A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered Response Model System

Chandra R. Bhat*
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu

Cristiano Varin
University Ca' Foscari
Department of Statistics
San Giobbe, Cannaregio, 873
30121 Venice, Italy
Phone: +39-041-2347439; Fax: +39-041-2347444
E-mail: sammy@unive.it

and

Nazneen Ferdous
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: nazneen.ferdous@mail.utexas.edu

*corresponding author

First Version: January 2010
Revised Version: March 2010

ABSTRACT

This paper compares the performance of the maximum-simulated likelihood (MSL) approach with the composite marginal likelihood (CML) approach in multivariate ordered-response situations. The ability of the two approaches to recover model parameters in simulated data sets is examined, as is the efficiency of estimated parameters and computational cost. Overall, the simulation results demonstrate the ability of the Composite Marginal Likelihood (CML) approach to recover the parameters very well in a 5-6 dimensional ordered-response choice model context. In addition, the CML recovers parameters as well as the MSL estimation approach in the simulation contexts used in the current study, while also doing so at a substantially reduced computational cost. Further, any reduction in the efficiency of the CML approach relative to the MSL approach is in the range of non-existent to small. When taken together with its conceptual and implementation simplicity, the CML approach appears to be a promising approach for the estimation of not only the multivariate ordered-response model considered here, but also for other analytically-intractable econometric models.

Keywords: Composite marginal likelihood, multivariate ordered-response model system, maximum simulated likelihood, pairwise marginal likelihood, statistical efficiency.

1. INTRODUCTION

Ordered response model systems are used when analyzing ordinal discrete outcome data that may be considered as manifestations of an underlying scale that is endowed with a natural ordering. Examples include 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), or count data (such as the number of trips made by a household, the number of episodes of physical activity pursued by an individual, and the number of cars owned by a household). In all of these situations, the observed outcome data may be considered as censored (or coarse) measurements of an underlying latent continuous random variable. The censoring mechanism is usually characterized as a partitioning or thresholding of the latent continuous variable into mutually exclusive (non-overlapping) intervals. The reader is referred to McKelvey and Zavoina (1971) 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. The reader is also referred to a forthcoming book by Greene and Hensher (2010) for a comprehensive history and treatment of the ordered-response model structure. These recent reviews indicate the abundance of applications of the ordered-response model in the sociological, biological, marketing, and transportation sciences, and the list of applications only continues to grow rapidly.

While the applications of the ordered response model are quite widespread, much of these are confined to the analysis of a single outcome, with a sprinkling of applications associated with two and three correlated ordered-response outcomes. Some very recent studies of two correlated ordered-response outcomes include Scotti (2006), Mitchell and Weale (2007), Scott and Axhausen (2006), and LaMondia and Bhat (2009).[1] The study by Scott and Kanaroglou (2002) represents an example of three correlated ordered-response outcomes. But the examination of more than two to three correlated outcomes is rare, mainly because the extension to an arbitrary number of correlated ordered-response outcomes entails, in the usual likelihood function approach, integration of dimensionality equal to the number of outcomes. On the other hand,

[1] The first three of these studies use the bivariate ordered-response probit (BORP) model in which the stochastic elements in the two ordered-response equations take a bivariate normal distribution, while the last study develops a more general and flexible copula-based bivariate ordered-response model that subsumes the BORP as but one special case.

there are many instances when interest may be centered around analyzing several ordered-response outcomes simultaneously, such as in the case of the number of episodes of each of several activities, or satisfaction levels associated with a related set of products/services, or multiple ratings measures regarding the state of health of an individual/organization (we will refer to such outcomes as cross-sectional multivariate ordered-response outcomes). There are also instances when the analyst may want to analyze time-series or panel data of ordered-response outcomes over time, and allow flexible forms of error correlations over these outcomes. For example, the focus of analysis may be to examine rainfall levels (measured in grouped categories) over time in each of several spatial regions, or individual stop-making behavior over multiple days in a week, or individual headache severity levels at different points in time (we will refer to such outcomes as panel multivariate ordered-response outcomes).

In the analysis of cross-sectional and panel ordered-response systems with more than three outcomes, the norm until very recently has been to apply numerical simulation techniques based on a maximum simulated likelihood (MSL) approach or a Bayesian inference approach. However, such simulation-based approaches become impractical in terms of computational time, or even infeasible, as the number of ordered-response outcomes increases. Even if feasible, the numerical simulation methods do get imprecise as the number of outcomes increase, leading to convergence problems during estimation. As a consequence, another approach that has seen some (though very limited) use recently is the 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. The CML method is easy to implement and has the advantage of reproducibility of results. Under usual regularity assumptions, the CML estimator is consistent and asymptotically normal distributed. The maximum CML estimator should lose some efficiency from a theoretical perspective relative to a full likelihood estimator, but this efficiency loss appears to be empirically minimal (see Zhao and Joe, 2005; Lele, 2006; Joe and

Lee, 2009).[2] Besides, the simulation estimation methods for evaluating the analytically intractable likelihood function also leads to a loss in estimator efficiency.

The objective of this paper is on introducing the CML inference approach to estimate general panel models of ordered-response. We also compare the performance of the maximum-simulated likelihood (MSL) approach with the composite marginal likelihood (CML) approach in ordered-response situations when the MSL approach is feasible. 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 the structures of the cross-sectional and panel multivariate ordered-response systems. Section 3 discusses the simulation estimation methods (with an emphasis on the MSL approach) and the CML estimation approach. Section 4 presents the experimental design for the simulation experiments, while Section 5 discusses the results. Section 6 concludes the paper by highlighting the important findings.

2. THE MULTIVARIATE ORDERED RESPONSE SYSTEM

2.1 The Cross-Sectional Multivariate Ordered-Response Probit (CMOP) Formulation

Let q be an index for individuals (q = 1, 2, …, Q, where Q denotes the total number of individuals in the data set), and let i be an index for the ordered-response variable (i = 1, 2, …, I, where I denotes the total number of ordered-response variables for each individual). Let the observed discrete (ordinal) level for individual q and variable i be m_{qi} (m_{qi} may take one of K_i values; *i.e.*, $m_{qi} \in \{1, 2, \ldots, K_i\}$ for variable i). In the usual ordered response framework notation, we write the latent propensity (y^*_{qi}) for each ordered-response variable as a function of relevant covariates and relate this latent propensity to the observed discrete level m_{qi} through threshold bounds (see McKelvey and Zavoina, 1975):

[2] 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 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 multivariate ordered-response system.

$$y_{qi}^* = \beta_i' x_{qi} + \varepsilon_{qi},\ y_{qi} = m_{qi} \text{ if } \theta_i^{m_{qi}-1} < y_{qi}^* < \theta_i^{m_{qi}}, \tag{1}$$

where x_{qi} is a $(L \times 1)$ vector of exogenous variables (not including a constant), β_i is a corresponding $(L \times 1)$ vector of coefficients to be estimated, ε_{qi} is a standard normal error term, and $\theta_i^{m_{qi}}$ is the upper bound threshold for discrete level m_{qi} of variable i ($\theta_i^0 < \theta_i^1 < \theta_i^2 ... < \theta_i^{K_i-1} < \theta_i^{K_i}$; $\theta_i^0 = -\infty$, $\theta_i^{K_i} = +\infty$ for each variable i). The ε_{qi} terms are assumed independent and identical across individuals (for each and all i). For identification reasons, the variance of each ε_{qi} term is normalized to 1. However, we allow correlation in the ε_{qi} terms across variables i for each individual q. Specifically, we define $\varepsilon_q = (\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \ldots, \varepsilon_{qI})'$. Then, ε_q is multivariate normal distributed with a mean vector of zeros and a correlation matrix as follows:

$$\varepsilon_q \sim N\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1I} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2I} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{I1} & \rho_{I2} & \rho_{I3} & \cdots & 1 \end{pmatrix}\right], \text{ or} \tag{2}$$

$$\varepsilon_q \sim N[\mathbf{0}, \mathbf{\Sigma}]$$

The off-diagonal terms of $\mathbf{\Sigma}$ capture the error covariance across the underlying latent continuous variables; that is, they capture the effects of common unobserved factors influencing the underlying latent propensities. These are the so-called polychoric correlations between pairs of observed ordered-response variables. Of course, if all the correlation parameters (*i.e.*, off-diagonal elements of $\mathbf{\Sigma}$), which we will stack into a vertical vector Ω, are identically zero, the model system in Equation (1) collapses to independent ordered response probit models for each variable. Note that the diagonal elements of $\mathbf{\Sigma}$ are normalized to one for identification purposes.

The parameter vector (to be estimated) of the cross-sectional multivariate probit model is $\delta = (\beta_1', \beta_2', ..., \beta_I';\ \theta_1', \theta_2', ..., \theta_I';\ \Omega')'$, where $\theta_i = (\theta_i^1, \theta_i^2, ..., \theta_i^{K_i-1})'$ for $i = 1, 2, ..., I$. The likelihood function for individual q may be written as follows:

$$L_q(\delta) = \Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI} = m_{qI}) \quad (3)$$

$$L_q(\delta) = \int_{v_1=\theta_1^{m_{q1}-1}-\beta_1' x_{q1}}^{\theta_1^{m_{q1}}-\beta_1' x_{q1}} \int_{v_2=\theta_2^{m_{q2}-1}-\beta_2' x_{q2}}^{\theta_2^{m_{q2}}-\beta_2' x_{q2}} \cdots \int_{v_I=\theta_I^{m_{qI}-1}-\beta_I' x_{qI}}^{\theta_I^{m_{qI}}-\beta_I' x_{qI}} \phi_I(v_1, v_2, ..., v_I \mid \Omega) dv_1 dv_2 ... dv_I,$$

where ϕ_I is the standard multivariate normal density function of dimension I. The likelihood function above involves an I-dimensional integral for each individual q.

2.2 The Panel Multivariate Ordered-Response Probit (PMOP) Formulation

Let q be an index for individuals as earlier (q = 1, 2, …, Q), but let j now be an index for the jth 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 jth observation be m_{qj} (m_{qj} may take one of K values; *i.e.*, $m_{qi} \in \{1, 2, …, K\}$). In the usual random-effects ordered response framework notation, we write the latent variable (y^*_{qj}) as a function of relevant covariates as:

$$y^*_{qj} = \beta' x_{qj} + u_q + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \theta^{m_{qj}-1} < y^*_{qj} < \theta^{m_{qj}}, \quad (4)$$

where x_{qj} is a $(L\times 1)$ vector of exogenous variables (not including a constant), β is a corresponding $(L\times 1)$ vector of coefficients to be estimated, ε_{qj} is a standard normal error term uncorrelated across observations j for individual q and also uncorrelated across individuals q, and $\theta^{m_{qj}}$ is the upper bound threshold for discrete level m_{qj} ($\theta^0 < \theta^1 < \theta^2 ... < \theta^{K-1} < \theta^K$; $\theta^0 = -\infty$, $\theta^K = +\infty$). The term u_q represents an individual-specific random term, assumed to be normally distributed with mean zero and variance σ^2. The term u_q is independent of $u_{q'}$ for $q \neq q'$. The net result of the specification above is that the joint

[3] In this paper, we assume 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. However, the efficiency of the composite marginal likelihood (CML) approach depends on the weights used for each individual in the case of varying number of observations across individuals (see Kuk and Nott, 2000; Joe and Lee, 2009 provide a recent discussion and propose new weighting techniques). But one can simply put a weight of one without any loss of generality for each individual in the case of equal number of panel observations for each individual. In our paper, the focus 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.

distribution of the latent variables $(y^*_{q1}, y^*_{q2}, ... y^*_{qJ})$ for the qth subject is multivariate normal with standardized mean vector $(\beta' x_{q1} / \mu, \beta' x_{q2} / \mu, ... \beta' x_{qJ} / \mu)$ and a correlation matrix with constant non-diagonal entries σ^2 / μ^2, where $\mu = \sqrt{1+\sigma^2}$.

The standard random-effects ordered-response model of Equation (4) allows easy estimation, since one can write the probability of the sequence of observed ordinal responses across the multiple observations on the same individual, conditional on u_q, as the product of standard ordered-response model probabilities, and then integrate the resulting probability over the range of normally distributed u_q values for each individual. This results in only a one-dimensional integral for each individual, which can be easily computed using numerical quadrature methods. However, the assumption of equal correlation across the multiple observations on the same individual is questionable, especially for medium-to-long individual-specific series. An alternative would be to allow serial correlation within each subject-specific series of observations, as proposed by Varin and Czado (2010). For instance, one may adopt an autoregressive structure of order one for the error terms of the same individual so that $corr(\varepsilon_{qj}, \varepsilon_{qk}) = \rho^{|t_{qj} - t_{qk}|}$, where t_{qj} is the measurement time of observation y_{qj}.[4] The autoregressive error structure specification results in a joint multivariate distribution of the latent variables $(y^*_{q1}, y^*_{q2}, ... y^*_{qJ})$ for the qth individual with standardized mean vector $(\beta' x_{q1} / \mu, \beta' x_{q2} / \mu, ... \beta' x_{qJ} / \mu)$ and a correlation matrix $\boldsymbol{R_q}$ with entries such that $corr(y^*_{qj}, y^*_{qg}) = (\sigma^2 + \rho^{|t_{qj} - t_{qg}|}) / \mu^2$, where $\mu = \sqrt{1+\sigma^2}$. The cost of the flexibility is paid dearly though in terms of computational difficulty in the likelihood estimator. Specifically, rather than a single dimension of integration, we now have an integral of dimension J for individual q. The parameter vector (to be estimated) of the panel multivariate probit model is $\delta = (\beta'; \theta^1, \theta^2, ..., \theta^{K-1}; \sigma, \rho)'$, and the likelihood for individual q becomes:

[4] Note that 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*. (2010) developed a specification where the correlation in physical activity between two individuals may be a function of several measures of spatial proximity and adjacency.

$$L_q(\delta) = \Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qJ} = m_{qJ})$$

$$L_q(\delta) = \int_{v_1=\alpha^{m_{q1}-1}}^{\alpha^{m_{q1}}} \int_{v_2=\alpha^{m_{q2}-1}}^{\alpha^{m_{q2}}} \cdots \int_{v_J=\alpha^{m_{qJ}-1}}^{\alpha^{m_{qJ}}} \phi_J(v_1, v_2, ..., v_J \mid R_q) dv_1 dv_2 ... dv_J \tag{5}$$

where $\alpha^{m_{qj}} = (\theta^{m_{qj}} - \beta' x_{qj}) / \mu$.

The likelihood function above entails a J-dimensional integral for each individual q. The above model is labeled as a mixed autoregressive ordinal probit model by Varin and Czado (2010).

3. OVERVIEW OF ESTIMATION APPROACHES

As indicated in Section 1, models that require integration of more than three dimensions in a multivariate ordered-response model are typically estimated using simulation approaches, though some recent studies have considered a composite marginal likelihood approach. Sections 3.1 and 3.2 provide an overview of each of these two approaches in turn.

3.1 Simulation Approaches

Two broad simulation approaches may be identified in the literature for multivariate ordered response modeling. One is based on a frequentist approach, while the other is based on a Bayesian approach. We provide an overview of these two approaches in the next two sections (Section 3.1.1 and Section 3.1.2), and then (in Section 3.1.3) discuss the specific simulation approaches used in the current paper for estimation of the multivariate ordered-response model systems.

3.1.1 The Frequentist Approach

In the context of a frequentist approach, Bhat and Srinivasan (2005) suggested a maximum simulated likelihood (MSL) method for evaluating the multi-dimensional integral in a cross-sectional multivariate ordered response model system, using quasi-Monte Carlo simulation methods proposed by Bhat (2001; 2003). In their approach, Bhat and Srinivasan (BS) partition the overall error term into one component that is independent across dimensions and another mixing component that generates the correlation across dimensions. The estimation proceeds by conditioning on the error components that cause correlation effects, writing the resulting conditional joint probability of the observed ordinal levels across the many dimensions for each individual, and then integrating out the mixing correlated error components. An important issue is to ensure that the covariance matrix of the mixing error terms remains in a correlation form (for identification reasons) and is positive definite, which BS maintain by writing the likelihood function in terms of the elements of the Cholesky decomposed-matrix of the correlation matrix of the mixing normally distributed elements and parameterizing the diagonal elements of the Cholesky matrix to guarantee unit values along the diagonal. Another alternative and related MSL method would be to consider the correlation across error terms directly without partitioning the error terms into two components. This corresponds to the formulation in Equations (1) and (2) of the current paper. Balia and Jones (2008) adopt such a formulation in their eight-dimensional multivariate probit model of lifestyles, morbidity, and mortality. They estimate their model using a Geweke-Hajivassiliou-Keane (GHK) simulator (the GHK simulator is discussed in more detail later in this paper). However, it is not clear how they accommodated the identification sufficiency condition that the covariance matrix be a correlation matrix and be positive definite. But one can use the GHK simulator combined with BS's approach to ensure unit elements along the diagonal of the covariance matrix. Yet another MSL method to approximate the multivariate rectangular (*i.e.*, truncated) normal probabilities in the likelihood functions of Equation (3) and (5) is based on the Genz-Bretz (GB) algorithm (also discussed in more detail later). In concept, all these MSL methods can be extended to any number of correlated ordered-response outcomes, but numerical stability, convergence, and precision problems start surfacing as the number of dimensions increase.

3.1.2 The Bayesian Approach

Chen and Dey (2000), Herriges *et al.* (2008), Jeliazkov *et al.* (2008), and Hasegawa (2010) have considered an alternate estimation approach for the multivariate ordered response system based on the posterior mode in an objective Bayesian approach. As in the frequentist case, a particular challenge in the Bayesian approach is to ensure that the covariance matrix of the parameters is in a correlation form, which is a sufficient condition for identification. Chen and Dey proposed a reparametization technique that involves a rescaling of the latent variables for each ordered-response variable by the reciprocal of the largest unknown threshold. Such an approach leads to an unrestricted covariance matrix of the re-scaled latent variables, allowing for the use of standard Markov Chain Monte Carlo (MCMC) techniques for estimation. In particular, the Bayesian approach is based on assuming prior distributions on the non-threshold parameters, reparameterizing the threshold parameters, imposing a standard conjugate prior on the reparameterized version of the error covariance matrix and a flat prior on the transformed threshold, obtaining an augmented posterior density using Baye's Theorem for the reparameterized model, and fitting the model using a Markov Chain Monte Carlo (MCMC) method. Unfortunately, the method remains cumbersome, requires extensive simulation, and is time-consuming. Further, convergence assessment becomes difficult as the number of dimensions increase. For example, Muller and Czado (2005) used a Bayesian approach for their panel multivariate ordered-response model, and found that the standard MCMC method exhibits bad convergence properties. They proposed a more sophisticated group move multigrid MCMC technique, but this only adds to the already cumbersome nature of the simulation approach. In this regard, both the MSL and the Bayesian approach are "brute force" simulation techniques that are not very straightforward to implement and can create convergence assessment problems.

3.1.3 Simulators Used in the Current Paper

In the current paper, we use the frequentist approach to compare simulation approaches with the composite marginal likelihood (CML) approach. Frequentist approaches are widely used in the literature, and are included in several software programs that are readily available. Within the frequentist approach, we test two MSL methods with the CML approach, just to have a comparison of more than one MSL method with the CML approach. The two MSL methods we select are among the most effective simulators for evaluating multivariate normal probabilities. Specifically, we consider the Geweke-Hajivassiliou-Keane (GHK) simulator for the CMOP model estimation in Equation (3), and the Genz-Bretz (GB) simulator for the PMOP model estimation in Equation (5).

3.1.3.1 The Geweke-Hajivassiliou-Keane Probability Simulator for the CMOP Model

The GHK is perhaps the most widely used probability simulator for integration of the multivariate normal density function, and is particularly well known in the context of the estimation of the multivariate unordered probit model. It is named after Geweke (1991), Hajivassiliou (Hajivassiliou and McFadden, 1998), and Keane (1990, 1994). Train (2003) provides an excellent and concise description of the GHK simulator in the context of the multivariate unordered probit model. In the current paper, we adapt the GHK simulator to the case of the multivariate ordered-response probit model.

The GHK simulator is based on directly approximating the probability of a multivariate rectangular region of the multivariate normal density distribution. To apply the simulator, we first write the likelihood function in Equation (3) as follows:

$$L_q(\delta) = \Pr(y_{q1} = m_{q1}) \Pr(y_{q2} = m_{q2} \mid y_{q1} = m_{q1}) \Pr(y_{q3} = m_{q3} \mid y_{q1} = m_{q1}, y_{q2} = m_{q2}) \ldots$$
$$\ldots \Pr(y_{qI} = m_{qI} \mid y_{q1} = m_{q1}, y_{q2} = m_{q2}, \ldots, y_{qI-1} = m_{qI-1}) \tag{6}$$

Also, write the error terms in Equation (2) as:

$$\begin{bmatrix} \varepsilon_{q1} \\ \varepsilon_{q2} \\ \vdots \\ \varepsilon_{qI} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{I1} & l_{I2} & l_{I3} & \cdots & l_{II} \end{bmatrix} \begin{bmatrix} v_{q1} \\ v_{q2} \\ \vdots \\ v_{qI} \end{bmatrix} \tag{7}$$

$$\varepsilon_q = \boldsymbol{L} v_q$$

where $\boldsymbol{L}$ is the lower triangular Cholesky decomposition of the correlation matrix $\mathbf{\Sigma}$, and v_q terms are independent and identically distributed standard normal deviates (*i.e.*, $v_q \sim N[\mathbf{0},\mathbf{I_I}]$). Each (unconditional/conditional) probability term in Equation (7) can be written as follows:

$$\Pr(y_{q1} = m_{q1}) = \Pr\left(\frac{\theta_1^{m_{q1}-1} - \beta_1' x_{q1}}{l_{11}} < v_{q1} < \frac{\theta_1^{m_{q1}} - \beta_1' x_{q1}}{l_{11}} \right)$$

$$\Pr(y_{q2} = m_{q2} \mid y_{q1} = m_{q1}) = \Pr\left(\frac{\theta_2^{m_{q2}-1} - \beta_2' x_{q2} - l_{21} v_{q1}}{l_{22}} < v_{q2} < \frac{\theta_2^{m_{q2}} - \beta_2' x_{q2} - l_{21} v_{q1}}{l_{22}} \middle| \frac{\theta_1^{m_{q1}-1} - \beta_1' x_{q1}}{l_{11}} < v_{q1} < \frac{\theta_1^{m_{q1}} - \beta_1' x_{q1}}{l_{11}} \right)$$

$$\Pr(y_{q3} = m_{q3} \mid y_{q1} = m_{q1}, y_{q2} = m_{q2}) = \Pr\left(\begin{array}{l} \dfrac{\theta_3^{m_{q3}-1} - \beta_3' x_{q3} - l_{31} v_{q1} - l_{32} v_{q2}}{l_{33}} < v_{q3} < \dfrac{\theta_3^{m_{q3}} - \beta_3' x_{q3} - l_{31} v_{q1} - l_{32} v_{q2}}{l_{33}} \\ \left| \dfrac{\theta_1^{m_{q1}-1} - \beta_1' x_{q1}}{l_{11}} < v_{q1} < \dfrac{\theta_1^{m_{q1}} - \beta_1' x_{q1}}{l_{11}}, \dfrac{\theta_2^{m_{q2}-1} - \beta_2' x_{q2} - l_{21} v_{q1}}{l_{22}} < v_{q2} < \dfrac{\theta_2^{m_{q2}} - \beta_2' x_{q2} - l_{21} v_{q1}}{l_{22}} \right. \end{array} \right)$$

$$\vdots \tag{8}$$

$$\Pr(y_{qI}=m_{qI}\,|\,y_{q1}=m_{q1},y_{q2}=m_{q2},...,y_{qI-1}=m_{qI-1})=$$

$$\Pr\left(\begin{array}{l}\dfrac{\theta_1^{m_{qI}-1}-\beta_I' x_{qI}-l_{I1}v_{q1}-l_{I2}v_{q2}-\cdots-l_{I(I-1)}v_{q(I-1)}}{l_{II}}<v_{qI}<\dfrac{\theta_1^{m_{qI}}-\beta_I' x_{qI}-l_{I1}v_{q1}-l_{I2}v_{q2}-\cdots-l_{I(I-1)}v_{q(I-1)}}{l_{II}}\\ \left|\dfrac{\theta_1^{m_{q1}-1}-\beta_1' x_{q1}}{l_{11}}<v_{q1}<\dfrac{\theta_1^{m_{q1}}-\beta_1' x_{q1}}{l_{11}},\dfrac{\theta_2^{m_{q2}-1}-\beta_2' x_{q2}-l_{21}v_{q1}}{l_{22}}<v_{q2}<\dfrac{\theta_2^{m_{q2}}-\beta_2' x_{q2}-l_{21}v_{q1}}{l_{22}},\cdots,\right.\\ \dfrac{\theta_{I-1}^{m_{q(I-1)}-1}-\beta_{I-1}' x_{qI-1}-l_{(I-1)1}v_{q1}-l_{(I-1)2}v_{q2}-\cdots-l_{(I-1)(I-2)}v_{q(I-2)}}{l_{(I-1)(I-1)}}<v_{q(I-1)}<\dfrac{\theta_{I-1}^{m_{q(I-1)}}-\beta_{I-1}' x_{qI-1}-l_{(I-1)1}v_{q1}-l_{(I-1)2}v_{q2}-\cdots-l_{(I-1)(I-2)}v_{q(I-2)}}{l_{(I-1)(I-1)}}\end{array}\right)$$

The error terms v_{qi} are drawn d times (d = 1, 2, …, D) from the univariate standard normal distribution with the lower and upper bounds as above. To be precise, we use a randomized Halton draw procedure to generate the d realizations of v_{qi}, where we first generate standard Halton draw sequences of size $D\times 1$ for each individual for each dimension i (i = 1, 2,…, I), and then randomly shift the $D\times 1$ integration nodes using a random draw from the uniform distribution (see Bhat, 2001 and 2003 for a detailed discussion of the use of Halton sequences for discrete choice models). These random shifts are employed because we generate 10 different randomized Halton sequences of size $D\times 1$ to compute simulation error. Gauss code implementing the Halton draw procedure is available for download from the home page of Chandra Bhat at http://www.caee.utexas.edu/prof/bhat/halton.html. For each randomized Halton sequence, the uniform deviates are translated to truncated draws from the normal distribution for v_{qi} that respect the lower and upper truncation points (see, for example, Train, 2003; page 210). An unbiased estimator of the likelihood function for individual q is obtained as:

$$L_{GHK,q}(\delta)=\frac{1}{D}\sum_{d=1}^{D}L_q^d(\delta) \tag{9}$$

where $L_q^d(\delta)$ is an estimate of Equation (6) for simulation draw d. A consistent and asymptotically normal distributed GHK estimator $\hat{\delta}_{GHK}$ is obtained by maximizing the logarithm of the simulated likelihood function $L_{GHK}(\delta)=\prod_q L_{GHK,q}(\delta)$. The covariance matrix of parameters is estimated using the inverse of the sandwich information matrix (*i.e.*, using the robust asymptotic covariance matrix estimator associated with quasi-maximum likelihood; see McFadden and Train, 2000).

The likelihood function (and hence, the log-likelihood function) mentioned above is parameterized with respect to the parameters of the Cholesky decomposition matrix $\boldsymbol{L}$ rather than the parameters of the original covariance parameter $\boldsymbol{\Sigma}$. This ensures the positive definiteness of $\boldsymbol{\Sigma}$, but also raises two new issues: (1) the parameters of the Cholesky matrix $\boldsymbol{L}$ should be such that $\boldsymbol{\Sigma}$ should be a correlation matrix, and (2) the estimated parameter values (and asymptotic covariance matrix) do not correspond to $\boldsymbol{\Sigma}$, but to $\boldsymbol{L}$. The first issue is overcome by parameterizing the diagonal terms of $\boldsymbol{L}$ as shown below (see Bhat and Srinivasan, 2005):

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & \sqrt{1-l_{21}^2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{I1} & l_{I2} & l_{I3} & \cdots & \sqrt{1-l_{I1}^2-l_{I2}^2\cdots-l_{I(I-1)}^2} \end{bmatrix} \tag{10}$$

The second issue is easily resolved by estimating $\boldsymbol{\Sigma}$ from the convergent values of the Cholesky decomposition parameters ($\boldsymbol{\Sigma}=\boldsymbol{LL'}$), and then running the parameter estimation procedure one more time with the likelihood function parameterized with the terms of $\boldsymbol{\Sigma}$.

3.1.3.2 The GB Simulator for the PMOP Model

An alternative simulation-based approximation of multivariate normal probabilities is provided by the Genz-Bretz algorithm (Genz and Bretz, 1999). At the first step, this method transforms the original hyper-rectangle integral region to an integral over a unit hypercube, as described in Genz (1992). The transformed integral region is filled in by randomized lattice rules using a number of points depending on the integral dimension and the desired precision. Robust integration error bounds are then derived by means of additional shifts of the integration nodes in random directions (this is similar to the generation of randomized Halton sequences, as described in Bhat, 2003, but with randomized lattice points rather than Halton points). The additional random shifts are employed to compute simulation errors using 10 sets of randomized lattice points for each individual. The interested reader is referred to Genz (2003) for details.

More recently, Genz's algorithm has been further developed by Genz and Bretz (2002). Fortran and Matlab code implementing the Genz-Bretz algorithm is available for download from the home page of Alan Genz http://www.math.wsu.edu/faculty/genz/homepage. Furthermore, the Fortran code has been included in an R (R Development Core Team, 2009) package called mvtnorm freely available from the repository http://cran.r-project.org/. For a brief description of package mvtnorm, see Hothorn *et al.* (2001) and Mi *et al.* (2009). Technically, the algorithm allows for computation of integrals up to 1,000 dimensions. However, the computational cost for reliable integral approximations explodes with the raising of the integral dimension, making the use of this algorithm impractical for likelihood inference except for low-dimensions.

In the PMOP model, a positive-definite correlation matrix $\boldsymbol{R}_q$ should result as long as $\sigma > 0$ and $0 < \rho < 1$. The GB approach implemented in the R routine is based on a check to ensure these conditions hold. If they do not hold (that is, the BHHH algorithm implemented in the R routine is trying to go outside the allowed parameter space), the algorithm reduces the "Newton-Raphson step" by half size to return the search direction within the parameter space.

3.2 The Composite Marginal Likelihood Technique – The Pairwise Marginal Likelihood Inference Approach

The composite marginal likelihood (CML) estimation approach is a relatively simple approach that can be used when the full likelihood function is near impossible or plain infeasible to evaluate due to the underlying complex dependencies. For instance, in a recent application, Varin and Czado (2010) examined the headache pain intensity of patients over several consecutive days. In this study, a full information likelihood estimator would have entailed as many as 815 dimensions of integration to obtain individual-specific likelihood contributions, an infeasible proposition using the computer-intensive simulation techniques. As importantly, the accuracy of simulation techniques is known to degrade rapidly at medium-to-high dimensions, and the simulation noise increases substantially. This leads to convergence problems during estimation. In contrast, 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 much easier-to-compute, lower-dimensional, marginal likelihoods. The CML approach can be applied using simple optimization software for likelihood estimation. It also represents a conceptually and pedagogically simpler simulation-free procedure relative to simulation techniques, and has the advantage of reproducibility of the results. Finally, as indicated by Varin and Vidoni (2009), it is possible that the "maximum CML estimator can be consistent when the ordinary full likelihood estimator is not". This is because the CML procedures are typically more robust and can represent the underlying low-dimensional process of interest more accurately than the low dimensional process implied by an assumed (and imperfect) high-dimensional multivariate model.

The simplest CML, formed by assuming independence across the latent variables underlying the ordinal outcome variables (in our paper context), entails the product of univariate probabilities for each variable. However, this approach does not provide estimates of correlation that are of interest in a multivariate context. Another approach is the pairwise likelihood function formed by the product of likelihood contributions of all or a selected subset of couplets (*i.e.*, pairs of variables or pairs of observations). Almost all earlier research efforts employing the CML technique have used the pairwise approach, including Apanasovich *et al.* (2008), Bellio and Varin (2005), de Leon (2005), Varin and Vidoni (2009), Varin *et al.* (2005), and Engle *et al.* (2007). Alternatively, the analyst can also consider larger subsets of observations, such as triplets or quadruplets or even higher dimensional subsets (see Engler *et al.*, 2006 and Caragea and Smith, 2007). In general, the issue of whether to use pairwise likelihoods or higher-dimensional likelihoods remains an open, and under-researched, area of research. However, it is generally agreed that the pairwise approach is a good balance between statistical and computation efficiency.

The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191), 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).[5] Of course, the maximum CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator (Lindsay, 1988; Zhao and Joe, 2005). On the other hand, there is also a loss in

[5] Intuitively, in the pairwise CML approach used in the current paper, the surrogate likelihood function represented by the CML function is the product of the marginal likelihood functions formed by each pair of ordinal variables. In general, maximization of the original likelihood function will result in parameters that tend to maximize each pairwise likelihood function. Since the CML is the product of pairwise likelihood contributions, it will therefore provide consistent estimates. Another equivalent way to see this is to assume we are discarding all but two randomly selected ordinal variables in the original likelihood function. Of course, we will not be able to estimate all the model parameters from two random ordinal variables, but if we could, the resulting parameters would be consistent because information (captured by other ordinal variables) is being discarded in a purely random fashion. The CML estimation procedure works similarly, but combines all ordinal variables observed two at a time, while ignoring the full joint distribution of the ordinal variables.

asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator (see McFadden and Train, 2000). Given the full likelihood estimator has to be approximated using simulation techniques in a multivariate ordered-response system of dimensionality more than 3, it is of interest to compare the MSL and CML estimators in terms of asymptotic efficiency.

Earlier applications of the CML approach (and specifically the pairwise likelihood approach) to multivariate ordered-response systems include de Leon (2005) and Ferdous *et al.* (2010) in the context of cross-sectional multivariate ordered-response probit (CMOP) systems, and Varin and Vidoni (2006) and Varin and Czado (2010) in the context of panel multivariate ordered-response probit (PMOP) systems. Bhat *et al.* (2010) also use a CML approach to estimate their multivariate ordered-response probit system in the context of a spatially dependent ordered response outcome variable. In this study, we do not use the high multivariate dimensionality of most of these earlier studies. Rather, we consider relatively lower multivariate dimensionality simulation situations, so that we are able to estimate the models using MSL techniques too.

3.2.1 Pairwise Likelihood Approach for the CMOP Model

The pairwise marginal likelihood function for individual q may be written for the CMOP model as follows:

$$L_{CML,q}^{CMOP}(\delta) = \prod_{i=1}^{I-1} \prod_{g=i+1}^{I} \Pr(y_{qi} = m_{qi}, y_{qg} = m_{qg})$$

$$= \prod_{i=1}^{I-1} \prod_{g=i+1}^{I} \left[\begin{array}{l} \Phi_2\left(\theta_i^{m_{qi}} - \beta_i' x_{qi}, \theta_g^{m_{qg}} - \beta_g' x_{qg}, \rho_{ig}\right) - \Phi_2\left(\theta_i^{m_{qi}} - \beta_i' x_{qi}, \theta_g^{m_{qg}-1} - \beta_g' x_{qg}, \rho_{ig}\right) \\ - \Phi_2\left(\theta_i^{m_{qi}-1} - \beta_i' x_{qi}, \theta_g^{m_{qg}} - \beta_g' x_{qg}, \rho_{ig}\right) + \Phi_2\left(\theta_i^{m_{qi}-1} - \beta_i' x_{qi}, \theta_g^{m_{qg}-1} - \beta_g' x_{qg}, \rho_{ig}\right) \end{array} \right], \quad (11)$$

where $\Phi_2(.,.,\rho_{ig})$ is the standard bivariate normal cumulative distribution function with correlation ρ_{ig}. The pairwise marginal likelihood function is $L_{CML}^{CMOP}(\delta) = \prod_q L_{CML,q}^{CMOP}(\delta)$.

The pairwise estimator $\hat{\delta}_{CML}$ obtained by maximizing the logarithm of the pairwise marginal likelihood function with respect to the vector δ is consistent and asymptotically normal distributed with asymptotic mean δ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix $G(\delta)$ (see Zhao and Joe, 2005):

$V_{CML}(\delta) = [G(\delta)]^{-1} = [H(\delta)]^{-1} J(\delta)[H(\delta)]^{-1}$, where

$$H(\delta) = E\left[-\frac{\partial^2 \log L_{CML}^{CMOP}(\delta)}{\partial\delta\partial\delta'} \right] \text{ and} \tag{12}$$

$$J(\delta) = E\left[\left(\frac{\partial \log L_{CML}^{CMOP}(\delta)}{\partial\delta} \right) \left(\frac{\partial \log L_{CML}^{CMOP}(\delta)}{\partial\delta'} \right) \right]$$

$H(\delta)$ and $J(\delta)$ can be estimated in a straightforward manner at the CML estimate ($\hat{\delta}_{CML}$):

$$\hat{H}(\hat{\delta}) = -\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{CML,q}^{CMOP}(\delta)}{\partial\delta\partial\delta'} \right]_{\hat{\delta}}$$

$$= -\left[\sum_{q=1}^{Q} \sum_{i=1}^{I-1} \sum_{g=i+1}^{I} \frac{\partial^2 \log \Pr(y_{qi} = m_{qi}, y_{qg} = m_{qg})}{\partial\delta\partial\delta'} \right]_{\hat{\delta}}, \text{ and} \tag{13}$$

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

In general, and as confirmed later in the simulation study, we expect that the ability to recover and pin down the parameters will be a little more difficult for the correlation parameters in $\mathbf{\Sigma}$ (when the correlations are low) than for the slope and threshold parameters, because the correlation parameters enter more non-linearly in the likelihood function.

3.2.2 Pairwise Likelihood Approach for the PMOP Model

The pairwise marginal likelihood function for individual q may be written for the PMOP model as follows:

$$L_{CML,q}^{PMOP}(\delta) = \prod_{j=1}^{J-1} \prod_{g=j+1}^{J} \Pr(y_{qj} = m_{qj}, y_{qg} = m_{qg})$$

$$= \prod_{j=1}^{J-1} \prod_{g=j+1}^{J} \left[\begin{array}{l} \Phi_2\left(\alpha^{m_{qj}}, \alpha^{m_{qg}}, \rho_{jg}\right) - \Phi_2\left(\alpha^{m_{qj}}, \alpha^{m_{qg-1}}, \rho_{jg}\right) \\ - \Phi_2\left(\alpha^{m_{qj}-1}, \alpha^{m_{qg}}, \rho_{jg}\right) + \Phi_2\left(\alpha^{m_{qj}-1}, \alpha^{m_{qg}-1}, \rho_{jg}\right) \end{array} \right], \tag{14}$$

where $\alpha^{m_{qj}} = (\theta^{m_{qj}} - \beta' x_{qj}) / \mu$, $\mu = \sqrt{1+\sigma^2}$, and $\rho_{jg} = (\sigma^2 + \rho^{|t_{qj}-t_{qg}|}) / \mu^2$. The pairwise marginal likelihood function is $L_{CML}^{PMOP}(\delta) = \prod_q L_{CML,q}^{PMOP}(\delta)$.

The pairwise estimator $\hat{\delta}_{CML}$ obtained by maximizing the logarithm of the pairwise marginal likelihood function with respect to the vector δ is consistent and asymptotically normal distributed with asymptotic mean δ. The covariance matrix of the estimator may be computed in a fashion similar to that for the CMOP case, with $L_{CML,q}^{CMOP}(\delta)$ being replaced by $L_{CML,q}^{PMOP}(\delta)$.

As in the CMOP case, we expect that the ability to recover and pin down the parameters will be a little more difficult for the correlation parameter ρ (when ρ is low) than for the slope and threshold parameters.

3.2.3 Positive-Definiteness of the Implied Multivariate Correlation Matrix

A point that we have not discussed thus far in the CML approach is how to ensure the positive-definiteness of the symmetric correlation matrix $\mathbf{\Sigma}$ (in the CMOP model) and $\boldsymbol{R}_q$ (in the PMOP model). This is particularly an issue for $\mathbf{\Sigma}$ in the CMOP model, so we will discuss this mainly in the context of the CMOP model. Maintaining a positive-definite matrix for $\boldsymbol{R}_q$ in the PMOP model is relatively easy, so we only briefly discuss the PMOP case toward the end of this section.

There are three ways that one can ensure the positive-definiteness of the $\mathbf{\Sigma}$ matrix. The first technique is to use Bhat and Srinivasan's technique of reparameterizing $\mathbf{\Sigma}$ through the Cholesky matrix, and then using these Cholesky-decomposed parameters as the ones to be estimated. Within the optimization procedure, one would then reconstruct the $\mathbf{\Sigma}$ matrix, and then “pick off” the appropriate elements of this matrix for the ρ_{ig} estimates at each iteration. This is probably the most straightforward and clean technique. The second technique is to undertake the estimation with a constrained optimization routine by requiring that the implied multivariate correlation matrix for any set of pairwise correlation estimates be positive definite. However, such a constrained routine can be extremely cumbersome. The third technique is to use an unconstrained optimization routine, but check for positive-definiteness of the implied multivariate correlation matrix. The easiest method within this third technique is to allow the estimation to proceed without checking for positive-definiteness at intermediate iterations, but check that the implied multivariate correlation matrix at the final converged pairwise marginal likelihood estimates is positive-definite. This will typically work for the case of a multivariate ordered-response model if one specifies exclusion restrictions (*i.e.*, zero correlations between some error terms) or correlation patterns that involve a lower dimension of effective parameters (such as in the PMOP model in the current paper). Also, the number of correlation parameters in the full multivariate matrix explodes quickly as the dimensionality of the matrix increases, and estimating all these parameters becomes almost impossible (with any estimation technique) with the usual sample sizes available in practice. So, imposing exclusion restrictions is good econometric practice. However, if the above simple method of allowing the pairwise marginal estimation approach to proceed without checking for positive definiteness at intermediate iterations does not work, then one can check the implied multivariate correlation matrix for positive definiteness at each and every iteration. If the matrix is not positive-definite during a direction search at a given iteration, one can construct a “nearest” valid correlation matrix (see Ferdous *et al.*, 2010 for a discussion).

In the CMOP CML analysis of the current paper, we used an unconstrained optimization routine and ensured that the implied multivariate correlation matrix at convergence was positive-definite. In the PMOP CML analysis of the current paper, we again employed an unconstrained optimization routine in combination with the following reparameterizations: $\rho = 1/[1+\exp(-\psi)]$, and $\sigma = \exp(\pi)$. These reparameterizations were used to guarantee $\sigma > 0$ and $0 < \rho < 1$, and therefore the positive-definiteness of the $\boldsymbol{R}_q$ multivariate correlation matrix. Once estimated, the ψ and π estimates were translated back to estimates of ρ and σ.

4. EXPERIMENTAL DESIGN

4.1 The CMOP Model

To compare and evaluate the performance of the GHK and the CML estimation techniques, we undertake a simulation exercise for a multivariate ordered response system with five ordinal variables. Further, to examine the potential impact of different correlation structures, we undertake the simulation exercise for a correlation structure with low correlations and another with high correlations. For each correlation structure, the experiment is carried out for 20 independent data sets with 1000 data points. Pre-specified values for the δ vector are used to generate samples in each data set.

In the set-up, we use three exogenous variables in the latent equation for the first, third, and fifth ordered-response variables, and four exogenous variables for the second and fourth ordered-response variables. The values for each of the exogenous variables are drawn from a standard univariate normal distribution. A fixed coefficient vector β_i $(i = 1, 2, 3, 4, 5)$ is assumed on the variables, and the linear combination $\beta_i' x_{qi}$ (q = 1, 2, …, Q, Q = 1000; i = 1, 2, 3, 4, 5) is computed for each individual q and category i. Next, we generate Q five-variate realizations of the error term vector $(\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \varepsilon_{q4}, \varepsilon_{q5})$ with predefined positive-definite low error correlation structure ($\mathbf{\Sigma}_{low}$) and high error correlation structure ($\mathbf{\Sigma}_{high}$) as follows:

$$\mathbf{\Sigma}_{low} = \begin{bmatrix} 1 & .30 & .20 & .22 & .15 \\ .30 & 1 & .25 & .30 & .12 \\ .20 & .25 & 1 & .27 & .20 \\ .22 & .30 & .27 & 1 & .25 \\ .15 & .12 & .20 & .25 & 1 \end{bmatrix}, \text{ and } \mathbf{\Sigma}_{high} = \begin{bmatrix} 1 & .90 & .80 & .82 & .75 \\ .90 & 1 & .85 & .90 & .72 \\ .80 & .85 & 1 & .87 & .80 \\ .82 & .90 & .87 & 1 & .85 \\ .75 & .72 & .80 & .85 & 1 \end{bmatrix} \quad (15)$$

The error term realization for each observation and each ordinal variable is then added to the systematic component $(\beta_i' x_{qi})$ as in Equation (1) and then translated to "observed" values of y_{qi} (0, 1, 2, ...) based on pre-specified threshold values. We assume four outcome levels for the first and the fifth ordered-response variables, three for the second and the fourth ordered-response variables, and five for the third ordered-response variable. Correspondingly, we pre-specify a vector of three threshold values [$(\theta_i = \theta_i^1, \theta_i^2, \theta_i^3)$, where i = 1 and 5] for the first and the fifth ordered-response equations, two for the second and the fourth equations [$(\theta_i = \theta_i^1, \theta_i^2)$, where i = 2 and 4], and four for the third ordered-response equation [$(\theta_i = \theta_i^1, \theta_i^2, \theta_i^3, \theta_i^4)$, where $i = 3$] .

As mentioned earlier, the above data generation process is undertaken 20 times with different realizations of the random error term to generate 20 different data sets. The CML estimation procedure is applied to each data set to estimate data-specific values of the δ vector. The GHK simulator is applied to each dataset using 100 draws per individual of the randomized Halton sequence.[6] In addition, to assess and to quantify simulation variance, the GHK simulator is applied to each dataset 10 times with different (independent) randomized Halton draw sequences. This allows us to estimate simulation error by computing the standard deviation of estimated parameters among the 10 different GHK estimates on the same data set.

A few notes are in order here. We chose to use a setting with five ordinal variables so as to keep the computation time manageable for the maximum simulated likelihood estimations (going to, for example, 10 ordinal variables will increase computation time substantially, especially since more number of draws per individual may have to be used; note also that we have a total of 400 MSL estimation runs just for the five ordinal variable case in our experimental design). At the same time, a system of five ordinal variables leads to a large enough dimensionality of integration in the likelihood function where simulation estimation has to be used. Of course, one can examine the effect of varying the number of ordinal variables on the performance of the MSL and CML estimation approaches. In this paper, we have chosen to focus on five dimensions, and examine the effects of varying correlation patterns and different model formulations (corresponding to cross-sectional and panel settings). A comparison with higher numbers of ordinal variables is left as a future exercise. However, in general, it is well known that MSL estimation gets more imprecise as the dimensionality of integration increases. On the other hand, our experience with CML estimation is that the performance does not degrade very much as the number of ordinal variables increases (see Ferdous *et al.*, 2010). Similarly, one can examine the effect of varying numbers of draws for MSL estimation. Our choice of 100 draws per individual was based on experimentation with different numbers of draws for the first data

[6] Bhat (2001) used Halton sequence to estimate mixed logit models, and found that the simulation error in estimated parameters is lower with 100 Halton draws than with 1000 random draws (per individual). In our study, we carried out the GHK analysis of the multivariate ordered-response model with 100 randomized Halton draws as well as 500 random draws per individual, and found the 100 randomized Halton draws case to be much more accurate/efficient as well as much less time-consuming. So, we present only the results of the 100 randomized Halton draws case here.

set. We found little improvement in ability to recover parameters or simulation variance beyond 100 draws per individual for this data set, and thus settled for 100 draws per individual for all data sets (as will be noted in the results section, the CMOP MSL estimation with 100 draws per individual indeed leads to negligible simulation variance). Finally, we chose to use three to four exogenous variables in our experimental design (rather than use a single exogenous variable) so that the resulting simulation data sets would be closer to realistic ones where multiple exogenous variables are employed.

4.2 The PMOP Model

For the panel case, we consider six observations (J = 6) per individual, leading to a six-dimensional integral per individual for the full likelihood function. Note that the correlation matrix $\boldsymbol{R_q}$ has entries such that $corr(y^*_{qj}, y^*_{qg}) = (\sigma^2 + \rho^{|t_{qj}-t_{qg}|}) / \mu^2$, where $\mu = \sqrt{1+\sigma^2}$. Thus, in the PMOP case, $\boldsymbol{R_q}$ is completely determined by the variance σ^2 of the individual-specific non-varying random term u_q and the single autoregressive correlation parameter ρ determining the correlation between the ε_{qj} and ε_{qk} terms: $corr(\varepsilon_{qj}, \varepsilon_{qk}) = \rho^{|t_{qj}-t_{qk}|}$. To examine the impact of different magnitudes of the autoregressive correlation parameter, we undertake the simulation exercise for two different values of ρ: 0.3 and 0.7. For each correlation parameter, the experiment is carried out for 100 independent data sets with 200 data points (*i.e.*, individuals).[7] Pre-specified values for the δ vector are used to generate samples in each data set.

In the set-up, we use two exogenous variables in the latent equation. One is a binary time-constant variable (x_{q1}) simulated from a Bernoulli variable with probability equal to 0.7, and another (x_{qj2}) is a continuous time-varying variable generated from the autoregressive model shown below:

---

[7] Note that we use more independent data sets for the panel case than the cross-sectional case, because the number of individuals in the panel case is fewer than the number of individuals in the cross-sectional case. Essentially, the intent is to retain the same order of sampling variability in the two cases across individuals and data sets (the product of the number of observations per data set and the number of data sets is 20,000 in the cross-sectional and the panel cases). Further, the lower number of data sets in the cross-sectional case is helpful because maximum simulated likelihood is more expensive relative to the panel case, given that the number of parameters to be estimated is substantially more than in the panel case. Note also that the dimensionality of the correlation matrices is about the same in the cross-sectional and panel cases. We use T = 6 in the panel case because the serial correlation gets manifested in the last five of the six observations for each individual. The first observation error term ε_{q1} for each individual q is randomly drawn from the normal distribution with variance σ^2.

$$(x_{qj2}-1)=0.6(x_{q,j-1,2}-1)+\gamma_{qj}, \gamma_{qj} \overset{iid}{\sim} N(0,0.2^2). \tag{16}$$

A fixed coefficient vector β is assumed, with $\beta_1 = 1$ (coefficient on x_{q1}) and $\beta_2 = 1$ (coefficient on x_{qj2}). The linear combination $\beta' x_{qj}$ ($x_{qj} = (x_{q1}, x_{qj2})'$; q = 1, 2, …, 200) is computed for each individual q's j^{th} observation. Next, we generate independent time-invariant values of u_q for each individual from a standard normal distribution (that is, we assume $\sigma^2 = 1$), and latent serially correlated errors for each individual q as follows:

$$\varepsilon_{qj} = \begin{cases} \eta_{q1} \overset{iid}{\sim} N(0,1) \text{ for } j=1 \\ \rho\varepsilon_{qj-1} + \left(\sqrt{1-\rho^2}\right)\eta_{qj}, \ \ \eta_{qj} \overset{iid}{\sim} N(0,1) \text{ for } j \geq 2. \end{cases} \tag{17}$$

The error term realizations for each individual's observation is then added to the systematic component ($\beta' x_{qj}$) as in Equation (4) and then translated to "observed" values of y_{qj} based on the following pre-specified threshold values: $\theta^1 = 1.5$, $\theta^2 = 2.5$, and $\theta^3 = 3.0$. The above data generation process is undertaken 100 times with different realizations of the random error terms u_q and ε_{qj} to generate 100 different data sets. The CML estimation procedure is applied to each data set to estimate data-specific values of the δ vector. The GB simulator is applied to each data set 10 times with different (independent) random draw sequences. This allows us to estimate simulation error by computing the standard deviation of estimated parameters among the 10 different GB estimates on the same data set. The algorithm is tuned with an absolute error tolerance of 0.001 for each six-dimensional integral forming the likelihood. The algorithm is adaptive in that it starts with few points and then increases the number of points per individual until the desired precision is obtained, but with the constraint that the maximal number of draws is 25,000.

5. 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 (GHK for CMOP and GB for PMOP) and the CML approach. The subsequent sections present the simulation and computational results.

5.1 Performance Measures

The steps discussed below for computing performance measures are for a specific correlation matrix pattern. For the CMOP model, we consider two correlation matrix patterns, one with low correlations and another with high correlations. For the PMOP model, we 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, …, 20 for CMOP and *s* = 1, 2, …, 100 for PMOP; *i.e.*, *S* = 20 for CMOP and *S* = 100 for PMOP) 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).

(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 **absolute percentage bias** (APB) as:

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

(4) Compute the standard deviation of the MED values across the data sets and label this as the **finite sample standard error** (essentially, this is the empirical standard error).

(5) For each data set *s*, compute the median s.e. for each model parameter across the draws. Call this MSED, and then take the mean of the MSED values across the *R* 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). Note that we compute the median s.e. for each model parameter across the draws and label it as MSED rather than computing the mean s.e. for each model parameter across the draws. This is because, for

some draws, the estimated standard errors turned out to be rather large relative to other independent standard error estimates for the same dataset. On closer inspection, this could be traced to the unreliability of the numeric Hessian used in the sandwich estimator computation. This is another bothersome issue with MSL -- it is important to compute the covariance matrix using the sandwich estimator rather than using the inverse of the cross-product of the first derivatives (due to the simulation noise introduced when using a finite number of draws per individual in the MSL procedure; see McFadden and Train, 2000). Specifically, using the inverse of the cross-product of the first derivatives can substantially underestimate the covariance matrix. But coding the analytic Hessian (as part of computing the sandwich estimator) is extremely difficult, while using the numeric Hessian is very unreliable. Craig (2008) also alludes to this problem when he states that "(...) the randomness that is inherent in such methods [referring here to the GB algorithm, but applicable in general to MSL methods] is sometimes more than a minor nuisance." In particular, even when the log-likelihood function is computed with good precision so that the simulation error in estimated parameters is very small, this is not always adequate to reliably compute the numerical Hessian. To do so, one will generally need to compute the log-likelihood with a substantial level of precision, which, however, would imply very high computational times even in low dimensionality situations. Finally, note that the mean asymptotic standard error is a theoretical approximation to the finite sample standard error, since, in practice, one would estimate a model on only one data set from the field.

(6) 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 deviation as SIMMED. For each parameter, take the mean of SIMMED across the different data sets. Label this as the **simulation s.e.** for each parameter.

(7) 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.[8]

(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 standard deviation of the CML parameter estimates across the data sets and label this as the **finite sample standard error** (essentially, this is the empirical standard error).

5.2 Simulation Results

5.2.1 THE CMOP Model

Table 1a presents the results for the CMOP model with low correlations, and Table 1b presents the corresponding results for the CMOP model with high correlations. The results indicate that both the MSL and CML approaches recover the parameters extremely well, as can be observed by comparing the mean estimate of the parameters with the true values (see the column titled "parameter estimates"). In the low correlation case, the absolute percentage bias (APB) ranges from 0.03% to 15.95% (overall mean value of 2.21% - see last row of table under the column titled "absolute percentage bias") across parameters for the MSL approach, and from 0.00% to 12.34% (overall mean value of 1.92%) across parameters for the CML approach. In the high correlation case, the APB ranges from 0.02% to 5.72% (overall mean value of 1.22% - see last row of table under the column titled "absolute percentage bias") across parameters for the MSL approach, and from 0.00% to 6.34% (overall mean value of 1.28%) across parameters for the CML approach. These are incredibly good measures for the ability to recover parameter estimates, and indicate that both the MSL and CML perform about evenly in the context of bias. Further, the ability to recover parameters does not seem to be affected at all by whether there is low correlation or high correlation (in fact, the overall APB reduces from the low correlation case to the high correlation case). Interestingly, the absolute percentage bias values are generally much higher for the correlation (ρ) parameters than for the slope (β) and threshold (θ) parameters in the low correlation case, but the situation is exactly reversed in the high correlation case where the absolute percentage bias values are generally higher for the slope (β) and

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

threshold (θ) parameters compared to the correlation (ρ) parameters (for both the MSL and CML approaches). This is perhaps because the correlation parameters enter more non-linearly in the likelihood function than the slope and threshold parameters, and need to be particularly strong before they start having any substantial effects on the log-likelihood function value. Essentially, the log-likelihood function tends to be relatively flat at low correlations, leading to more difficulty in accurately recovering the low correlation parameters. But, at high correlations, the log-likelihood function shifts considerably in value with small shifts in the correlation values, allowing them to be recovered accurately.[9]

The standard error measures provide several important insights. First, the finite sample standard error and asymptotic standard error values are quite close to one another, with very little difference in the overall mean values of these two columns (see last row). This holds for both the MSL and CML estimation approaches, and for both the low and high correlation cases, and confirms that the inverses of the sandwich information estimator (in the case of the MSL approach) and the Godambe information matrix estimator (in the case of the CML approach) recover the finite sample covariance matrices remarkably well. Second, the empirical and asymptotic standard errors for the threshold parameters are higher than for the slope and correlation parameters (for both the MSL and CML cases, and for both the low and high correlation cases). This is perhaps because the threshold parameters play a critical role in the partitioning of the underlying latent variable into ordinal outcomes (more so than the slope and correlation parameters), and so are somewhat more difficult to pin down. Third, a comparison of the standard errors across the low and high correlation cases reveals that the empirical and asymptotic standard errors are much lower for the correlation parameters in the latter case than in the former case. This reinforces the finding earlier that the correlation parameters are much easier to recover at high values because of the considerable influence they have on the log-likelihood function at high values; consequently, not only are they recovered accurately, but they are also recovered more precisely at high correlation values. Fourth, across all parameters, there is a reduction in the empirical and asymptotic standard errors for both the MSL and CML cases

[9] One could argue that the higher absolute percentage bias values for the correlation parameters in the low correlation case compared to the high correlation case is simply an artifact of taking percentage differences from smaller base correlation values in the former case. However, the sum of the absolute values of the deviations between the mean estimate and the true value is 0.0722 for the low correlation case and 0.0488 for the high correlation case. Thus, the correlation values are indeed being recovered more accurately in the high correlation case compared to the low correlation case.

between the low and high correlation cases (though the reduction is much more for the correlation parameters than for the non-correlation parameters). Fifth, the simulation error in the MSL approach is negligible to small. On average, based on the mean values in the last row of the table, the simulation error is about 3.9% of the sampling error for the low correlation case and 10.3% of the sampling error for the high correlation case. The higher simulation error for the high correlation case is not surprising, since we use the same number of Halton draws per individual in both the low and high correlation cases, and the multivariate integration is more involved with a high correlation matrix structure. Thus, as the levels of correlations increase, the evaluation of the multivariate normal integrals can be expected to become less precise at a given number of Halton draws per individual. However, overall, the results suggest that our MSL simulation procedure is well tuned, and that we are using adequate numbers of Halton draws per individual for the accurate evaluation of the log-likelihood function and the accurate estimation of the model parameters (this is also reflected in the negligible difference in the simulation-adjusted standard error and the mean asymptotic standard error of parameters in the MSL approach).

The final two columns of each of Tables 1a and 1b provide a relative efficiency factor between the MSL and CML approaches. The first of these columns provides the ratio of the asymptotic standard error of parameters from the MSL approach and the asymptotic standard error of the corresponding parameters from the CML approach. The second of these columns 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 expected, the second column provides slightly higher values of efficiency, indicating that CML efficiency increases when one also considers the presence of simulation standard error in the MSL estimates. However, this efficiency increase is negligible in the current context because of very small MSL simulation error. The more important and interesting point though is that the relative efficiency of the CML approach is as good as the MSL approach in the low correlation case. This is different from the relative efficiency results obtained in Renard *et al.* (2004), Zhao and Joe (2005), and Kuk and Nott (2000) in other model contexts, where the CML has been shown to lose efficiency relative to a maximum likelihood approach. However, note that all these other earlier studies focus on a comparison of a CML approach vis-à-vis a maximum likelihood (ML) approach, while, in our setting, we must resort to MSL to approximate the likelihood function.

To our knowledge, this is the first comparison of the CML approach to an MSL approach, applicable to situations when the full information maximum likelihood estimator cannot be evaluated analytically. In this regard, it is not clear that the earlier theoretical result that 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 would extend to our case because the asymptotic covariance of MSL is computed as the inverse of the sandwich information matrix.[10] 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. Our results regarding the efficiency of individual parameters suggests that any reduction in efficiency of the CML (because of using only pairwise likelihoods rather than the full likelihood) is balanced by the reduction in efficiency because of using MSL rather than ML, so that there is effectively no loss in asymptotic efficiency in using the CML approach (relative to the MSL approach) in the CMOP case for low correlation. However, for the high correlation case, the MSL does provide slightly better efficiency than the CML. However, even in this case, the relative efficiency of parameters in the CML approach ranges between 90%-99% (mean of 95%) of the efficiency of the MSL approach, without considering simulation standard error. When considering simulation error, the relative efficiency of the CML approach is even better at about 96% of the MSL efficiency (on average across all parameters). Overall, there is little to no drop in efficiency because of the use of the CML approach in the CMOP simulation context.

5.2.2 The PMOP Model

Most of the observations made from the CMOP model results also hold for the PMOP model results presented in Table 2. Both the MSL and CML approaches recover the parameters extremely well. In the low correlation case, the absolute percentage bias (APB) ranges from

---

[10] 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.

0.26% to 4.29% (overall mean value of 1.29%) across parameters for the MSL approach, and from 0.65% to 5.33% (overall mean value of 1.84%) across parameters for the CML approach. In the high correlation case, the APB ranges from 0.45% to 6.14% (overall mean value of 2.06%) across parameters for the MSL approach, and from 0.41% to 5.71% (overall mean value of 2.40%) across parameters for the CML approach. Further, the ability to recover parameters does not seem to be affected too much in an absolute sense by whether there is low correlation or high correlation. The CML approach shows a mean value of absolute percentage bias that increases about 1.3 times (from 1.84% to 2.40%) between the low and high ρ values compared to an increase of about 1.6 times (from 1.29% to 2.06%) for the MSL approach. It in indeed interesting that the PMOP results indicate a relative increase in the APB values from the low to high correlation case, while there was actually a corresponding relative decrease in the CMOP case. Another result is that the APB increases from the low to the high correlation case for the threshold (θ) and variance (σ^2) parameters in both the MSL and CML approaches. On the other hand, the APB decreases from the low to the high correlation case for the correlation (ρ) parameter, and remains relatively stable between the low and high correlation cases for the slope (β) parameters. That is, the recovery of the slope parameters appears to be less sensitive to the level of correlation than is the recovery of other parameters.

The finite sample standard error and asymptotic standard error values are close to one another, with very little difference in the overall mean values of these two columns (see last row). This holds for both the MSL and CML approaches. Also, as in the CMOP case, the empirical and asymptotic standard errors for the threshold parameters are generally higher than for the other parameters. The simulation error in the MSL approach is negligible, at about 0.1% or less than the sampling error for both the low and high correlation cases. Note that, unlike in the CMOP case, the PMOP MSL estimation did not involve the same number of draws per individual for the low and high correlation cases; rather, the number of draws varied to ensure an absolute error tolerance of 0.001 for each six-dimensional integral forming the likelihood. Thus, it is no surprise that the simulation error does not increase much between the low and high correlation cases as it did in the CMOP case. A significant difference with the CMOP case is that the empirical standard errors and asymptotic standard errors are consistently larger for the high correlation case than for the low correlation case, with a particularly substantial increase in the standard error of σ^2.

The final two columns provide a relative efficiency factor between the MSL and CML approaches. The values in these two columns are identical because of the very low simulation error. As in the CMOP case, the estimated efficiency of the CML approach is as good as the MSL approach in the low correlation case (the relative efficiency ranges between 90%-103%, with a mean of 97%). For the high correlation case, the relative efficiency of parameters in the CML approach ranges between 82%-96% (mean of 91%) of the efficiency of the MSL approach, indicating a reduction in efficiency as the dependence level goes up (again, consistent with the CMOP case). Overall, however, the efficiency of the CML approach remains high for all the parameters.

5.3 Non-Convergence and Computational Time

The simulation estimation of multivariate ordered response model can involve numerical instability because of possible unstable operations such as large matrix inversions and imprecision in the computation of the Hessian. This can lead to convergence problems. On the other hand, the CML approach is a straightforward approach that should be easy to implement and should not have any convergence-related problems. In the current empirical study, we classified any estimation run that had not converged in 5 hours as having non-converged.

We computed non-convergence rates in two ways for the MSL approach. For the CMOP model, we computed the non-convergence rates in terms of the starting seeds that led to failure in a complete estimation of 10 simulation runs (using different randomized Halton sequences) for each data set. If a particular starting seed led to failure in convergence for any of the 10 simulation runs, that seed was classified as a failed seed. Otherwise, the seed was classified as a successful seed. This procedure was applied for each of the 20 data sets generated for each of the low and high correlation matrix structures until we had a successful seed.[11] The non-convergence rate was then computed as the number of failed seeds divided by the total number of seeds

[11] Note that we use the terminology "successful seed" to simply denote if the starting seed led to success in a complete estimation of the 10 simulation runs. In MSL estimation, it is not uncommon to obtain non-convergence (because of a number of reasons) for some sets of random sequences. There is, however, nothing specific to be learned here in terms of what starting seeds are likely to be successful and what starting seeds are likely to be unsuccessful. The intent is to use the terminology "successful seed" simply as a measure of non-convergence rates.

considered. Note that this would be a good reflection of non-convergence rates if the analyst ran the simulation multiple times on a single data set to recognize simulation noise in statistical inferences. But, in many cases, the analyst may run the MSL procedure only once on a single data set, based on using a high level of accuracy in computing the multivariate integrals in the likelihood function. For the PMOP model, which was estimated based on as many draws as needed to obtain an absolute error tolerance of 0.001 for each six-dimensional integral forming the likelihood, we therefore consider another way of computing non-convergence. This is based on the number of unsuccessful runs out of the 1000 simulated estimation runs considered (100 data sets times 10 simulated estimation runs). The results indicated a non-convergence rate of 28.5% for the low correlation case and 35.5% for the high correlation case in the CMOP model, and a non-convergence rate of 4.2% for the low correlation case and 2.4% for the high correlation case in the PMOP model (note, however, that the rates cannot be compared between the CMOP and PMOP models because of very different ways of computing the rates, as discussed above). For both the CMOP and PMOP models, and both the low and high correlation cases, we always obtained convergence with the CML approach.

Next, we examined the time to convergence per converged estimation run for the MSL and CML procedures (the time to convergence included the time to compute the standard error of parameters). For the CMOP model, we had a very well-tuned and efficient MSL procedure with an analytic gradient (written in Gauss matrix programming language). We used naïve independent probit starting values for the MSL as well as the CML in the CMOP case (the CML is very easy to code relative to the MSL, and was also undertaken in the GAUSS language for the CMOP model). The estimations were run on a desktop machine. But, for the PMOP model, we used an MSL code written in the R language without an analytic gradient, and a CML code written using a combination of C and R languages. However, we used the CML convergent values (which are pretty good) as the MSL start values in the PMOP model to compensate for the lack of analytic MSL gradients. The estimations were run on a powerful server machine. As a consequence of all these differences, one needs to be careful in the computational time comparisons. Here, we only provide a relative computational time factor (RCTF), computed as the mean time needed for an MSL run divided by the mean time needed for a CML run. In addition, we present the standard deviation of the run times as a percentage of mean run time (SDR) for the MSL and CML estimations.

The RCTF for the CMOP model for the case of the low correlation matrix is 18, and for the case of the high correlation matrix is 40. The substantially higher RCTF for the high correlation case is because of an increase in the mean MSL time between the low and high correlation cases; the mean CML time hardly changed. The MSL SDR in the CMOP model for the low correlation case is 30% and for the high correlation case is 47%, while the CML SDR is about 6% for both the low and high correlation cases. The RCTF for the PMOP model for the case of low correlation is 332, and for the case of high correlation is 231. The MSL SDR values for the low and high correlation cases in the PMOP model are in the order of 16-24%, though this small SDR is also surely because of using the CML convergent values as the start values for the MSL estimation runs. The CML SDR values in the PMOP model are low (6-13%) for both the low and high correlation cases. Overall, the computation time results do very clearly indicate the advantage of the CML over the MSL approach – the CML approach estimates parameters in much less time than the MSL, and the stability in the CML computation time is substantially higher than the stability in the MSL computation times. As the number of ordered-response outcomes increase, one can only expect a further increase in the computational time advantage of the CML over the MSL estimation approach.

6. CONCLUSIONS

This paper compared the performance of the maximum-simulated likelihood (MSL) approach with the composite marginal likelihood (CML) approach in multivariate ordered-response situations. We used simulated data sets with known underlying model parameters to evaluate the two estimation approaches in the context of a cross-sectional ordered-response setting as well as a panel ordered-response setting. The ability of the two approaches to recover model parameters was examined, as was 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 were also presented.

Overall, the simulation results demonstrate the ability of the Composite Marginal Likelihood (CML) approach to recover the parameters in a multivariate ordered-response choice model context, independent of the correlation structure. In addition, the CML approach recovers parameters as well as the MSL estimation approach in the simulation contexts used in the current study, while also doing so at a substantially reduced computational cost and improved computational stability. Further, any reduction in the efficiency of the CML approach relative to the MSL approach is in the range of non-existent to small. All these factors, combined with the conceptual and implementation simplicity of the CML approach, makes it a promising and simple approach not only for the multivariate ordered-response model considered here but also for other analytically-intractable econometric models. Also, as the dimensionality of the model explodes, the CML approach remains practical and feasible, while the MSL approach becomes impractical and/or infeasible. Additional comparisons of the CML approach with the MSL approach for high dimensional model contexts and alternative covariance patterns are directions for further research.

ACKNOWLEDGEMENTS

The authors are grateful to Lisa Macias for her help in formatting this document. Two referees provided important input on an earlier version of the paper.

REFERENCES

Apanasovich, T.V., D. Ruppert, J.R. Lupton, N. Popovic, N.D. Turner, R.S. Chapkin, and R.J. Carroll (2008) Aberrant crypt foci and semiparametric modelling of correlated binary data. *Biometrics*, 64(2), 490-500.

Balia, S., and A.M. Jones (2008) Mortality, lifestyle and socio-economic status. *Journal of Health Economics*, 27(1), 1-26.

Bellio, R., and C. Varin (2005) A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, 5(3), 217-227.

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., and S. Srinivasan (2005) A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B*, 39(3), 255-278.

Bhat, C.R., I.N. Sener, and N. Eluru (2010) 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.

Caragea, P.C., and R.L. Smith (2007) Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98(7), 1417- 1440.

Chen, M.-H., and D.K. Dey (2000) Bayesian analysis for correlated ordinal data models. In *Generalized Linear Models: A Bayesian Perspective*, D.K. Dey, S.K. Gosh, and B.K. Mallick (eds), Marcel Dekker, New York.

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

Craig, P. (2008) A new reconstruction of multivariate normal orthant probabilities. *Journal of the Royal Statistical Society: Series B*, 70(1), 227-243.

de Leon, A.R. (2005) Pairwise likelihood approach to grouped continuous model and its extension. *Statistics & Probability Letters*, 75(1), 49-57.

Engle, R.F., N. Shephard, and K. Sheppard (2007) Fitting and testing vast dimensional time-varying covariance models. Finance Working Papers, FIN-07-046, Stern School of Business, New York University.

Engler, D.A., M. Mohapatra, D.N. Louis, and R.A. Betensky (2006) A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics*, 7(3), 399-421.

Ferdous, N., N. Eluru, C.R. Bhat, and I. Meloni (2010) A multivariate ordered-response model system for adults' weekday activity episode generation by activity purpose and social context. *Transportation Research Part B*, 44(8-9), 922-943.

Genz, A. (1992) Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2), 141-149.

Genz, A. (2003) Fully symmetric interpolatory rules for multiple integrals over hyper-spherical surfaces. *Journal of Computational and Applied Mathematics*, 157(1), 187-195.

Genz, A., and F. Bretz (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.

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

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.

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

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

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

Hasegawa, H. (2010) Analyzing tourists' satisfaction: A multivariate ordered probit approach. *Tourism Management*, 31(1), 86-97.

Herriges, J.A., D.J. Phaneuf, and J.L. Tobias (2008) Estimating demand systems when outcomes are correlated counts. *Journal of Econometrics*, 147(2), 282-298.

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

Hothorn, T., F. Bretz, and A. Genz (2001) On multivariate *t* and gauss probabilities in *R*. *R News*, 1(2), 27-29.

Jeliazkov, I., J. Graves, and M. Kutzbach (2008) Fitting and comparison of models for multivariate ordinal outcomes. *Advances in Econometrics*, 23, 115-156.

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

Keane, M. (1990) Four essays in empirical macro and labor economics. PhD Thesis, Brown University.

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

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

LaMondia, J., and C.R. Bhat (2009) A conceptual and methodological framework of leisure activity loyalty accommodating the travel context: Application of a copula-based bivariate ordered-response choice model. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

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

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

Liu, I., and A. Agresti (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., J.T. Kent, G. Hughes, and C.C. Taylor (2009) Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, 96(4), 975-982.

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

McKelvey, R., and W. Zavoina (1971) An IBM Fortran IV program to perform n-chotomous multivariate probit analysis. *Behavioral Science*, 16, 186-187.

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

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

Mitchell, J., and M. Weale (2007) The reliability of expectations reported by British households: Micro evidence from the BHPS. National Institute of Economic and Social Research discussion paper.

Molenberghs, G., and G. Verbeke (2005) *Models for Discrete Longitudinal Data*. Springer Series in Statistics, Springer Science + Business Media, Inc., New York

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

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

Scott, D.M., and K.W. Axhausen (2006) Household mobility tool ownership: modeling interactions between cars and season tickets. *Transportation*, 33(4), 311-328.

Scott, D.M., and P.S. Kanaroglou (2002) An activity-episode generation model that captures interactions between household heads: development and empirical analysis. *Transportation Research Part B*, 36(10), 875-896.

Scotti, C. (2006) A bivariate model of Fed and ECB main policy rates. International Finance Discussion Papers 875, Board of Governors of the Federal Reserve System (U.S.).

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

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

Varin, C., and P. Vidoni (2006) Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics & Data Analysis* 51(4), 2365-2373.

Varin, C., and P. Vidoni (2009) Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28(1-3), 170-185.

Varin, C., G. Host, and O. Skare (2005) Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics & Data Analysis*, 49(4), 1173-1191.

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

Zhao, Y., and H. Joe (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 – With Low Error Correlation Structure

Parameter	True Value	MSL Approach						CML Approach				Relative Efficiency	
		Parameter Estimates		Standard Error Estimates				Parameter Estimates		Standard Error Estimates			
		Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{MASE_{MSL}}{MASE_{CML}}$	$\frac{SASE_{MSL}}{MASE_{CML}}$
Coefficients													
β_{11}	0.5000	0.5167	3.34%	0.0481	0.0399	0.0014	0.0399	0.5021	0.43%	0.0448	0.0395	1.0109	1.0116
β_{21}	1.0000	1.0077	0.77%	0.0474	0.0492	0.0005	0.0492	1.0108	1.08%	0.0484	0.0482	1.0221	1.0222
β_{31}	0.2500	0.2501	0.06%	0.0445	0.0416	0.0010	0.0416	0.2568	2.73%	0.0252	0.0380	1.0957	1.0961
β_{12}	0.7500	0.7461	0.52%	0.0641	0.0501	0.0037	0.0503	0.7698	2.65%	0.0484	0.0487	1.0283	1.0311
β_{22}	1.0000	0.9984	0.16%	0.0477	0.0550	0.0015	0.0550	0.9990	0.10%	0.0503	0.0544	1.0100	1.0104
β_{32}	0.5000	0.4884	2.31%	0.0413	0.0433	0.0017	0.0434	0.5060	1.19%	0.0326	0.0455	0.9518	0.9526
β_{42}	0.2500	0.2605	4.19%	0.0372	0.0432	0.0006	0.0432	0.2582	3.30%	0.0363	0.0426	1.0149	1.0150
β_{13}	0.2500	0.2445	2.21%	0.0401	0.0346	0.0008	0.0346	0.2510	0.40%	0.0305	0.0342	1.0101	1.0104
β_{23}	0.5000	0.4967	0.66%	0.0420	0.0357	0.0021	0.0358	0.5063	1.25%	0.0337	0.0364	0.9815	0.9833
β_{33}	0.7500	0.7526	0.34%	0.0348	0.0386	0.0005	0.0386	0.7454	0.62%	0.0441	0.0389	0.9929	0.9930
β_{14}	0.7500	0.7593	1.24%	0.0530	0.0583	0.0008	0.0583	0.7562	0.83%	0.0600	0.0573	1.0183	1.0184
β_{24}	0.2500	0.2536	1.46%	0.0420	0.0486	0.0024	0.0487	0.2472	1.11%	0.0491	0.0483	1.0067	1.0079
β_{34}	1.0000	0.9976	0.24%	0.0832	0.0652	0.0017	0.0652	1.0131	1.31%	0.0643	0.0633	1.0298	1.0301
β_{44}	0.3000	0.2898	3.39%	0.0481	0.0508	0.0022	0.0508	0.3144	4.82%	0.0551	0.0498	1.0199	1.0208
β_{15}	0.4000	0.3946	1.34%	0.0333	0.0382	0.0014	0.0382	0.4097	2.42%	0.0300	0.0380	1.0055	1.0061
β_{25}	1.0000	0.9911	0.89%	0.0434	0.0475	0.0016	0.0475	0.9902	0.98%	0.0441	0.0458	1.0352	1.0358
β_{35}	0.6000	0.5987	0.22%	0.0322	0.0402	0.0007	0.0402	0.5898	1.69%	0.0407	0.0404	0.9959	0.9961
Correlation Coefficients													
ρ_{12}	0.3000	0.2857	4.76%	0.0496	0.0476	0.0020	0.0476	0.2977	0.77%	0.0591	0.0467	1.0174	1.0184
ρ_{13}	0.2000	0.2013	0.66%	0.0477	0.0409	0.0019	0.0410	0.2091	4.56%	0.0318	0.0401	1.0220	1.0231
ρ_{14}	0.2200	0.1919	12.76%	0.0535	0.0597	0.0035	0.0598	0.2313	5.13%	0.0636	0.0560	1.0664	1.0682
ρ_{15}	0.1500	0.1739	15.95%	0.0388	0.0439	0.0040	0.0441	0.1439	4.05%	0.0419	0.0431	1.0198	1.0239
ρ_{23}	0.2500	0.2414	3.46%	0.0546	0.0443	0.0040	0.0445	0.2523	0.92%	0.0408	0.0439	1.0092	1.0133
ρ_{24}	0.3000	0.2960	1.34%	0.0619	0.0631	0.0047	0.0633	0.3013	0.45%	0.0736	0.0610	1.0342	1.0372
ρ_{25}	0.1200	0.1117	6.94%	0.0676	0.0489	0.0044	0.0491	0.1348	12.34%	0.0581	0.0481	1.0154	1.0194
ρ_{34}	0.2700	0.2737	1.37%	0.0488	0.0515	0.0029	0.0516	0.2584	4.28%	0.0580	0.0510	1.0094	1.0110
ρ_{35}	0.2000	0.2052	2.62%	0.0434	0.0378	0.0022	0.0378	0.1936	3.22%	0.0438	0.0391	0.9662	0.9678
ρ_{45}	0.2500	0.2419	3.25%	0.0465	0.0533	0.0075	0.0538	0.2570	2.78%	0.0455	0.0536	0.9937	1.0034

Table 1a: (Continued) Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches – With Low Error Correlation Structure

Parameter	True Value	MSL Approach						CML Approach				Relative Efficiency	
		Parameter Estimates		Standard Error Estimates				Parameter Estimates		Standard Error Estimates			
		Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{MASE_{MSL}}{MASE_{CML}}$	$\frac{SASE_{MSL}}{MASE_{CML}}$
Threshold Parameters													
θ_1^1	-1.0000	-1.0172	1.72%	0.0587	0.0555	0.0007	0.0555	-1.0289	2.89%	0.0741	0.0561	0.9892	0.9893
θ_1^2	1.0000	0.9985	0.15%	0.0661	0.0554	0.0011	0.0554	1.0010	0.10%	0.0536	0.0551	1.0063	1.0065
θ_1^3	3.0000	2.9992	0.03%	0.0948	0.1285	0.0034	0.1285	2.9685	1.05%	0.1439	0.1250	1.0279	1.0282
θ_2^1	0.0000	-0.0172	-	0.0358	0.0481	0.0007	0.0481	-0.0015	-	0.0475	0.0493	0.9750	0.9751
θ_2^2	2.0000	1.9935	0.32%	0.0806	0.0831	0.0030	0.0831	2.0150	0.75%	0.0904	0.0850	0.9778	0.9784
θ_3^1	-2.0000	-2.0193	0.97%	0.0848	0.0781	0.0019	0.0781	-2.0238	1.19%	0.0892	0.0787	0.9920	0.9923
θ_3^2	-0.5000	-0.5173	3.47%	0.0464	0.0462	0.0005	0.0462	-0.4968	0.64%	0.0519	0.0465	0.9928	0.9928
θ_3^3	1.0000	0.9956	0.44%	0.0460	0.0516	0.0011	0.0516	1.0014	0.14%	0.0584	0.0523	0.9877	0.9879
θ_3^4	2.5000	2.4871	0.52%	0.0883	0.0981	0.0040	0.0982	2.5111	0.44%	0.0735	0.1002	0.9788	0.9796
θ_4^1	1.0000	0.9908	0.92%	0.0611	0.0615	0.0031	0.0616	1.0105	1.05%	0.0623	0.0625	0.9838	0.9851
θ_4^2	3.0000	3.0135	0.45%	0.1625	0.1395	0.0039	0.1396	2.9999	0.00%	0.1134	0.1347	1.0356	1.0360
θ_5^1	-1.5000	-1.5084	0.56%	0.0596	0.0651	0.0032	0.0652	-1.4805	1.30%	0.0821	0.0656	0.9925	0.9937
θ_5^2	0.5000	0.4925	1.50%	0.0504	0.0491	0.0017	0.0492	0.5072	1.44%	0.0380	0.0497	0.9897	0.9903
θ_5^3	2.0000	2.0201	1.01%	0.0899	0.0797	0.0017	0.0798	2.0049	0.24%	0.0722	0.0786	1.0151	1.0154
Overall mean value across parameters		-	2.21%	0.0566	0.0564	0.0022	0.0564	-	1.92%	0.0562	0.0559	1.0080	1.0092

Table 1b: Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches – With High Error Correlation Structure

Parameter	True Value	MSL Approach						CML Approach				Relative Efficiency	
		Parameter Estimates		Standard Error Estimates				Parameter Estimates		Standard Error Estimates			
		Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{MASE_{MSL}}{MASE_{CML}}$	$\frac{SASE_{MSL}}{MASE_{CML}}$
Coefficients													
β_{11}	0.5000	0.5063	1.27%	0.0300	0.0294	0.0020	0.0294	0.5027	0.54%	0.0292	0.0317	0.9274	0.9294
β_{21}	1.0000	1.0089	0.89%	0.0410	0.0391	0.0026	0.0392	1.0087	0.87%	0.0479	0.0410	0.9538	0.9560
β_{31}	0.2500	0.2571	2.85%	0.0215	0.0288	0.0017	0.0289	0.2489	0.42%	0.0251	0.0290	0.9943	0.9961
β_{12}	0.7500	0.7596	1.27%	0.0495	0.0373	0.0028	0.0374	0.7699	2.65%	0.0396	0.0395	0.9451	0.9477
β_{22}	1.0000	1.0184	1.84%	0.0439	0.0436	0.0036	0.0437	1.0295	2.95%	0.0497	0.0463	0.9419	0.9451
β_{32}	0.5000	0.5009	0.17%	0.0343	0.0314	0.0023	0.0315	0.5220	4.39%	0.0282	0.0352	0.8931	0.8955
β_{42}	0.2500	0.2524	0.96%	0.0284	0.0294	0.0021	0.0294	0.2658	6.34%	0.0263	0.0315	0.9318	0.9343
β_{13}	0.2500	0.2473	1.08%	0.0244	0.0233	0.0015	0.0234	0.2605	4.18%	0.0269	0.0251	0.9274	0.9293
β_{23}	0.5000	0.5084	1.67%	0.0273	0.0256	0.0020	0.0256	0.5100	2.01%	0.0300	0.0277	0.9221	0.9248
β_{33}	0.7500	0.7498	0.02%	0.0302	0.0291	0.0019	0.0291	0.7572	0.96%	0.0365	0.0318	0.9150	0.9170
β_{14}	0.7500	0.7508	0.11%	0.0416	0.0419	0.0039	0.0420	0.7707	2.75%	0.0452	0.0450	0.9302	0.9341
β_{24}	0.2500	0.2407	3.70%	0.0311	0.0326	0.0033	0.0327	0.2480	0.80%	0.0234	0.0363	0.8977	0.9022
β_{34}	1.0000	1.0160	1.60%	0.0483	0.0489	0.0041	0.0491	1.0000	0.00%	0.0360	0.0513	0.9532	0.9566
β_{44}	0.3000	0.3172	5.72%	0.0481	0.0336	0.0028	0.0337	0.3049	1.62%	0.0423	0.0368	0.9133	0.9165
β_{15}	0.4000	0.3899	2.54%	0.0279	0.0286	0.0026	0.0288	0.4036	0.90%	0.0274	0.0301	0.9516	0.9554
β_{25}	1.0000	0.9875	1.25%	0.0365	0.0391	0.0036	0.0393	1.0008	0.08%	0.0452	0.0398	0.9821	0.9862
β_{35}	0.6000	0.5923	1.28%	0.0309	0.0316	0.0030	0.0317	0.6027	0.45%	0.0332	0.0329	0.9607	0.9649
Correlation Coefficients													
ρ_{12}	0.9000	0.8969	0.34%	0.0224	0.0177	0.0034	0.0180	0.9019	0.21%	0.0233	0.0183	0.9669	0.9845
ρ_{13}	0.8000	0.8041	0.51%	0.0174	0.0201	0.0035	0.0204	0.8009	0.11%	0.0195	0.0203	0.9874	1.0023
ρ_{14}	0.8200	0.8249	0.60%	0.0284	0.0265	0.0061	0.0272	0.8151	0.60%	0.0296	0.0297	0.8933	0.9165
ρ_{15}	0.7500	0.7536	0.49%	0.0248	0.0243	0.0046	0.0247	0.7501	0.01%	0.0242	0.0251	0.9678	0.9849
ρ_{23}	0.8500	0.8426	0.87%	0.0181	0.0190	0.0081	0.0207	0.8468	0.38%	0.0190	0.0198	0.9606	1.0438
ρ_{24}	0.9000	0.8842	1.75%	0.0187	0.0231	0.0097	0.0251	0.9023	0.26%	0.0289	0.0244	0.9484	1.0284
ρ_{25}	0.7200	0.7184	0.22%	0.0241	0.0280	0.0072	0.0289	0.7207	0.09%	0.0295	0.0301	0.9298	0.9600
ρ_{34}	0.8700	0.8724	0.27%	0.0176	0.0197	0.0036	0.0200	0.8644	0.65%	0.0208	0.0220	0.8972	0.9124
ρ_{35}	0.8000	0.7997	0.04%	0.0265	0.0191	0.0039	0.0195	0.7988	0.15%	0.0193	0.0198	0.9645	0.9848
ρ_{45}	0.8500	0.8421	0.93%	0.0242	0.0231	0.0128	0.0264	0.8576	0.89%	0.0192	0.0252	0.9156	1.0480

Table 1b: (Continued) Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches – With High Error Correlation Structure

Parameter	True Value	MSL Approach						CML Approach				Relative Efficiency	
		Parameter Estimates		Standard Error Estimates				Parameter Estimates		Standard Error Estimates			
		Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{MASE_{MSL}}{MASE_{CML}}$	$\frac{SASE_{MSL}}{MASE_{CML}}$
Threshold Parameters													
θ_1^1	-1.0000	-1.0110	1.10%	0.0600	0.0520	0.0023	0.0520	-1.0322	3.22%	0.0731	0.0545	0.9538	0.9548
θ_1^2	1.0000	0.9907	0.93%	0.0551	0.0515	0.0022	0.0515	1.0118	1.18%	0.0514	0.0528	0.9757	0.9766
θ_1^3	3.0000	3.0213	0.71%	0.0819	0.1177	0.0065	0.1179	2.9862	0.46%	0.1185	0.1188	0.9906	0.9921
θ_2^1	0.0000	-0.0234	-	0.0376	0.0435	0.0028	0.0436	0.0010	-	0.0418	0.0455	0.9572	0.9592
θ_2^2	2.0000	2.0089	0.44%	0.0859	0.0781	0.0066	0.0784	2.0371	1.86%	0.0949	0.0823	0.9491	0.9525
θ_3^1	-2.0000	-2.0266	1.33%	0.0838	0.0754	0.0060	0.0757	-2.0506	2.53%	0.0790	0.0776	0.9721	0.9752
θ_3^2	-0.5000	-0.5086	1.73%	0.0305	0.0440	0.0030	0.0441	-0.5090	1.80%	0.0378	0.0453	0.9702	0.9725
θ_3^3	1.0000	0.9917	0.83%	0.0516	0.0498	0.0035	0.0499	0.9987	0.13%	0.0569	0.0509	0.9774	0.9798
θ_3^4	2.5000	2.4890	0.44%	0.0750	0.0928	0.0066	0.0930	2.5148	0.59%	0.1144	0.0956	0.9699	0.9724
θ_4^1	1.0000	0.9976	0.24%	0.0574	0.0540	0.0050	0.0542	1.0255	2.55%	0.0656	0.0567	0.9526	0.9566
θ_4^2	3.0000	3.0101	0.34%	0.1107	0.1193	0.0125	0.1200	3.0048	0.16%	0.0960	0.1256	0.9498	0.9550
θ_5^1	-1.5000	-1.4875	0.84%	0.0694	0.0629	0.0056	0.0632	-1.5117	0.78%	0.0676	0.0649	0.9699	0.9737
θ_5^2	0.5000	0.4822	3.55%	0.0581	0.0465	0.0041	0.0467	0.4968	0.64%	0.0515	0.0472	0.9868	0.9906
θ_5^3	2.0000	1.9593	2.03%	0.0850	0.0741	0.0064	0.0744	2.0025	0.12%	0.0898	0.0761	0.9735	0.9771
Overall mean value across parameters		-	1.22%	0.0429	0.0428	0.0044	0.0432	-	1.28%	0.0455	0.0449	0.9493	0.9621

Table 2: Evaluation of Ability to Recover "True" Parameters by the MSL and CML Approaches – The Panel Case

Parameter	True Value	MSL Approach						CML Approach				Relative Efficiency	
		Parameter Estimates		Standard Error Estimates				Parameter Estimates		Standard Error Estimates			
		Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{MSL}$)	Simulation Standard Error	Simulation Adjusted Standard Error ($SASE_{MSL}$)	Mean Estimate	Absolute Percentage Bias	Finite Sample Standard Error	Asymptotic Standard Error ($MASE_{CML}$)	$\frac{MASE_{MSL}}{MASE_{CML}}$	$\frac{SASE_{MSL}}{MASE_{CML}}$
ρ = 0.30													
β_1	1.0000	0.9899	1.01%	0.1824	0.1956	0.0001	0.1956	0.9935	0.65%	0.1907	0.1898	1.0306	1.0306
β_2	1.0000	1.0093	0.93%	0.1729	0.1976	0.0001	0.1976	1.0221	2.21%	0.1955	0.2142	0.9223	0.9223
ρ	0.3000	0.2871	4.29%	0.0635	0.0605	0.0000	0.0605	0.2840	5.33%	0.0632	0.0673	0.8995	0.8995
σ^2	1.0000	1.0166	1.66%	0.2040	0.2072	0.0002	0.2072	1.0142	1.42%	0.2167	0.2041	1.0155	1.0155
θ^1	1.5000	1.5060	0.40%	0.2408	0.2615	0.0001	0.2615	1.5210	1.40%	0.2691	0.2676	0.9771	0.9771
θ^2	2.5000	2.5129	0.52%	0.2617	0.2725	0.0002	0.2725	2.5272	1.09%	0.2890	0.2804	0.9719	0.9719
θ^3	3.0000	3.0077	0.26%	0.2670	0.2814	0.0002	0.2814	3.0232	0.77%	0.2928	0.2882	0.9763	0.9763
Overall mean value across parameters		-	1.29%	0.1989	0.2109	0.0001	0.2109	-	1.84%	0.2167	0.2159	0.9705	0.9705
ρ = 0.70													
β_1	1.0000	1.0045	0.45%	0.2338	0.2267	0.0001	0.2267	1.0041	0.41%	0.2450	0.2368	0.9572	0.9572
β_2	1.0000	1.0183	1.83%	0.1726	0.1812	0.0001	0.1812	1.0304	3.04%	0.1969	0.2199	0.8239	0.8239
ρ	0.7000	0.6854	2.08%	0.0729	0.0673	0.0001	0.0673	0.6848	2.18%	0.0744	0.0735	0.9159	0.9159
σ^2	1.0000	1.0614	6.14%	0.4634	0.4221	0.0004	0.4221	1.0571	5.71%	0.4864	0.4578	0.9220	0.9220
θ^1	1.5000	1.5192	1.28%	0.2815	0.2749	0.0002	0.2749	1.5304	2.03%	0.3101	0.3065	0.8968	0.8968
θ^2	2.5000	2.5325	1.30%	0.3618	0.3432	0.0003	0.3432	2.5433	1.73%	0.3904	0.3781	0.9076	0.9076
θ^3	3.0000	3.0392	1.31%	0.4033	0.3838	0.0003	0.3838	3.0514	1.71%	0.4324	0.4207	0.9123	0.9123
Overall mean value across parameters		-	2.06%	0.2842	0.2713	0.0002	0.2713	-	2.40%	0.3051	0.2990	0.9051	0.9051