, J. P. (2013). Multiple-shrinkage multinomial probit models with applications to simulating geographies in public use data. *Bayesian Analysis*, 8(2), 453-478.

Chib, S., Greenberg, E., and Chen, Y. (1998). MCMC methods for fitting and comparing multinomial response models. *Economics Working Paper Archive, Econometrics, 9802001*.

Connors, R. D., Hess, S., and Daly, A. (2014). Analytic approximations for computing probit choice probabilities. *Transportmetrica A: Transport Science*, 10(2), 119-139.

Daganzo, C. (1979). *Multinomial Probit: the Theory and its Application to Demand Forecasting*. Academic Press, New York.

Gelman, A., and Shirley, K. (2011). Inference from simulations and monitoring convergence. In Brooks, S. P., Gelman, A., Jones, G., and Meng, X-L. (Eds.) *Handbook of Markov Chain Monte Carlo*, Ch. 6, 163-174. Chapman Hall, Boca Raton, FL.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 1208-1211.

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

Heiss, F. (2010). The panel probit model: Adaptive integration on sparse grids. In Greene, W.H. and Hill, R.C. (Eds.) *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Vol. 26, 41-64, Emerald Group Publishing Limited.

Heiss, F., and Winschel, V. (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1), 62-80.

Imai, K., and van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124(2), 311-334.

Jiao, X., and van Dyk, D. A. (2015). A corrected and more efficient suite of MCMC samplers for the multinomal probit model. *arXiv preprint arXiv:1504.07823*.

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90(431), 957-964.

Johndrow, J., Dunson, D. B., and Lum, K. (2013). Diagonal orthant multinomial probit models. Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS), Scottsdale, AZ, Journal of Machine Learning Research (JMLR) Workshop and Conference Proceedings, Vol. 31, 29-38.

McCulloch, R. R., Polson, N. G., and Rossi, P. E. (2000). Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99, 173-193.

McFadden, D., and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.

Richard, J. F., and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2), 1385-1411.

Sándor, Z., and András, P. (2004). Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics*, 120(2), 207-234.

Solow, A. R. (1990). A method for approximating multivariate normal orthant probabilities. *Journal of Statistical Computation and Simulation*, 37(3-4), 225-229.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics Doklady*, 4, 240-243.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, Cambridge.

Trinh, G., and Genz, A.(2015). Bivariate conditioning approximations for multivariate normal probabilities. *Statistics and Computing*, 25, 989-996.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1), 1-28.

Wang, Y., Kockelman, K. M., and Damien, P. (2014). A spatial autoregressive multinomial probit model for anticipating land-use change in Austin, Texas. *The Annals of Regional Science*, 52(1), 251-278.

Zhang, X., Boscardin, W. J., and Belin, T. R. (2008). Bayesian analysis of multivariate nominal measures using multivariate multinomial probit models. *Computational Statistics & Data Analysis*, 52(7), 3697-3708.