The Composite Marginal Likelihood (CML) Estimation of Panel Ordered-Response Models

Rajesh Paleti
The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin, Texas 78712
Tel: 512-471-4535, Fax: 512-475-8744
Email: rajeshp@mail.utexas.edu

and

Chandra R. Bhat*
The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin, Texas 78712
Tel: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu

*corresponding author

July 27, 2010

ABSTRACT

In the context of panel ordered-response structures, the current paper compares the performance of the maximum-simulated likelihood (MSL) inference approach and the composite marginal likelihood (CML) inference approach. The panel structures considered include the pure random coefficients (RC) model with no autoregressive error component, as well as the more general case of random coefficients combined with an autoregressive error component. The ability of the MSL and CML approaches to recover the true parameters is examined using simulated datasets. The results indicate that the performances of the MSL approach (with 150 scrambled and randomized Halton draws) and the simulation-free CML approach are of about the same order in all panel structures in terms of the absolute percentage bias (APB) of the parameters and econometric efficiency. However, the simulation-free CML approach exhibits no convergence problems of the type that affect the MSL approach. At the same time, the CML approach is about 6-8 times faster than the MSL approach for the simple random coefficients panel structure, and about 75 times faster than the MSL approach when an autoregressive error component is added. As the number of random coefficients increases, or if higher order autoregressive error structures are considered, one can expect even higher computational efficiency factors for the CML over the MSL approach. These results are promising for the use of the CML method for the quick, accurate, and practical estimation of panel ordered-response models with flexible and rich stochastic specifications.

1. INTRODUCTION

Ordinal discrete data arise in several empirical contexts, including ratings data (of consumer products, bonds, credit evaluation, movies, *etc.*), or likert-scale type attitudinal/opinion data (of air pollution levels, traffic congestion levels, school academic curriculum satisfaction levels, teacher evaluations, *etc.*), or grouped data (such as bracketed income data in surveys or discretized rainfall data). The traditional "workhorse" multinomial logit (MNL) model is not very appropriate to characterize the data generating process for such ordinal discrete data because the MNL ignores the potential correlation in errors between proximal ordinal response categories (Train, 2003, page 163; Greene, 2000, page 875). While more advanced unordered-response models (such as the ordered generalized extreme value logit or a mixed logit) may be considered to account for the proximal error correlation in alternatives, such models intrinsically assume that each ordinal category is assigned a utility, and the observed ordinal choice corresponds to the category with the highest utility (we will refer to this unordered decision process as a "vertical comparison" mechanism, where the latent utilities of the ordinal categories are arranged vertically and compared with one another, and the category with the highest utility is the observed outcome). While such unordered models may provide reasonable statistical fit and results, a more natural characterization of the generating mechanism is that the observed ordinal data correspond to a partitioning of a single latent continuous variable into mutually exclusive (non-overlapping) intervals (we will refer to this ordered decision process as a "horizontal partitioning" mechanism, where a single latent utility or propensity variable is horizontally partitioned by thresholds, and the observed ordinal category corresponds to the section of the real line in which the measure of the latent variable falls). The reader is referred to McKelvey and Zavoina (1975) and Winship and Mare (1984) for some early expositions of the ordered-response model formulation, and Liu and Agresti (2005) for a survey of recent developments. A recently published book by Greene and Hensher (2010) is another excellent and comprehensive resource on ordered-response models.

The number of applications of the ordered-response model has been burgeoning in several fields, including sociology, education, biology, marketing, and transportation sciences. Several of these applications have modeled the case of either repeated ordinal choice data (such as would be obtained from a stated preference exercise in which each respondent is asked to provide, at the same cross-sectional point in time, her/his opinion of a product multiple times

based on varying the attributes of the product) or panel-based ordinal data (similar to repeated choice data, except that these are actual revealed choices made by individuals over a period of time). In this paper, the focus is on the latter case because restricted versions of the models for panel data may be applied to repeated choice data. Within this panel context, the norm in the literature is to introduce random effects and/or random parameter heterogeneity to accommodate panel effects. Such terms lead to integration in the likelihood function during estimation, resulting, in general, in the need to use numerical simulation techniques based on a maximum simulated likelihood (MSL) approach (for example, see Bhat and Zhao, 2002, Greene, 2005, Greene and Hensher, 2009) or a Bayesian inference approach (for example, see Müller and Czado, 2005, Girard and Parent, 2001). However, such simulation-based approaches can become infeasible for some panel model specifications and for long panel data. Even if feasible, the numerical simulation methods can be time-consuming and can lead to convergence problems during estimation. For instance, Bhat *et al.*, (2010a) find that standard classical MSL approaches can be imprecise and have poor convergence properties, and Müller and Czado (2005) find that standard Bayesian MCMC approaches can be useless for panel ordered response model estimations due to bad convergence properties. As a consequence, another inference approach that has seen some use recently is the simulation-free composite marginal likelihood (CML) approach. This is an estimation technique that is gaining substantial attention in the statistics field, though there has relatively little coverage of this method in econometrics and other fields. The CML method, which belongs to the more general class of composite likelihood function approaches, is based on forming a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods. Under usual regularity assumptions, and based on the theory of estimating equations (see Lindsay, 1988, Cox and Reid, 2004), the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood). The maximum CML estimator should lose some efficiency from a theoretical perspective relative to a full likelihood estimator (if this is feasible), but this efficiency loss appears to be empirically small (see Zhao and Joe, 2005, Lele, 2006, and Joe and Lee, 2009).[1] Besides, the MSL approach also loses

[1] A handful of studies (see Hjort and Varin, 2008; Mardia *et al.*, 2009; Cox and Reid, 2004) have also theoretically examined the limiting normality properties of the CML approach, and compared the asymptotic variance matrices

efficiency since it involves simulation of the true analytically intractable likelihood function (see McFadden and Train, 2000). The CML approach also represents a conceptually, pedagogically, and implementationally simpler procedure relative to simulation techniques, and also has the advantage of reproducibility of results.

The focus of this paper is on comparing the performance of the maximum-simulated likelihood (MSL) approach with the composite marginal likelihood (CML) approach in panel ordered-response situations when the MSL approach is feasible.[2] We use simulated data sets with known underlying model parameters to evaluate the two estimation approaches. The ability of the two approaches to recover model parameters is examined, as is the sampling variance and the simulation variance of parameters in the MSL approach relative to the sampling variance in the CML approach. The computational costs of the two approaches are also presented.

The rest of this paper is structured as follows. In the next section, we present alternative model structures for panel ordered-response models, and discuss the maximum simulated likelihood (MSL) estimation method and the maximum CML estimation methods in the context of each of the alternative panel structures. Section 3 presents the experimental design for the simulation experiments. Section 4 presents the performance measures used for the comparison of the MSL and CML approaches, while Section 5 discusses the results. Section 6 concludes the paper by highlighting the important findings.

2. MODEL STRUCTURE

Let q be an index for individuals (q = 1, 2, …, Q), and let j be an index for the j^{th} observation (say at time t_{qi}) on individual q (j = 1, 2, …, J, where J denotes the total number of observations on individual q).[3] Let the observed discrete (ordinal) level for individual q at the j^{th} observation be m_{qj} (m_{qj} may take one of K values; *i.e.*, $m_{qi} \in \{1, 2, \ldots, K\}$). In the usual random-effects

from this approach with the maximum likelihood approach. However, such a precise theoretical analysis is possible only for very simple models, and becomes much harder for models such as a panel ordered-response system.

[2] Note that our discussions in the paper for the panel ordered-response situation are immediately applicable to the panel binary response situation, because the latter is but a special case of the former.

[3] We assume here that the number of panel observations is the same across individuals. Extension to the case of different numbers of panel observations across individuals does not pose any substantial challenges, and will be discussed later.

ordered response framework notation, we write the latent variable (y_{qj}^{*}) as a function of relevant covariates as:

$$y_{qj}^{*} = \boldsymbol{\beta}_q' \boldsymbol{x}_{qj} + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \psi^{m_{qj}-1} < y_{qj}^{*} < \psi^{m_{qj}}, \quad (1)$$

where $\boldsymbol{x}_{qj}$ is a (H×1)-vector of exogenous variables (including a constant), $\boldsymbol{\beta}_q$ is an individual-specific (H×1)-vector of coefficients to be estimated that is a function of unobserved individual attributes, ε_{qj} is a standard normal (or logistic) error term uncorrelated across individuals q (but it may be correlated across observations j (j = 1, 2, …, J) of the same individual, depending upon the analyst's specification) , and $\psi^{m_{qj}}$ is the upper bound threshold for discrete level m_{qj} ($\psi^0 < \psi^1 < \psi^2 ... < \psi^{K-1} < \psi^K$; $\psi^0 = -\infty, \psi^1 = 0, \psi^K = +\infty$).[4] Assume that the $\boldsymbol{\beta}_q$ vector in Equation (1) is a realization from a multivariate normal distribution $\phi(\beta)$ with a mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \boldsymbol{LL}'$, where $\boldsymbol{L}$ is the lower-triangular Choleski factor of $\boldsymbol{\Omega}$. Also, assume that the ε_{qj} term, which captures the idiosyncratic effect of all omitted variables for individual q at the j^{th} choice occasion, is independent of the elements of the $\boldsymbol{\beta}_q$ and $\boldsymbol{x}_{qj}$ vectors. We now discuss four different model structures, based on different assumptions about the $\boldsymbol{\beta}_q$ vector.

2.1 Random-Effects Model

The simplest panel model is one that includes an individual-specific constant term, but does not consider heterogeneity in other parameters in $\boldsymbol{\beta}_q$ across individuals q. Thus, we write $\boldsymbol{\beta}_q' \boldsymbol{x}_{qj} = \alpha_q + \boldsymbol{\gamma}' \boldsymbol{z}_q$, where the vector $\boldsymbol{z}_q$ now includes all the variables but no constant, and $\boldsymbol{\gamma}$ is a

[4] The model can be generalized in many ways, though the model as written is the most familiar and common panel version of the ordered-response model. For instance, the mean vector $\boldsymbol{b}$ of $\boldsymbol{\beta}_q$ can be a function of observed individual attributes. However, this can be accommodated without any complications by redefining $\boldsymbol{x}_{qj}$ to include interaction terms. Also, one can label the coefficient vector on the $\boldsymbol{x}_{qj}$ variable vector as $\boldsymbol{\beta}_{qj}$ (rather than $\boldsymbol{\beta}_q$) to allow for fixed or random elements of response specific to period j, and to capture random heterogeneity in response across individuals and choice occasions (see Bhat and Sardesai, 2006). This relabeling also then allows observed individual and choice occasion specific variable effects to be introduced in the covariance matrix of $\boldsymbol{\beta}_{qj}$. Finally, one can also include heterogeneity in the variance of ε_{qj} and accommodate heterogeneity in the thresholds $\psi^2, \psi^3, \ldots < \psi^{K-1}$ through careful parameterizations to ensure the ordinality conditions on the thresholds (see Eluru *et al*., 2008). However, all these generalizations cause an explosion in parameters, and need very rich data sets to estimate parameters.

fixed coefficient vector to be estimated. Substituting this expression in Equation (1), and writing α_q in random effects form as $\alpha_q = \alpha + \eta_q$, we get the following equation:

$$y^*_{qj} = \alpha + \boldsymbol{\gamma}'\boldsymbol{z_{qj}} + \eta_q + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \psi^{m_{qj}-1} < y^*_{qj} < \psi^{m_{qj}} \quad (2)$$

η_q in the above equation is an individual-specific random term that generates a correlation in the propensity across all of individuals *q*'s *J* observed choice occasions. It is typical to consider the heterogeneity term η_q to be normally distributed, since the central limit theorem can be invoked assuming the term is the sum of several small influences. However, other distributions may also be empirically tested, such as the logistic distribution with fatter tails. But the consideration of a normally distributed η_q with a standard normally distributed ε_{qj} is natural and convenient here, which is what we will assume. The result is the standard textbook random-effects ordered-response model, which takes the same form as the random-effects binary choice model proposed by Butler and Moffitt (1982).

2.1.1 Maximum Simulated Likelihood (MSL) Estimation of Random-Effects Model

The MSL estimation of the random-effects model is relatively straightforward. The probability of the observed vector m_q of the sequence of ordinal choices $(m_{q1}, m_{q2}, m_{q3}, ..., m_{qJ})$ for individual *q*, conditional on the heterogeneity term η_q, can be written as:

$$\text{Prob}(m_q)\big|\eta_q = \prod_{j=1}^{J}\left\{\Phi(\psi^{m_{qj}} - \alpha - \boldsymbol{\gamma}'\boldsymbol{z_{qj}} - \eta_q) - \Phi(\psi^{m_{qj}-1} - \alpha - \boldsymbol{\gamma}'\boldsymbol{z_{qj}} - \eta_q)\right\} \quad (3)$$

The unconditional likelihood of the observed choice sequence is obtained by integrating out the term η_q:

$$L_q(\boldsymbol{\psi}, \alpha, \boldsymbol{\gamma}, \sigma) = \int_{v=-\infty}^{\infty}\left[\prod_{j=1}^{J}\left\{\Phi(\psi^{m_{qj}} - \alpha - \boldsymbol{\gamma}'\boldsymbol{z_{qj}} - \sigma v) - \Phi(\psi^{m_{qj}-1} - \alpha - \boldsymbol{\gamma}'\boldsymbol{z_{qj}} - \sigma v)\right\}\phi(v)\,dv\right] \quad (4)$$

where $v = \frac{\eta_q}{\sigma}$, $\eta_q \sim N(0, \sigma^2)$, $\boldsymbol{\psi}$ is the vector of all threshold bounds, $\Phi(\cdot)$ is the univariate standard normal cumulative distribution, and $\phi(\cdot)$ is the corresponding univariate standard normal density function. Finally, the log-likelihood function may be written as:

$$\log L\ (\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma) = \sum_{q} \log L_q(\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma) \tag{5}$$

The log-likelihood function above can be maximized using Gauss-Hermite Quadrature or using a simulation method. Since, the function entails only a one dimensional integral, estimation is generally very fast and there is no convergence-related problems.

2.1.2 Composite Marginal Likelihood (CML) Estimation of Random-Effects Model

The composite marginal likelihood (CML) estimation approach (see Varin, 2008 and Varin *et al*., 2010 for good reviews) is a relatively simple approach that can be used when the full likelihood function is cumbersome or plain infeasible to evaluate due to the underlying complex dependencies, as is the case with certain specifications of panel models that entail high dimensional integration in the likelihood function. While there have been recent advances in simulation techniques within a classical or Bayesian framework that assist with such model estimation situations (see Bhat, 2003, Beron and Vijverberg, 2004, and LeSage, 2000), these techniques are impractical and/or infeasible in situations in some panel ordered-response situations (see, for example, Varin and Czado, 2010). Further, even when the integration is of low dimension, the CML method may have a substantial edge in terms of computation speed. The CML method, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), is based on forming a surrogate likelihood function that compounds easier-to-compute, lower-dimensional, marginal likelihoods. In panel data, the simplest CML, formed by assuming independence across observations from the same individual, entails the product of univariate densities (for continuous data) or probability mass functions (for discrete data). However, this approach does not provide estimates of dependence among the individual observations. Another approach is the pairwise likelihood function formed by the product of power-weighted likelihood contributions of all or a selected subset of couplets (*i.e.*, pairs of observations). This pairwise method corresponds to a composite marginal approach based on bivariate marginals. For individual q, the pairwise likelihood function is:

$$L_{CML,q}(\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma) = \left(\prod_{j=1}^{J-1} \prod_{g=j+1}^{J} \left[\Pr\left(y_{qj} = m_{qj}, y_{qg} = m_{qg} \right) \right] \right)^{w_q}, \tag{6}$$

where w_q is a power weight to be chosen based on efficiency considerations (see Kuk and Nott, 2000; Zhao and Joe, 2005; Joe and Lee, 2009). When the number of choice occasions are the

same across individuals, as we assume in the current paper, this power weight term may be ignored and arbitrarily set to one for each individual. However in the more general case when the number of observations from individual q is J_q, setting w_q to be one for all individuals will give more weight to individuals who have more choice occasions than to individuals who have fewer choice occasions. Le Cessie and Van Houwelingen (1994) suggest, based on their correlated binary model analysis, that each individual should contribute about equally to the CML function. This may be achieved by power-weighting each individual's likelihood contribution by a factor that is the inverse of the number of choice occasions minus one (in our context, this is $[J_q - 1]^{-1}$.). The net result is that the composite likelihood contribution of individual q collapses to the likelihood contribution of the individual under the case of independence across choice occasions. In a recent paper, Joe and Lee (2009) theoretically studied the issue of efficiency in the context of a simple random-effects binary choice model. They indicate that the weights suggested by Le Cessie and Van Houwelingen (1994) can provide poor efficiency when the correlation between pairs of the underlying latent variables for the repeated binary choices over time is moderate to high. Intuitively, Joe and Lee's discussion is based on the concept that, when there is perfect dependence between each pair of inter-temporal binary choices, each pairing should contribute the same amount of information to the CML function. While Joe and Lee's theoretical analysis is confined to a simple random-effects binary model, it may be extended to the random-effects ordered panel case (and also to other ordered-response panel models discussed later). Joe and Lee (2009) proposed the optimal power weight for individual q in the unbalanced panel case as $w_q = (J_q - 1)^{-1}[1 + 0.5(J_q - 1)]^{-1}$. In the rest of this paper, we will ignore the weight term, since we are focusing our simulation experiments on the case of the same number of choice occasions from each individual.[5]

To write the pairwise likelihood function in terms of the parameters to be estimated in the simple random-effects model, note that the joint distribution of the latent variables $(y_{q1}^*, y_{q2}^*, ..., y_{qJ}^*)$ for the q^{th} individual is multivariate normal with standardized mean vector

[5] The focus in the current paper is on comparing the performance of the maximum simulated likelihood approach with the CML approach, so we steer clear of issues related to optimal weights for the CML approach by considering the "equal observations across individuals" case.

$\frac{\alpha+\gamma' z_{q1}}{\mu}, \frac{\alpha+\gamma' z_{q2}}{\mu}, \ldots\ldots \frac{\alpha+\gamma' z_{qJ}}{\mu}$ and a correlation matrix with constant non-diagonal entries $\frac{\sigma^2}{\mu^2}$, where $\mu = \sqrt{1+\sigma^2}$. Then, we can write

$$L_{CML,q}(\psi,\alpha,\gamma,\sigma) = \prod_{j=1}^{J-1}\prod_{g=j+1}^{J}\begin{bmatrix}\Phi_2(\delta^{m_{qj}},\delta^{m_{qg}},\rho_{jg}) - \Phi_2(\delta^{m_{qj}},\delta^{m_{qg}-1},\rho_{jg}) \\ -\Phi_2(\delta^{m_{qj}-1},\delta^{m_{qg}},\rho_{jg}) + \Phi_2(\delta^{m_{qj}-1},\delta^{m_{qg}-1},\rho_{jg})\end{bmatrix}, \quad (7)$$

where $\delta^{m_{qj}} = \frac{\psi^{m_{qj}} - \alpha - \gamma' z_{qj}}{\mu}, \mu = \sqrt{1+\sigma^2}, \rho_{jg} = \frac{\sigma^2}{\mu^2}$

The logarithm of the pairwise likelihood function is:

$$\log L_{CML}(\psi,\alpha,\gamma,\sigma) = \sum_q \log L_{CML,q}(\psi,\alpha,\gamma,\sigma) \quad (8)$$

The CML estimator $\hat{\theta} = (\psi,\alpha,\gamma',\sigma)'$ obtained by maximizing the above function is consistent and asymptotically normally distributed with the asymptotic variance matrix vector given by the inverse of the Godambe's (1960) sandwich information matrix:

$$V(\theta) = [G(\theta)]^{-1} = [H(\theta)]^{-1} J(\theta) [H(\theta)]^{-1}, \quad (9)$$

where

$$H(\theta) = E\left[\frac{-\partial^2 \log L_{CML}(\theta)}{\partial\theta\,\partial\theta'}\right] \text{ and } J(\theta) = E\left[\left(\frac{\partial \log L_{CML}(\theta)}{\partial\theta}\right)\left(\frac{\partial \log L_{CML}(\theta)}{\partial\theta'}\right)\right]$$

The above matrices can be estimated at the CML estimate $\hat{\theta}$ in the sample as follows:

$$\hat{H}(\hat{\theta}) = \left[\sum_{q=1}^{Q}\frac{-\partial^2 \log L_{CML,q} \mid \theta}{\partial\theta\,\partial\theta'}\right]_{\hat{\theta}} = -\left[\sum_{q=1}^{Q}\sum_{j=1}^{J-1}\sum_{g=j+1}^{J}\frac{\partial^2 \log \Pr(y_{qj}=m_{qj}, y_{qg}=m_{qg})}{\partial\theta\,\partial\theta'}\right]_{\hat{\theta}},$$

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

2.2 Random Coefficients Model

In this model, the coefficients on the exogenous variables are also considered to be randomly distributed. Going back to Equation (1), assume that β_q is multivariate normal distributed with

mean vector $\boldsymbol{b}$ and covariance $\boldsymbol{\Omega} = \boldsymbol{L}\boldsymbol{L}'$. Note that it is not necessary that all elements of the $\boldsymbol{\beta}_q$ be random. That is, the analyst may specify fixed coefficients on some exogenous variables in the model, though it will be convenient in presentation to assume that all elements of $\boldsymbol{\beta}_q$ are random.

2.2.1 Maximum Simulated Likelihood Estimation

The likelihood function contribution of individual q for the random coefficients model is:

$$L_q(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega}) = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left[\prod_{j=1}^{J} \left\{ \Phi(\psi^{m_{qj}} - \boldsymbol{\beta}' \boldsymbol{x}_{qj}) - \Phi(\psi^{m_{qj}-1} - \boldsymbol{\beta}' \boldsymbol{x}_{qj}) \right\} f(\boldsymbol{\beta} \mid \boldsymbol{b}, \boldsymbol{\Omega}) \, \mathrm{d}\boldsymbol{\beta} \right] \tag{11}$$

where $f(.)$ is multivariate normal density function with mean vector $\boldsymbol{b}$ and covariance $\boldsymbol{\Omega}$

The log-likelihood function is:

$$\log L(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega}) = \sum_{q} \log L_q(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega}) \tag{12}$$

The expression $\sum_{q} \log L_q(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega})$ entails integration of dimension equal to the number of elements of $\boldsymbol{\beta}_q$. The estimation of the log-likelihood function cannot, in general, be pursued using quadrature techniques due to the curse of dimensionality. Instead, it is typical to use quasi-Monte Carlo (QMC) techniques for simulation estimation (Bhat, 2001, 2003). To ensure the positive definiteness of the covariance matrix $\boldsymbol{\Omega}$, the likelihood function contribution of individual q of Equation (11) is rewritten in terms of the Cholesky-decomposed matrix $\boldsymbol{L}$ of $\boldsymbol{\Omega}$. The maximum simulated likelihood approach then proceeds by optimizing with respect to the elements of $\boldsymbol{L}$ rather than $\boldsymbol{\Omega}$. Once convergence is achieved, the implied covariance matrix $\boldsymbol{\Omega}$ may be reconstructed from the estimated matrix $\boldsymbol{L}$.

While there have been important advances in terms of the QMC based simulation of the mixed panel models for random coefficients, these QMC methods continue to be quite expensive for the usual sample sizes encountered in practice. Besides, even for low to moderate dimensions of integration (of the order of four to seven), the numerical simulators can lead to numerical instability, non-convergence, and imprecision problems as the number of dimensions increases. Bhat *et al.*, 2010a find another bothersome issue with these MSL simulation methods even for

low to moderate dimensions in that even if the log-likelihood function is computed with good precision, so that the simulation error in the estimated parameters is small, the computation of the numerical Hessian is not very reliable. But a good estimate the Hessian is needed for the sandwich estimator of the covariance matrix in the MSL method (the alternative of using the inverse of the cross product of the first derivatives is not appropriate in the MSL because of simulation noise introduced when using a finite number of draws per individual, see McFadden and Train, 2000). The only way out of the problem is to compute the log likelihood function with a very high level of precision, which can lead to high computational times even at low dimensions.

2.2.2 CML Estimation

The pairwise marginal likelihood function for the random coefficients panel ordered-response model is much simpler than the full likelihood function in Equation (11), as also suggested by Renard *et al*., (2004) in the context of a panel binary choice model. In particular, based on the joint distribution of the latent variables $(y_{q1}^*, y_{q2}^*, ..., y_{qJ}^*)$ for the q^{th} individual, one can write the contribution of the q^{th} individual to the pairwise-likelihood function as:

$$L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega}) = \left(\prod_{j=1}^{J-1} \prod_{g=j+1}^{J} \begin{bmatrix} \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}}, \rho_{qjg}) - \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}-1}, \rho_{qjg}) \\ -\Phi_2(\delta^{m_{qj}-1}, \delta^{m_{qg}}, \rho_{qjg}) + \Phi_2(\delta^{m_{qj}-1}, \delta^{m_{qg}-1}, \rho_{qjg}) \end{bmatrix} \right)^{w_q}, \tag{13}$$

where $\delta^{m_{qj}} = \dfrac{\psi^{m_{qj}} - \boldsymbol{b}' \boldsymbol{x}_{qj}}{\sqrt{\text{Var}(y_{qj}^*)}}$ and $\rho_{qjg} = \dfrac{\text{Cov}(y_{qj}^*, y_{qg}^*)}{\sqrt{\text{Var}(y_{qj}^*)}\sqrt{\text{Var}(y_{qg}^*)}}$

In the above expression, the $\text{Var}(y_{qj}^*)$, $\text{Var}(y_{qg}^*)$ and ρ_{qjg} terms are obtained by picking off the appropriate 2×2 sub-matrix of the covariance matrix of $(y_{q1}^*, y_{q2}^*, ..., y_{qJ}^*)$ given by $(\boldsymbol{x_q \Omega x_q'} + \boldsymbol{I_J})$ where $\boldsymbol{x_q}$ is a $J \times K$ matrix corresponding to the J choice occasions and K exogenous variables (including the constant) obtained by vertically concatenating the transpose of the $K \times 1$ vector $\boldsymbol{x_{qj}}$ $(j = 1, 2, ..., J)$, and $\boldsymbol{I_J}$ is the identity matrix of size J. The logarithm of the pairwise likelihood function is:

$$L_{CML}(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\Omega}) = \sum_q \log L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\Omega}), \tag{14}$$

The asymptotic variance expression is given by the sandwich estimator, as discussed earlier in Section 2.1.2.

The random coefficients model is commonly referred to as the mixed model in the literature, and the CML approach above is an alternative to the commonly used MSL approach. As in the MSL case, one can ensure the positive-definiteness of $\mathbf{\Omega}$ in the CML method by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of $\mathbf{\Omega}$ and maximizing with respect to these elements of the Cholesky factor. Essentially, this entails passing the Cholesky elements as parameters to the optimization routine, constructing the $\mathbf{\Omega}$ matrix internal to the optimization routine, and then picking off the appropriate sub-matrix for the pairwise likelihood components.

2.3 Random Effects Autoregressive Structure

The standard random-effects ordered-response model of Equation (1) allows easy estimation, since there is only a one-dimensional integral for each individual. However, the assumption of equal correlation across the multiple observations on the same individual is questionable, especially for medium-to-long panels. An alternative would be to allow a time-stationary error component, but also allow serial correlation within each subject-specific series of observations (see Varin and Czado (2010) and Bhat *et al.* 2010a). For instance, one may adopt an autoregressive structure of order one for the error terms of the same individual, so that $corr(\varepsilon_{qj}, \varepsilon_{qg}) = \rho^{|t_{qj} - t_{qg}|}$ $(0 < \rho < 1)$, where t_{qj} is the measurement time of observation y_{qj}. This is in addition to the equal correlation across observations of the same individual, due to the individual specific random term η_q in the Equation (2). The autoregressive error structure specification results in a joint multivariate distribution of the latent variables $(y^*_{q1}, y^*_{q2}, ..., y^*_{qJ})$ for the q^{th} individual with standardized mean vector $\frac{\alpha + \boldsymbol{\gamma}' \boldsymbol{z}_{q1}}{\mu}, \frac{\alpha + \boldsymbol{\gamma}' \boldsymbol{z}_{q2}}{\mu}, ... \frac{\alpha + \boldsymbol{\gamma}' \boldsymbol{z}_{qJ}}{\mu}$ and a correlation matrix $\mathbf{\Sigma}$ with entries such that $corr(y^*_{qj}, y^*_{qg}) = (\sigma^2 + \rho^{|t_{qj} - t_{qg}|}) / \mu^2$, where $\mu = \sqrt{1 + \sigma^2}$.

2.3.1 Maximum Simulated Likelihood Estimation

The random effects autoregressive model structure, while much more realistic than the simple random effects, also costs dearly in terms of computational time. In particular, rather than a single dimension of integration, we now have an integral of dimension J for individual q. The likelihood function for individual q is:

$$L_q(\boldsymbol{\psi}, \alpha, \boldsymbol{\gamma}, \sigma, \rho) = \Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qJ} = m_{qJ})$$

$$L_q(\boldsymbol{\psi}, \alpha, \boldsymbol{\gamma}, \sigma, \rho) = \int_{w_1=\delta^{m_{q1}-1}}^{\delta^{m_{q1}}} \int_{w_2=\delta^{m_{q2}-1}}^{\delta^{m_{q2}}} \cdots \int_{w_J=\delta^{m_{qJ}-1}}^{\delta^{m_{qJ}}} \phi_J(w_1, w_2, ..., w_J \mid \boldsymbol{\Sigma}) dw_1 dw_2 ... dw_J \quad (15)$$

where $\delta^{m_{qj}} = \dfrac{(\psi^{m_{qj}} - \alpha - \boldsymbol{\gamma}' \mathbf{z}_{qj})}{\mu}$, $\mu = \sqrt{1+\sigma^2}$, and ϕ_J is the J-variate standard multivariate normal density function. The integral above may be evaluated using the Geweke-Hajivassililiou-Keane (GHK) simulator (see Geweke, 1991, Hajivassiliou and McFadden, 1998, and Keane, 1994) or the Genz-Bretz (GB) simulator (Genz and Bretz, 1999; 2002, and Mi *et al.*, 2009), which are among the most effective simulators for evaluating rectangular multivariate normal probabilities (*i.e.*, bounded as opposed to unbounded limits of integration). Positive definiteness of the correlation matrix is guaranteed as long as $\sigma > 0$, and $0 < \rho < 1$, which can be easily imposed through appropriate parameterizations.

The problems with the MSL approach for the random-effects autoregressive model are similar to the ones discussed earlier in the context of the random coefficients model. However, while the dimension of random coefficients and therefore the dimensionality of the integration may be relatively low for the random coefficients model, the number of observations per individual, and therefore the dimensionality of integration, can be very high for the random-effects autoregressive model. For instance, in Varin and Czado (2010), the authors examine the headache pain intensity of patients at different points of time during the day and across several consecutive days. In this study, the full information likelihood estimation has of the order of 800 dimensions of integration for some of the individual-specific likelihood contributions, an infeasible proposition for model parameter estimation using any computer intensive simulation procedure.

2.3.2 Composite Marginal Likelihood Estimator

The pairwise marginal likelihood function for individual q in the random-effects autoregressive structure is:

$$L_{CML,q}(\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma,\rho)=\prod_{g=j+1}^{J}\prod_{j=1}^{J-1}\left[\begin{array}{c}\Phi_2\left(\delta^{m_{qj}},\delta^{m_{qg}},\rho_{jg}\right)-\Phi_2\left(\delta^{m_{qj}},\delta^{m_{qg-1}},\rho_{jg}\right)\\-\Phi_2\left(\delta^{m_{qj}-1},\delta^{m_{qg}},\rho_{jg}\right)+\Phi_2\left(\delta^{m_{qj}-1},\delta^{m_{qg}-1},\rho_{jg}\right)\end{array}\right], \tag{16}$$

where $\delta^{m_{qj}}=(\psi^{m_{qj}}-\alpha-\boldsymbol{\gamma}'\boldsymbol{z}_{qj})/\mu,\ \ \mu=\sqrt{1+\sigma^2}$, and $\rho_{jg}=(\sigma^2+\rho^{|t_{qj}-t_{qg}|})/\mu^2$.

The logarithm of the likelihood function is:

$$LL_{CML}(\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma,\rho)=\sum_q \log L_{CML,q}(\boldsymbol{\psi},\alpha,\boldsymbol{\gamma},\sigma,\rho)\,. \tag{17}$$

Compared to the MSL technique, the pairwise approach only entails bivariate normal distributions, which can be evaluated rapidly. The asymptotic covariance matrix may be obtained as the inverse of the Godambe sandwich information estimator, as in Section 2.1.2.

2.4 Random Coefficients Autoregressive Structure

This general structure combines the random coefficient structure with the autoregressive structure. The form of the model is as follows (using the same notations as earlier):

$$y^*_{qj}=\boldsymbol{\beta}'_{\boldsymbol{q}}\boldsymbol{x}_{\boldsymbol{qj}}+\varepsilon_{qj},\ y_{qj}=m_{qj}\ \text{ if }\ \psi^{m_{qj}-1}<y^*_{qj}<\psi^{m_{qj}} \tag{18}$$

with $\boldsymbol{\beta}_{\boldsymbol{q}}\sim N(\boldsymbol{b},\boldsymbol{\Omega}),\varepsilon_{qj}\sim N(0,1)$, and $corr(\varepsilon_{qj},\varepsilon_{qg})=\rho^{|t_{qj}-t_{qg}|},0<\rho<1$

2.4.1 Maximum Simulated Likelihood Estimation

$$L_q(\boldsymbol{\psi},\boldsymbol{b},\boldsymbol{\Omega},\rho)=\int_{\omega_1=\delta^{m_{q1}-1}}^{\delta^{m_{q1}}}\int_{\omega_2=\delta^{m_{q2}-1}}^{\delta^{m_{q2}}}\cdots\int_{\omega_J=\delta^{m_{qJ}-1}}^{\delta^{m_{qJ}}}\phi_J(\omega_1,\omega_2,...,\omega_J\,|\,\boldsymbol{\Sigma}_{\boldsymbol{q}})d\omega_1 d\omega_2...d\omega_J \tag{19}$$

where $\delta^{m_{qj}}=\dfrac{(\psi^{m_{qj}}-\boldsymbol{b}'\boldsymbol{x}_{\boldsymbol{qj}})}{\sqrt{\mathrm{Var}(y^*_{qj})}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{q}}$ is a correlation matrix with entries such that $\Sigma_{qjg}=\dfrac{\mathrm{Cov}(y^*_{qj},y^*_{qg})}{\sqrt{\mathrm{Var}(y^*_{qj})\mathrm{Var}(y^*_{qg})}}$. The covariance matrix of the latent $(y^*_{q1},y^*_{q2},...,y^*_{qJ})$ dependent

variables is given by $(\boldsymbol{x_q \Omega x_q'} + \boldsymbol{R_q})$ where $\boldsymbol{R_q}$ is a $J \times J$ matrix with the entry ($\rho^{|t_{qj} - t_{qg}|}$) for the jg^{th} element of the matrix (j = 1, 2,…, J; g = 1, 2,…, J). This covariance matrix is positive definite as long as $\boldsymbol{\Omega}$ can be written as $\boldsymbol{LL'}$ (where $\boldsymbol{L}$ is the lower triangular Cholesky factor of $\boldsymbol{\Omega}$) and $0 < \rho < 1$. The log-likelihood function is finally computed as:

$$\sum_q \log L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{b}, \boldsymbol{\Omega}, \rho) \tag{20}$$

The MSL estimation of the random coefficients autoregressive model is particularly cumbersome, and we are not aware of any earlier literature considering such a model, even though it naturally arises as a combination of random coefficients and a first-order autocorrelation process.

2.4.2 Composite Marginal Likelihood Estimation

The pairwise function for individual q in the random coefficients autoregressive structure takes the same form as for the random coefficients structure (see Equation 13), but with the important difference that the $\text{Var}(y^*_{qj})$, $\text{Var}(y^*_{qg})$ and ρ_{qjg} are obtained by picking off the appropriate (2×2) sub-matrix of the covariance matrix of $(y^*_{q1}, y^*_{q2}, ..., y^*_{qJ})$ given by $(\boldsymbol{x_q \Omega x_q'} + \boldsymbol{R_q})$ (rather than $(\boldsymbol{x_q \Omega x_q'} + \boldsymbol{I_J})$ in the random coefficients models). The logarithm of the pairwise likelihood function and the asymptotic variance expression of the estimator are obtained in the usual way.

3. SIMULATION STUDY

In the current paper, we assess the performance of the CML technique for panel ordered response models in the context of the random coefficients structure and the random coefficients autoregressive structure. This is because of three reasons. First, the random coefficients structure subsumes the random effects structure as a special case. Second, the random coefficients structure (without autoregressive error terms) has been extensively used in binary and ordered response modeling (for example, see Bhat and Zhao, 2002, Greene, 2000). The use of the CML technique can lead to a reduction in computational time for these mixed models, and may be the only practical approach if there are numerous random coefficients. Third, the random coefficients autoregressive structure subsumes all other structures as special cases, and is a general panel specification that, to our knowledge, has not been considered in the literature.

3.1 Experimental Set-up

In the simulation set-up, we consider five choice occasions ($J = 5$) per individual and 500 individuals, with five independent variables per choice occasion. While the number of choice occasions per individual and the number of independent variables per choice occasion can be much larger, we use five choice occasions and five independent variables so that the MSL estimation (which entails a five dimensional integral) is manageable in the context of the many MSL runs we undertake in the paper, while also being reasonably realistic of the kinds of panel setting encountered in practice. The intent is to compare the MSL estimation results and the CML results in terms of the ability to recover parameters as well as computational time. For all the datasets generated in the experimental design, the values of each of the five independent variables are drawn from a standard univariate normal distribution. In the subsequent two sections, we discuss the set-up for each of the random coefficients (RC) and the random coefficients autoregressive (RCA) structures in detail.

3.1.1 The RC Structure

In the RC structure, a coefficient vector $\boldsymbol{\beta}_q$ (specific to each individual) is assumed and is drawn from a multivariate normal distribution with a mean vector of $\boldsymbol{b}$ (= 1.5, 1, 2, 1, 2). We then consider both independent realizations as well as correlated realizations for the coefficient vector $\boldsymbol{\beta}_q$ (across exogenous variables for each individual q):

$$\boldsymbol{\Omega}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\Omega}_2 = \begin{bmatrix} 1 & -0.5 & 0.25 & 0.75 & 0 \\ -0.5 & 1 & 0.25 & -0.5 & 0 \\ 0.25 & 0.25 & 1 & 0.33 & 0 \\ 0.75 & -0.5 & 0.33 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (21)$$

For each of the above two positive-definite covariance matrices, the random vector realization of $\boldsymbol{\beta}_q$ is applied to the qth individual's choice occasions, and is linearly combined with the corresponding vector of independent variables ($\boldsymbol{\beta}_q' \boldsymbol{x}_{qj}$). The result is added to an independent standard normal error term draw (ε_{qj}) as in Equation (1) to obtain a value of y_{qj}^*. This is then translated to "observed" values of y_{qj} based on the following pre-specified threshold

values: $\psi^1 = 0,\ \psi^2 = 1.0,\ \psi^3 = 2.0$ (thus we assume four outcome levels for the ordinal variable). The above data generation process is undertaken 50 times with different realizations of the $\boldsymbol{\beta}_q$ vector and the error term ε_{qj} to generate 50 different data sets (for each of the $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ specifications).

The MSL and CML estimation procedures are applied to each data set to estimate data specific values of $\boldsymbol{b}$, $\boldsymbol{L}_1$ ($\boldsymbol{\Omega}_1 = \boldsymbol{L}_1\boldsymbol{L}_1'$) or $\boldsymbol{L}_2$ ($\boldsymbol{\Omega}_2 = \boldsymbol{L}_2\boldsymbol{L}_2'$) (as appropriate), and the threshold vector $\boldsymbol{\psi} = \left(\psi^1, \psi^2, \psi^3\right)'$.[6] Note that $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are the lower Cholesky decompositions of the covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, respectively. We estimate the Choleski parameters to ensure the positive definiteness of the variance-covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$. For the MSL estimation, we use draws from the randomized Halton sequence for the random coefficients vector $\boldsymbol{\beta}_q$, because it is the most commonly used QMC sequence in the literature (Halton, 1960; see Bhat, 2003 for a discussion). Within the context of the Halton draws, we experimented with different kinds of scramblings and randomizations of the Halton sequence. This included the following: (a) scrambling the Halton draws for different dimensions using the Bratten-Weller approach to break correlations across dimensions arising from the periodic cycling of the Halton draws, (b) scrambling the Halton draws using a randomization approach to break correlations across dimensions, (c) randomizing the Halton draws along each dimension by adding a uniform random number that still preserves the uniformly distributed and equi-distribution properties of the underlying Halton sequence (referred to as the Tuffin-randomization in the literature; see Tuffin, 1996), and (d) randomizing the assignment of Halton dimensions to the random coefficients (so that, for example, the Halton dimension that is based off the prime number two is assigned to say the first random coefficient in one of the simulation runs, while the same Halton dimension is assigned to a different random coefficient in another simulation run). Our experiments suggested that the best performance was obtained using a procedure that combined Bratten-Weller scrambling with the Tuffin randomization as well as the random assignment of Halton dimensions to coefficients. Finally, while a higher number of draws per individual (based on the combination scrambling/randomization discussed above) generally provided improved results, we used 150 draws per individual, which is about what is typically used in most

[6] Since the vector $\boldsymbol{x}_{qj}$ does not include a constant in the simulation set-up, the first threshold ψ^1 is estimable.

applications of the MSL procedure for ordered-response models. As we will indicate later, we also undertook the MSL estimation with 250 draws per individual, and the results were not substantially different. To assess and quantify simulation variance, the randomized and scrambled Halton-based simulation procedure is applied to each dataset 10 times with different (independent) randomized Halton draw sequences.

3.1.2 The RCA Structure

For the RCA structure, we generate $\boldsymbol{\beta}_q$ vectors for each individual q based on the mean vector $\boldsymbol{b}$ (= 1.5, 1, 2, 1, 1) and a covariance matrix given by $\boldsymbol{\Omega}_2$ (we use the more general non-diagonal covariance matrix used in the RC structure for the RCA structure). The rest of the procedure is the same as the RC structure, except for the generation of the standard normal error terms (ε_{qj}). Specifically, these error terms are now serially correlated for each individual q. We assume that this serial correlation gets manifested in the last four of the five observations for each individual, with the first observation error term ε_{q1} for each individual randomly drawn from a standard normal distribution. That is, ε_{qj} (j = 1, 2 …J) is generated for each individual q as follows:

$$\varepsilon_{qj} = \begin{cases} \eta_{q1} \overset{iid}{\sim} N(0,1) \text{ for } j = 1 \\ \rho\varepsilon_{qj-1} + \left(\sqrt{1-\rho^2}\right)\eta_{qj}, \quad \eta_{qj} \overset{iid}{\sim} N(0,1) \text{ for } j \geq 2. \end{cases} \tag{22}$$

The resulting correlation matrix of ε_{qj} is $\boldsymbol{R}_q$ which is a $J \times J$ dimension matrix with its jg^{th} element being $\rho^{|j-g|}$, $g,\ j = 1, 2, ..., J$.[7] In the current paper, we undertake the simulation exercise for low autocorrelation $(\rho = 0.3)$ and high autocorrelation $(\rho = 0.7)$. For each of these correlation values, error term realizations of $\boldsymbol{\varepsilon}_q = (\varepsilon_{q1}, \varepsilon_{q2}, ..., \varepsilon_{qJ})'$ are drawn and used to generate data for the RCA structure. To examine the impact of different magnitudes of the autoregressive correlation parameter, the process is undertaken 20 times with different realizations of the $\boldsymbol{\beta}_q$ and $\boldsymbol{\varepsilon}_q$ vectors to generate 20 different data sets (for each value of ρ of 0.3 and 0.7). We used

[7] Note that this is the autoregressive structure of order 1. One can also use more complicated autoregressive structures of order p for the error terms, or use more general structures for the error correlation. For instance, while we focus on a time series context, in spatial contexts related to ordered-response modeling, Bhat *et al.* (2010b) developed a specification where the correlation in physical activity between two individuals may be a function of several measures of spatial proximity and adjacency.

fewer data sets for the RCA case compared to the RC case because of the substantially increased computational cost for the RCA case with the MSL method. Further, for the MSL approach, it was too expensive to run estimations multiple times with the same data using 150 Halton draws per individual. At the same time, the results from the RA non-diagonal structure (to be discussed later) clearly indicate that one cannot ignore away the simulation error, because it is a sizeable fraction of the asymptotic error in the RA case. Thus, for the two non-diagonal RCA structures (with low and high autocorrelation), we applied the average magnitude of simulation standard error (as a percentage of the asymptotic standard error) as obtained from the non-diagonal RA structure to the asymptotic standard errors of the RCA non-diagonal structures to estimate the mean simulation standard error for each parameter. The simulation errors estimated in this way for the RCA non-diagonal cases may be regarded as lower bounds of the true simulation errors, since the simulation errors are expected to be higher for the case with autocorrelation than without autocorrelation

The MSL and CML estimation procedures are applied to each data set to estimate data-specific values of $\boldsymbol{b}, \boldsymbol{L}_2, \boldsymbol{\psi}$, and ρ. To ensure that $0 < \rho < 1$, we re-parameterized ρ as follows: $\rho = 1/[1+\exp(-\Delta)]$. The MSL estimation procedure uses the GHK simulation procedure using 150 draws per individual of the randomized and scrambled Halton sequence (see Bhat *et al*., 2010a for a discussion of the GHK simulator in the context of ordered response models).

4. PERFORMANCE COMPARISON BETWEEN THE MSL AND CML APPROACHES

In this section, we first identify a number of performance measures and discuss how these are computed for the MSL approach and the CML approach. The subsequent sections present the simulation and computational results.

4.1 Performance Measures

The steps discussed below for computing performance measures are for a specific covariance matrix pattern. For the RC model, we consider zero covariance across the random coefficients (diagonal covariance specification) and non-zero covariance across the random coefficients (non-diagonal covariance specification). For the RCA structure, we retain the non-diagonal covariance matrix specification of the RC structure and then consider two correlation patterns, corresponding to the autoregressive correlation parameter values of 0.3 and 0.7.

MSL Approach

(1) Estimate the MSL parameters for each data set s ($s = 1, 2, ..., 50$) and for each of 10 independent draws, and obtain the time to get the convergent values and the standard errors. Note combinations for which convergence is not achieved. Everything below refers to cases when convergence is achieved. Obtain the mean time for convergence (TMSL) and standard deviation of convergence time across the converged runs and across all data sets (the time to convergence includes the time to compute the covariance matrix of parameters and the corresponding parameter standard errors). All estimations are started with the true parameter values as the starting values. While multiple computers had to be used for the many different runs undertaken in this paper, all the run times were carefully scaled to the equivalent time on a desktop computer with 2.66GHz Core2Duo processor and 3.25GB of RAM. The scaling was based on extensive experimentation on different computers.

(2) For each data set *s* and draw combination, estimate the standard errors (s.e.) of parameters (using the sandwich estimator).

(3) For each data set *s*, compute the mean estimate for each model parameter across the draws. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the finite sample **absolute percentage bias** (APB) as: $\text{APB} = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$ [8]

(4) For each data set *s*, compute the median standard error for each model parameter across the draws. Call this MSED, and then take the mean of the MSED values across the data sets and label this as **the asymptotic standard error** (essentially this is the standard error of the distribution of the estimator as the sample size gets large). Next, for each data set *s*, compute the simulation standard deviation for each parameter as the standard deviation in the estimated values across the independent draws (about the MED value**). Call this standard deviatio**n as SIMMED. For each parameter, take the mean of SIMMED across the different data sets. Label this as the **simulation standard error** for each parameter. The above procedure is used to obtain the simulation standard error values for the two RC

[8] If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1in the denominator, and multiplying by 100.

cases. For the two RCA cases, the simulation standard error values are estimated as discussed in the previous section.

(5) For each parameter, compute a **simulation adjusted standard error** as **follows:** $\sqrt{(\text{asymptotic standard error})^2 + (\text{simulation standard error})^2}$

CML Approach

(1) Estimate the CML parameters for each data set *s* and obtain the time to get the convergent values (including the time to obtain the Godambe matrix-computed covariance matrix and corresponding standard errors). Determine the mean time for convergence (TCML) across the *S* data sets.[9]

(2) For each data set *s*, estimate the standard errors (s.e.) (using the Godambe estimator).

(3) Compute the **mean estimate** for each model parameter across the *R* data sets. Compute **absolute percentage bias** as in the MSL case.

(4) Compute the median standard error for each model parameter across the *R* data sets and label this as the **asymptotic standard error.**

5. RESULTS

5.1 RC Structure

Tables 1a and 1b provide the results of the RC structure for the diagonal and non-diagonal cases, respectively. The tables provide the true value of the parameters (second column), followed by the maximum simulated likelihood (MSL) estimation results and the composite marginal likelihood (CML) estimation results.

The Diagonal Covariance Matrix Case

The columns under "parameter estimates" in Table 1a provides the mean parameter estimates across data sets and runs, as well as the absolute percentage bias (APB) values. These results indicate that both the MSL and CML methods perform reasonably well in recovering parameters. Specifically, the APB values for the parameters range from 1.05% to 16.08 % for the MSL method and 0.71% to 10.38 % for the CML method. The mean APB using 150 Halton draws per individual in the MSL case is 7.55% (see the row of the table labeled "Overall mean value across

[9] The CML estimator always converged in our simulations, unlike the MSL estimator.

parameters" and the column titled "absolute percentage bias"), while the mean APB from the CML approach is lower at 5.51%. There are two interesting observations from the mean estimate and APB values. First, the MSL estimation underestimates parameter values (note that the mean estimate values from the MSL are consistently lower than the true values), while the CML estimation overestimates parameter values (except, for the first parameter). This behavior, even if very small in magnitude based on the APB values, is worthy of further study. Second, the APB values for the Choleski parameters (*i.e.*, the l values in the table, which in this diagonal case are the standard deviations of the distributions of each of the five random coefficients) are generally somewhat higher relative to the threshold parameters (*i.e.*, the ψ values in the table) and the mean values of the distributions of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the b values in the table), especially so in the MSL estimation. Also, there is more variation in the APB values among the Cholesky parameters than among the b and ψ values in both the MSL and CML estimations. This is perhaps because the Choleski parameters enter the likelihood function in a more complex non-linear fashion than other parameters, leading to a relatively flat log-likelihood function for different values of standard deviations of the random coefficients and more difficulty in accurately recovering these standard deviation parameters.

The sampling standard error values of the parameters indicate good efficiency of both the MSL and CML estimators, with the asymptotic standard error ranging from 7-20% of the mean values of the MSL estimator and from 10-22% of the mean values of the CML estimator.[10] The magnitudes of the asymptotic standard error values are certainly lower in the MSL (mean asymptotic standard error of 0.11) compared to the CML (mean asymptotic standard error of 0.16). However, note also that this direct comparison of the asymptotic standard error values is somewhat deceptive, because of the underestimation (overestimation) in recovering the true values of the parameters in the MSL (CML). This translates to consistently lower values of the asymptotic standard error estimates from the MSL approach relative to the CML approach. Further, one also needs to consider simulation error in the MSL estimation. These simulation standard errors for the MSL method are lower than the asymptotic standard errors, but not an insignificant fraction of the asymptotic standard errors. In particular, the simulation standard

[10] We do not include the first parameter ψ^1 in computing these ranges, because the true value of this parameter is zero, and the mean estimate is also very close to zero. Thus, percentages taken with respect to the mean estimate will be very high.

errors vary from 39%-58% of the asymptotic standard errors. Of course, even though the simulation errors are quite high as a percentage of asymptotic standard errors, the simulation adjusted standard errors in the table do not increase too much relative to the asymptotic standard errors because of the higher values of the asymptotic standard errors. Another observation from the standard error estimates is that these estimates (as a percentage of the mean estimates) are generally higher for the Cholesky parameters relative to the other parameters, reinforcing the finding earlier that the Choleski parameters are more difficult to recover than other parameters.

The final column of Table 1a provides a relative efficiency factor between the MSL and CML approaches. In particular, this column provides the ratio of the simulation-adjusted standard error of parameters from the MSL approach and the asymptotic standard error of parameters from the CML approach. As indicated earlier, the CML approach should lose some efficiency relative to the full maximum likelihood (ML) approach, because the CML approach compounds pairs of observations from the same individual, and does not consider all the panel observations from the same individual simultaneously. Theoretically speaking, therefore, the difference between the asymptotic covariance matrix of the CML estimator (obtained as the inverse of the Godambe matrix) and of the ML estimator (obtained as the inverse of the cross-product matrix of derivatives) should be positive semi-definite. However, note that the procedure being used in the current paper is the maximum simulated likelihood (MSL) approach, not the ML approach. In the MSL approach, the asymptotic covariance is computed as the inverse of the sandwich information matrix.[11] Basically, the presence of simulation noise, even if very small in the estimates of the parameters as in our case, can lead to a significant drop in the amount of information available in the sandwich matrix, resulting in increased standard errors of parameters when using MSL. Besides, one has to contend with the simulation error too introduced by the MSL, while the CML is simulation-free. Thus, one does not know *a priori* whether the MSL estimator will be more efficient than the CML estimator, and, if so, by how much. No theoretical results are derivable, and one has to consider this as an empirical issue. The ratio of the

[11] McFadden and Train (2000) indicate, in their use of independent number of random draws across observations, that the difference between the asymptotic covariance matrix of the MSL estimator obtained as the inverse of the sandwich information matrix and the asymptotic covariance matrix of the MSL estimator obtained as the inverse of the cross-product of first derivatives should be positive definite for finite number of draws per observation. Consequently, for the case of independent random draws across observations, the relationship between the MSL sandwich covariance matrix estimator and the CML Godambe covariance matrix is unclear. The situation gets even more unclear in our case because of the use of Halton or Lattice point draws that are not based on independent random draws across observations.

simulation adjusted standard errors from the MSL and the asymptotic standard errors from the CML (*i.e.*, the values in the last columns of the tables) provide an empirical estimate of the relative efficiency of the CML compared to the MSL. Relative efficiency values lower than one indicate a lower CML efficiency relative to the MSL, while values higher than one indicate that the CML is more efficient than the MSL. The results in the final column do indicate that the efficiency of the CML approach is about 63-87% (mean of 73%) of the MSL approach for this simple random coefficients case with a diagonal covariance matrix and no autoregressive error.

The time to convergence for the MSL estimation has a mean value of 7.55 minutes with a standard deviation of about 3 minutes. On the other hand, the time to convergence for the CML estimation has a mean value of 0.88 minutes with a standard deviation of about 0.1 minutes. This indicates that the CML method is about six times faster than the MSL estimation. Further, note that the CML method is actually more effective than suggested by this factor of six, because it produces more accurate estimates than the MSL estimates. In fact, it took about 250 Halton draws per individual to reach about the same level of mean APB value as for the CML approach, and the mean time for convergence with 250 Halton draws is about 7.25 minutes, suggesting a time efficiency factor of over 8 for the CML method relative to the MSL method. Also, while the APB improved as we moved from 150 Halton draws to 250 Halton draws in the MSL approach, there was surprisingly little change in the simulation error. Further, with 250 draws, the asymptotic standard errors increased because there was less underestimation in recovering parameters, with the net result that the relative efficiency of the CML actually marginally improved relative to the MSL method. One other problem we found even in this simple random coefficients MSL estimation was that 62 of the 500 runs did not converge (that is, 11.6% of total runs did not converge). This confirms that, even for the low to moderate dimensions of integration, numerical simulators can lead to numerical instability and convergence problems. On the other hand, no convergence issues whatsoever were encountered with the CML estimation.

Overall, for the specific case of the panel random-coefficients ordered-response model (with no autoregressive error structure), the results here indicate that the MSL estimator is more efficient than the CML estimator, but also that the CML estimator has a computational cost efficiency gain by a factor of about 8. Of course, the CML has the advantage of reproducibility of results, since it is simulation-free (and, as indicated earlier, the simulation errors are not

insignificant). It may be expected that, as the number of random coefficients increase, the econometric efficiency gains of the MSL will slip and the computational efficiency gains of the CML will increase. Besides, with large data sets and several specifications to potentially test, even a computational efficiency gain of 8 can be quite substantial. Of course, as we will see in the rest of this paper, as soon as one introduces more realistic and flexible specifications (such as non-diagonal random coefficients and autoregressive error structures,), the MSL estimation approach shows little to no econometric efficiency gains over the CML approach and/or literally become infeasible in practice from a computation cost standpoint.

The Non-Diagonal Covariance Matrix Case

The results in Table 1b provide information on the true values, the mean estimates, and the standard errors for the threshold parameters (the ψ parameters), the mean values of the distribution of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the b values in the table) and the Cholesky-decomposed parameters characterizing the covariance matrix of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the l values in the table).

As in the diagonal case, both the MSL and CML approaches do very well in recovering the parameters, with the APB values ranging from 0.2% to 30.5% (mean of 5.33%) for the MSL and from 0.12% to 29% (mean of 5.83%) for the CML approach. Though these are very good recovery statistics, the MSL method, in general, continues to under-estimate the magnitudes of parameters, while the CML method over-estimates the magnitudes of parameters. Also, similar to the diagonal case, there is more stability in the APB values across the ψ and b parameters than for the Choleski parameters (the l values), with the APB values for some of the Choleski elements being rather high. However, these high APB values are also somewhat deceptive, because the estimated values of the Choleski parameters are not too far away from the true values. But the small magnitudes of the true Choleski parameter values tend to inflate the APB values. For instance, the highest APB of 29% for the CML method is for the l_{42} parameter, even though the estimated value of -0.1862 is not far from the true value of -0.1443. Also, with a limited sample size and several Cholesky parameters to estimate, one would only expect a little more difficulty in accurately and precisely recovering the Cholesky parameters with finite samples. This is also noticeable in the asymptotic standard errors that are in the order of about

10% of the mean estimates for the ψ and b parameters (in both the MSL and CML cases), but much higher for the l parameters. Between the MSL and the CML estimators, the asymptotic standard errors are more similar in this non-diagonal case, with the mean standard error being 0.14 in the MSL case and 0.17 in the CML case. This is because the MSL provides estimates that are closer to the true values, and to the values from the CML estimation. However, the MSL estimator has very large simulation errors in this non-diagonal case, especially for the Cholesky elements, with one of these simulation standard errors being as high as 110% of the asymptotic standard error. The mean simulation standard error as a percentage of the asymptotic standard error is 67%. The net result is that the simulation-adjusted standard errors from the MSL approach are about the same magnitude as the CML standard errors. This is also reflected in the final column, where the relative efficiency values are close to 1, with some values being higher than 1. The mean relative efficiency value is 1.00, indicating that the CML estimator is as efficient as the MSL estimator.[12]

The mean computational time for the MSL method is about 18 minutes (with a standard deviation of 9.53 minutes) compared to 2.96 minutes (with a standard deviation of 0.88 minutes) for the CML method. So, the CML method is again about 6 times faster compared to the MSL method, for about the same level of accuracy in recovering parameters. At the same time, the econometric efficiency of the CML estimator is as good as the MSL estimator. Besides, 76 of the 500 runs did not converge in the MSL approach, with no such problems with the CML approach.

5.2 RCA Structure

As discussed earlier, the simulations for the case of random coefficients with an autoregressive error structure (*i.e.*, the RCA structure) is undertaken with only 20 datasets. Further, we did not undertake multiple runs for each of the 20 datasets because of the extremely high computational costs of doing so. The simulation errors were estimated based on those obtained from the RC structure. Tables 2a and 2b provide the results for the RCA non-diagonal case with low and high auto-correlations respectively.

[12] We also examined the results with 250 Halton draws per individual. However, there was little reduction in the simulation errors. To be specific, the mean simulation error as a percentage of the asymptotic standard error was still about 57% with 250 draws compared to 67% with 150 draws. The mean relative efficiency turned out to be 0.97 with 250 draws, indicating that the CML is pretty much as efficient as the MSL even if the number of draws is increased. Further, the APB values did not improve very substantially -- the mean APB value turned out to be 3.28% with 250 draws compared to 5.33% with 150 draws.

Low Autoregressive Correlation Case

The mean APB values are about the same in the MSL and the CML methods, with the APB ranging from 0.16% to 19.74% (mean of 5.47%) for the MSL and ranging from 0.17% to 23.3% (mean of 6.85%) for the CML. The parameters, including the ρ correlation parameter, are recovered well in both the MSL and CML cases. The autoregressive correlation parameter ρ is estimated reasonably well both in MSL and CML methods with APB values of 4.67% and 7.21%, respectively.

The estimated simulation-adjusted standard error for the MSL approach (mean value of 0.18) and the asymptotic standard errors from the CML approach (mean value of 0.21) are quite close to one another. The relative efficiency values (last column of table) range from 80% to 118%, with a mean value of 91%. This indicates relatively little overall econometric efficiency loss in using the CML approach relative to the MSL. Note also that because we had to estimate the simulation errors based on the results from the random coefficients (RC) case, the relative efficiency values in Table 2 may be considered as lower bounds, implying that the mean relative efficiency value is 91% or higher. At the same time, the mean time to convergence is 286.4 minutes or about 5 hours (standard deviation of 34.7 minutes) for the MSL compared to only about 3.80 minutes for the CML method (with a standard deviation of 1.24 minutes). This is a phenomenal computation efficiency leap, with the CML method being about 75 times faster than the MSL method. Moreover, only 90% of the MSL runs converged compared to a 100% convergence rate for the CML method.

High Autoregressive Correlation Case

The APB for the MSL approach ranges from 0.48% to 8.52% (mean of 3.82%), while the CML APB ranges from 0.93% to 17.95% (mean of 6.75%). While both of these APBs are of about the same order, the MSL approach does provide a marginally better APB in this high correlation case. However, in general, the ability to recover parameters does not seem to be affected at all by whether there is low correlation or high correlation. With specific regard to the autoregressive correlation parameter (ρ), the result indicates that the CML approach (but not the MSL approach) recovers this parameter very well in the high correlation case relative to the low correlation case (a difference of 0.0065 between the mean estimate of the correlation and the true value with an associated APB of 0.93% in the high correlation case, compared to a

corresponding difference of 0.1140 and an associated 7.21% in the low correlation case). This is perhaps because the correlation parameter needs to be particularly strong before it starts having any substantial effects on the log-likelihood function value. Essentially, the log-likelihood function can be relatively flat at low correlation, leading to more difficulty in accurately recovering the low correlation parameter. But, at a high correlation level, the log-likelihood function shifts considerably in value with small shifts in the correlation value, allowing it to be recovered accurately. Why this does not play out in the MSL case is an open question. It is possible that, at high correlations, there is considerable instability in the search direction and the convergence process in the MSL approach. This is reinforced by the fact that, while 18 of the 20 runs converged (90% convergence rate) in the MSL approach in the low correlation case, only 11 of the 20 runs converged (55% convergence rate) in the MSL approach in the high correlation case.[13]

The relative efficiency values in the last column reveal that there is an even lower efficiency loss with the CML approach relative to the MSL approach in this high correlation case compared to the low correlation case. The MSL runs that converged have a mean time to convergence of 252.71 minutes (more than 4.5 hours) relative to only 3.55 minutes for the CML method; that is, the CML method is about 71 times faster than the MSL method. Note also that there is a huge standard deviation in the time to convergence of the MSL method, which is consistent with the convergence-related instability problems discussed earlier. In cases with more than five random coefficients and more general auto-regressive structures than the simple first order structure considered in this study, the convergence problems and the high computational times of the MSL make it literally infeasible. The CML method, on the other hand, should be able to accommodate such panel structures with relative ease.

6. SUMMARY AND CONCLUSIONS

This paper focuses on panel ordered-response model structures, and compares the performance of the maximum simulated likelihood (MSL) estimation approach with that of the composite marginal likelihood (CML) estimation approach. The panel structures considered in the paper

[13] We undertook an independent simulation exercise to see if the same convergence problems held up with a diagonal instead of a non-diagonal covariance structure for the random coefficients in this RCA case. The results showed the same trend as in the non-diagonal case discussed here, with the convergence rates being 95% for the low correlation case compared to only 65% for the high correlation case. Thus, the convergence problem seems quite definitely associated with the correlation level.

include the pure random coefficients (RC) model with no autoregressive error component, as well as the more general case of random coefficients combined with an autoregressive error component. The ability of the MSL and CML approaches to recover the true parameters is examined using simulated datasets.

Overall, the results suggest that the CML method is able to recover the true parameters in all the cases considered in the paper, irrespective of the type of covariance matrix (diagonal versus non-diagonal) of the random coefficients and the level of the autoregressive correlation (low versus high). In fact, the performance of the MSL approach with 150 randomized and scrambled Halton draws and the simulation-free CML approach are about the same order in all cases in terms of the absolute percentage bias (APB) of the parameters estimated. The MSL approach has an edge in terms of econometric efficiency in the context of the pure RC model with a diagonal covariance matrix for the random coefficients, but is also riddled with convergence problems even in this simple panel case. The CML approach does not exhibit any convergence problems, and is about eight times faster. Any econometric efficiency gains of the MSL approach vanishes as soon as a non-diagonal covariance matrix of the random coefficients is introduced, due to the presence of large simulation noise in the MSL runs. This continues to hold when an autoregressive error structure is added to the non-diagonal random coefficients structure. Increasing the number of randomized/scrambled Halton draws from 150 per individual to 250 per individual has little impact on these results. Further, when an autoregressive error structure is added, the convergence times of the MSL runs start getting very high, while there is little to no effect on the convergence times for the CML runs. In fact, the CML method is about 70-75 times faster than the MSL method when there is an autoregressive error component. The MSL approach breaks down in particular when the autoregressive error correlation magnitude is high, with a substantial number of runs failing to converge. At the same time, there is literally no loss in econometric efficiency of the CML estimator compared to the MSL estimator.

Future research efforts should consider varying numbers of random coefficients, autoregressive error structures of order higher than one, and varying numbers of observations per individual. However, the results in this paper paint a very encouraging picture for the use of the CML pairwise likelihood method for the quick, accurate, and practical estimation of panel ordered-response models with flexible and rich stochastic specifications.

ACKNOWLEDGEMENTS

This research was partially funded by a Southwest Region University Transportation Center grant. The authors are grateful to Lisa Macias for her help in typesetting and formatting this document.

REFERENCES

Beron, K.J., Vijverberg, W.P.M., 2004. Probit in a spatial context: A Monte Carlo analysis. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), Springer-Verlag, Berlin, 169-196.

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., Sardesai, R., 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B* 40(9), 709-730.

Bhat, C.R., Zhao, H., 2002. The spatial analysis of activity stop generation. *Transportation Research Part B* 36(6), 557-575.

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

Bhat, C.R., Sener, I.N., Eluru, N., 2010b. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* 44(8-9), 903-921.

Butler, J.S., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* 50(3), 761-764.

Cox, D.R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3), 729-737.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention* 40(3), 1033-1054.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computer Science and Statistics: Proceedings of the Twenty Third Symposium on the Interface*, 571-578, Foundation of North America Inc., Fairfax.

Genz, A., Bretz, F., 2002. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* 11(4), 950-971.

Genz, A., Bretz, F., 1999. Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* 63(4), 361-378.

Girard, P., Parent, E., 2001. Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. *Technometrics* 43(2), 180-191.

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208-1211.

Greene, W.H., 2000. *Econometric Analysis* (Fourth ed.), Prentice-Hall Inc., New Jersey.

Greene, W.H., 2005. Censored data and truncated distributions. In Mills, T.C., Patterson, K. (eds.), *Palgrave Handbook of Econometrics*, Volume 1: Econometric Theory, Palgrave Macmillan, Hampshire.

Greene, W.H., Hensher, D.A., 2009. Ordered choice, heterogeneity, and attribute processing. *Journal of Transport Economics and Policy* (in press).

Greene, W.H., Hensher, D.A., 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.

Hajivassiliou, V., McFadden, D. 1998. The method of simulated scores for the estimation of LDV models. *Econometrica* 66(4), 863-896.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik* 2, 84-90.

Hjort, N.L., Varin, C., 2008. ML, PL, QL in Markov chain models. *Scandinavian Journal of Statistics* 35(1), 64-82.

Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis* 100(4), 670-685.

Keane, M., 1994. A computationally practical simulation estimator for panel data. *Econometrica*, 62(1), 95-116.

Kuk, A.Y.C., Nott, D.J., 2000. A pairwise likelihood approach to analyzing correlated binary data. *Statistics and. Probability Letters* 47(4), 329-335.

Le Cessie, S., Van Houwelingen, J.C., 1994. Logistic regression for correlated binary data. *Applied. Statistics* 43(1), 95-108.

Lele, S.R., 2006. Sampling variability and estimates of density dependence: a composite-likelihood approach. *Ecology* 87(1), 189-202.

LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32(1), 19-35.

Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Liu, I., Agresti, A., 2005. The analysis of ordered categorical data: An overview and a survey of recent developments. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research* 14(1), 1-73.

Mardia, K.V., Kent, J.T., Hughes, G., Taylor, C.C., 2009. Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika* 96(4), 975-982.

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

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4(summer), 103-120.

Mi, X., Miwa, T., Hothorn, T., 2009. mvtnorm: new numerical algorithm for multivariate normal probabilities. *The R Journal* 1(1), 37-39.

Müller, G., Czado, C. 2005. An autoregressive ordered probit model with application to high frequency financial data. *Journal of Computational and Graphical Statistics* 14(2), 320-338.

Renard, D., Molenberghs, G., Geys, H., 2004. A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis* 44(4), 649-667.

Train, K., 2003. *Discrete Choice Methods with Simulation*. 1st edition, Cambridge University Press, Cambridge.

Tuffin, B., 1996. On the use of low discrepancy sequences in Monte Carlo methods. *Monte Carlo Methods and Applications* 2(4), 295-320.

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

Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11(1), 127-138.

Varin, C., Reid, N., Firth, D, 2010. An overview of composite marginal likelihoods. Working Paper 09-45, Centre for Research in Statistical Methodology (CRiSM), University of Warwick, U.K.

Winship, C., Mare, R.D., 1984. Regression models with ordinal variables. *American Sociological Review* 49(4), 512-525.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3), 335-356.

LIST OF TABLES

Table 1a Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches –Diagonal Case without Auto-Correlation Random Coefficients (RC) Structure

Parameter	True Value	MSL					CML Approach			Relative Efficiency
		Parameter Estimates		Standard Error Estimates			Parameter Estimates			
		Mean Estimate	Absolute Percentage Bias (*APB*)*	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias (*APB*)*	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{SASE_{MSL}}{MASE_{CML}}$
ψ_1	0.0000	-0.0105	1.05	0.0562	0.0086	0.0569	-0.0071	0.71	0.0651	0.87
ψ_2	1.0000	0.9305	6.95	0.0790	0.0308	0.0848	1.0522	5.22	0.1217	0.70
ψ_3	2.0000	1.8716	6.42	0.1228	0.0601	0.1367	2.1123	5.62	0.2153	0.63
b_1	1.5000	1.3970	6.87	0.1109	0.0506	0.1219	1.5949	6.33	0.1811	0.67
b_2	1.0000	0.9338	6.62	0.0899	0.0388	0.0979	1.0683	6.83	0.1345	0.73
b_3	2.0000	1.8484	7.58	0.1309	0.0664	0.1468	2.1096	5.48	0.2287	0.64
b_4	1.0000	0.9360	6.40	0.0844	0.0339	0.0909	1.0671	6.71	0.1294	0.70
b_5	2.0000	1.8430	7.85	0.1350	0.0647	0.1497	2.0889	4.45	0.2291	0.65
l_{11}	1.0000	0.9247	7.53	0.1125	0.0520	0.1239	1.0705	7.05	0.1641	0.76
l_{22}	1.0000	0.9164	8.36	0.1107	0.0467	0.1201	1.0587	5.87	0.1649	0.73
l_{33}	0.7071	0.5934	16.08	0.1209	0.0708	0.1401	0.7151	1.13	0.1616	0.87
l_{44}	0.7071	0.6440	8.93	0.1067	0.0453	0.1159	0.7799	10.30	0.1469	0.79
l_{55}	1.0000	0.9252	7.48	0.1166	0.0534	0.1282	1.0596	5.96	0.1694	0.76
Overall mean value across parameters			7.55	0.11	0.05	0.12		5.51	0.16	0.73
Mean Time		5.46					0.88			
Std. Dev. Of Time		2.79					0.11			
% of Runs Converged		87.6%					100%			

* If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1in the denominator, and multiplying by 100.

Table 1b Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches –Non-Diagonal Case without Auto-Correlation

Random Coefficients (RC) Structure

Parameter	True Value	MSL					CML Approach			Relative Efficiency
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Asymptotic standard error ($MASE_{CML}$)	
		Mean Estimate	Absolute Percentage Bias *(APB)**	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute percentage bias *(APB)**		$\frac{SASE_{MSL}}{MASE_{CML}}$
ψ_1	0.0000	-0.0020	0.20	0.0583	0.0103	0.0592	-0.0044	0.44	0.0646	0.92
ψ_2	1.0000	0.9771	2.28	0.0896	0.0336	0.0957	1.0582	5.82	0.1176	0.81
ψ_3	2.0000	1.9521	2.39	0.1473	0.0656	0.1613	2.1142	5.71	0.2118	0.76
b_1	1.5000	1.4598	2.68	0.1318	0.0593	0.1445	1.5948	6.32	0.1749	0.83
b_2	1.0000	0.9917	0.83	0.1068	0.0455	0.1161	1.0817	8.17	0.1372	0.84
b_3	2.0000	1.9456	2.72	0.1666	0.0739	0.1822	2.1153	5.77	0.2251	0.81
b_4	1.0000	0.9690	3.10	0.1065	0.0411	0.1141	1.0563	5.63	0.1331	0.86
b_5	2.0000	1.9385	3.07	0.1688	0.0738	0.1843	2.0897	4.49	0.2257	0.82
l_{11}	1.0000	0.9798	2.02	0.1317	0.0615	0.1453	1.0867	8.67	0.1594	0.91
l_{21}	-0.5000	-0.4860	2.79	0.1275	0.0914	0.1568	-0.4987	0.27	0.1452	1.08
l_{22}	0.8660	0.8176	5.58	0.1442	0.0995	0.1752	0.9265	6.98	0.1754	0.99
l_{31}	0.2500	0.2543	1.72	0.1280	0.0934	0.1585	0.2219	11.24	0.1488	1.06
l_{32}	0.4330	0.4829	11.52	0.1484	0.1209	0.1915	0.4497	3.85	0.1769	1.08
l_{33}	0.8660	0.7423	14.29	0.1491	0.1272	0.1960	0.8812	1.75	0.1746	1.12
l_{41}	0.7500	0.7298	2.69	0.1172	0.0707	0.1369	0.7503	0.04	0.1391	0.98
l_{42}	-0.1443	-0.1298	10.09	0.1325	0.1142	0.1749	-0.1862	28.98	0.1724	1.01
l_{43}	0.2367	0.2446	3.35	0.1405	0.1296	0.1912	0.2692	13.73	0.1674	1.14
l_{44}	0.6005	0.4174	30.50	0.1610	0.1768	0.2391	0.5851	2.56	0.1914	1.24
l_{51}	0.0000	0.0251	2.51	0.1275	0.0893	0.1556	-0.0013	0.19	0.1444	1.08
l_{52}	0.0000	0.0087	0.87	0.1444	0.1232	0.1898	-0.0129	1.29	0.1728	1.10
l_{53}	0.0000	-0.0012	0.12	0.1564	0.1546	0.2199	-0.0159	1.59	0.1986	1.11
l_{54}	0.0000	0.0206	2.06	0.1857	0.2848	0.3400	0.0153	1.53	0.2663	1.28
l_{55}	1.0000	0.8477	15.22	0.1580	0.1674	0.2302	0.9088	9.12	0.2068	1.11
Overall mean value across parameters			5.33	0.14	0.10	0.17		5.83	0.17	1.00
Mean Time		18.09					2.96			
Std. Dev. Of Time		9.53					0.88			
% of Runs Converged		84.8%					100%			

*If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1in the denominator, and multiplying by 100.

Table 2a Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches –Non-Diagonal Case with Low Auto-Correlation Random Coefficients Autoregressive (RCA) Structure

Parameter Value	True Value	MSL Approach					CML Approach			Relative Efficiency
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Asymptotic Standard Error ($MASE_{CML}$)	$\frac{SASE_{MSL}}{MASE_{CML}}$
		Mean Estimate	Absolute Percentage Bias *(APB)**	Asymptotic Standard Error ($MASE_{MSL}$)	Estimated Simulation Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias *(APB)**		
ψ_1	0.0000	0.0016	0.16	0.0631	0.0440	0.0769	-0.0017	0.17	0.0669	1.15
ψ_2	1.0000	1.0435	4.35	0.0972	0.0678	0.1185	1.0639	6.39	0.1290	0.92
ψ_3	2.0000	2.0628	3.14	0.1580	0.1102	0.1926	2.1092	5.46	0.2282	0.84
b_1	1.5000	1.5276	1.84	0.1399	0.0975	0.1706	1.5649	4.33	0.1908	0.89
b_2	1.0000	1.0657	6.57	0.1141	0.0795	0.1391	1.0813	8.13	0.1487	0.93
b_3	2.0000	2.0568	2.84	0.1811	0.1262	0.2208	2.1047	5.23	0.2466	0.90
b_4	1.0000	1.0369	3.69	0.1118	0.0779	0.1362	1.0555	5.55	0.1470	0.93
b_5	2.0000	2.0581	2.90	0.1791	0.1248	0.2183	2.0794	3.97	0.2482	0.88
l_{11}	1.0000	1.0487	4.87	0.1281	0.0893	0.1561	1.0945	9.45	0.1734	0.90
l_{21}	-0.5000	-0.5314	6.27	0.1315	0.0917	0.1603	-0.5555	11.10	0.1550	1.03
l_{22}	0.8660	0.9250	6.80	0.1428	0.0995	0.1741	0.9205	6.29	0.2097	0.83
l_{31}	0.2500	0.2084	16.63	0.1283	0.0894	0.1564	0.2186	12.54	0.1508	1.04
l_{32}	0.4330	0.5117	18.16	0.1478	0.1030	0.1801	0.5340	23.33	0.2263	0.80
l_{33}	0.8660	0.8704	0.50	0.1494	0.1041	0.1821	0.8912	2.91	0.2517	0.72
l_{41}	0.7500	0.7482	0.25	0.1165	0.0812	0.1420	0.7421	1.05	0.1444	0.98
l_{42}	-0.1443	-0.1159	19.74	0.1446	0.1008	0.1762	-0.1336	7.45	0.1713	1.03
l_{43}	0.2367	0.2446	3.34	0.1409	0.0982	0.1717	0.2536	7.13	0.1797	0.96
l_{44}	0.6005	0.5507	8.30	0.1475	0.1028	0.1798	0.5371	10.56	0.1950	0.92
l_{51}	0.0000	0.0189	1.89	0.1459	0.1017	0.1778	0.0139	1.39	0.1505	1.18
l_{52}	0.0000	0.0366	3.66	0.1857	0.1294	0.2264	0.0230	2.30	0.2321	0.98
l_{53}	0.0000	-0.0460	4.60	0.2411	0.1680	0.2939	0.0022	0.22	0.3648	0.81
l_{54}	0.0000	0.0568	5.68	0.2708	0.1888	0.3302	0.0408	4.08	0.3520	0.94
l_{55}	1.0000	0.9958	0.42	0.1628	0.1135	0.1984	0.8187	18.13	0.5906	0.34
ρ	0.3000	0.2860	4.67	0.1117	0.0779	0.1362	0.2784	7.21	0.1308	1.04
Overall mean value across parameters			5.47	0.15	0.10	0.18		6.85	0.21	0.91
Mean Time		286.40					3.80			
Std. Dev. Of Time		34.73					1.24			
% of Runs Converged		90%					100%			

*If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1in the denominator, and multiplying by 100.

Table 2b Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches –Non-Diagonal Case with High Auto-Correlation Random Coefficients Autoregressive (RCA) Structure

Parameter Value	True Value	MSL Approach					CML Approach			Relative Efficiency
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Asymptotic standard error ($MASE_{CML}$)	$\frac{SASE_{MSL}}{MASE_{CML}}$
		Mean Estimate	Absolute percentage bias *(APB)**	Asymptotic standard error ($MASE_{MSL}$)	Estimated Simulation error	Simulation Adjusted standard error ($SASE_{MSL}$)	Mean Estimate	Absolute percentage bias *(APB)**		
ψ_1	0.0000	0.0382	3.82	0.0676	0.0471	0.0824	0.0130	1.30	0.0722	1.14
ψ_2	1.0000	1.0627	6.27	0.0990	0.0690	0.1206	1.0704	7.04	0.1212	1.00
ψ_3	2.0000	2.0861	4.30	0.1569	0.1094	0.1913	2.1011	5.06	0.2054	0.93
b_1	1.5000	1.5535	3.57	0.1394	0.0972	0.1700	1.5819	5.46	0.1764	0.96
b_2	1.0000	1.0659	6.59	0.1095	0.0763	0.1335	1.0724	7.24	0.1334	1.00
b_3	2.0000	2.0482	2.41	0.1741	0.1214	0.2122	2.0963	4.82	0.2243	0.95
b_4	1.0000	1.0386	3.86	0.1103	0.0769	0.1345	1.0644	6.44	0.1356	0.99
b_5	2.0000	2.0582	2.91	0.1764	0.1229	0.2150	2.0914	4.57	0.2263	0.95
l_{11}	1.0000	1.0752	7.52	0.1269	0.0885	0.1547	1.1077	10.77	0.1611	0.96
l_{21}	-0.5000	-0.4739	5.23	0.1225	0.0854	0.1494	-0.5528	10.55	0.1484	1.01
l_{22}	0.8660	0.8983	3.73	0.1287	0.0897	0.1569	0.9058	4.60	0.1713	0.92
l_{31}	0.2500	0.2462	1.52	0.1238	0.0863	0.1509	0.2239	10.44	0.1479	1.02
l_{32}	0.4330	0.4556	5.20	0.1417	0.0988	0.1727	0.5076	17.24	0.1809	0.95
l_{33}	0.8660	0.8767	1.23	0.1350	0.0941	0.1646	0.8812	1.76	0.1810	0.91
l_{41}	0.7500	0.7536	0.48	0.1137	0.0792	0.1385	0.7588	1.18	0.1375	1.01
l_{42}	-0.1443	-0.1320	8.52	0.1312	0.0914	0.1599	-0.1203	16.64	0.1699	0.94
l_{43}	0.2367	0.2469	4.28	0.1302	0.0908	0.1588	0.2792	17.95	0.1696	0.94
l_{44}	0.6005	0.5770	3.92	0.1384	0.0965	0.1687	0.5504	8.35	0.1900	0.89
l_{51}	0.0000	0.0740	3.70	0.1254	0.0874	0.1528	0.0546	5.46	0.1478	1.03
l_{52}	0.0000	0.0412	2.06	0.1567	0.1092	0.1910	0.0059	0.59	0.1859	1.03
l_{53}	0.0000	-0.0448	2.24	0.1673	0.1166	0.2039	-0.0467	4.67	0.2092	0.97
l_{54}	0.0000	0.0323	1.62	0.2427	0.1692	0.2958	0.0209	2.09	0.3239	0.91
l_{55}	1.0000	0.9932	0.68	0.1522	0.1061	0.1855	0.9299	7.01	0.2812	0.66
ρ	0.7000	0.6577	6.05	0.0769	0.0536	0.0937	0.7065	0.93	0.0959	0.98
Overall mean value across parameters			3.82	0.14	0.09	0.16		6.75	0.17	0.96
Mean Time		252.71					3.55			
Std. Dev. Of Time		54.43					1.61			
% of Runs Converged		55%					100%			

*If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1in the denominator, and multiplying by 100.