The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Models

Chandra R. Bhat
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Phone: 512-471-4535; Fax: 512-475-8744; Email: bhat@mail.utexas.edu
and
King Abdulaziz University, Jeddah 21589, Saudi Arabia

ABSTRACT

This paper presents the basics of the composite marginal likelihood (CML) inference approach, discussing the asymptotic properties of the CML estimator and the advantages and limitations of the approach. The composite marginal likelihood (CML) inference approach is a relatively simple approach that can be used when the full likelihood function is practically infeasible to evaluate due to underlying complex dependencies. The history of the approach may be traced back to the pseudo-likelihood approach of Besag (1974) for modeling spatial data, and has found traction in a variety of fields since, including genetics, spatial statistics, longitudinal analyses, and multivariate modeling. However, the CML method has found little coverage in the econometrics field, especially in discrete choice modeling. This paper fills this gap by identifying the value and potential applications of the method in discrete dependent variable modeling as well as mixed discrete and continuous dependent variable model systems. In particular, the paper develops a blueprint (complete with matrix notation) to apply the CML estimation technique to a wide variety of discrete and mixed dependent variable models.

1. INTRODUCTION

1.1. Background

The need to accommodate underlying complex interdependencies in decision-making for more accurate policy analysis as well as for good forecasting, combined with the explosion in the quantity of data available for the multidimensional modeling of inter-related choices of a single observational unit and/or inter-related decision-making across multiple observational units, has resulted in a situation where the traditional frequentist full likelihood function becomes near impossible or plain infeasible to evaluate. As a consequence, another approach that has seen some (though very limited) use recently is the composite likelihood (CL) approach. This is an estimation technique that is gaining substantial attention in the statistics field, though there has been relatively little coverage of this method in econometrics and other fields. While the method has been suggested in the past under various pseudonyms such as quasi-likelihood (Hjort and Omre, 1994; Hjort and Varin, 2008), split likelihood (Vandekerkhove, 2005), and pseudolikelihood or marginal pseudo-likelihood (Molenberghs and Verbeke, 2005), Varin (2008) discusses reasons why the term composite likelihood is less subject to literary confusion.

At a basic level, a composite likelihood (CL) refers to the product of a set of lower-dimensional component likelihoods, each of which is a marginal or conditional density function. The maximization of the logarithm of this CL function is achieved by setting the composite score equations to zero, which are themselves linear combinations of valid lower-dimensional likelihood score functions. Then, from the theory of estimating equations, it can be shown that the CL score function (and, therefore, the CL estimator) is unbiased (see Varin *et al.*, 2011). In this paper, we discuss these theoretical aspects of CL methods, with an emphasis on an overview of developments and applications of the CL inference approach in the context of discrete dependent variable models.

The history of the CL method may be traced back to the pseudo-likelihood approach of Besag (1974) for modeling spatial data, and has found traction in a variety of fields since, including genetics, spatial statistics, longitudinal analyses, and multivariate modeling (see Varin *et al.*, 2011 and Larribe and Fearnhead, 2011 for reviews). However, the CL method has, as indicated earlier, found little coverage in the econometrics field, and it is the hope that this paper will fill this gap by identifying the value and potential applications of the method in econometrics.

1.2. Types of CL Methods

To present the types of CL methods, assume that the data originate from a parametric underlying model based on a random ($\widetilde{H} \times 1$) vector $\boldsymbol{Y}$ with density function $f(\boldsymbol{y},\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown $\widetilde{K}$-dimensional parameter vector (technically speaking, the density function $f(\boldsymbol{y},\boldsymbol{\theta})$ refers to the conditional density function $f_{Y|X}(\boldsymbol{y},\boldsymbol{\theta})$ of the random variable $\boldsymbol{Y}$ given a set of explanatory variables $\boldsymbol{X}$, though we will use the simpler notation $f(\boldsymbol{y},\boldsymbol{\theta})$ for the conditional density function). Each element of the random variable vector $\boldsymbol{Y}$ may be observed directly, or

may be observed in a truncated or censored form. Assume that the actual observation vector corresponding to $\boldsymbol{Y}$ is given by the vector $\boldsymbol{m} = (m_1, m_2, m_3, \ldots, m_{\tilde{H}})'$, some of which may take a continuous form and some of which may take a limited-dependent form. Let the likelihood corresponding to this observed vector be $L(\boldsymbol{\theta}; \boldsymbol{m})$. Now consider the situation where computing $L(\boldsymbol{\theta}; \boldsymbol{m})$ is very difficult. However, suppose evaluating the likelihood functions of a set of $\tilde{E}$ observed marginal or conditional events determined by marginal or conditional distributions of the sub-vectors of $\boldsymbol{Y}$ is easy and/or computationally expedient. Let these observed marginal events be characterized by ($A_1(\boldsymbol{m}), A_2(\boldsymbol{m}), ..., A_{\tilde{E}}(\boldsymbol{m})$). Let each event $A_e(\boldsymbol{m})$ be associated with a likelihood object $L_e(\boldsymbol{\theta}; \boldsymbol{m}) = L[\boldsymbol{\theta}; A_e(\boldsymbol{m})]$, which is based on a lower-dimensional marginal or conditional joint density function corresponding to the original high-dimensional joint density of $\boldsymbol{Y}$. Then, the general form of the composite likelihood function is as follows:

$$L_{CL}(\boldsymbol{\theta}, \boldsymbol{m}) = \prod_{e=1}^{\tilde{E}} \left[L_e(\boldsymbol{\theta}; \boldsymbol{m})\right]^{\omega_e} = \prod_{e=1}^{\tilde{E}} \left[L(\boldsymbol{\theta}; A_e(\boldsymbol{m})\right]^{\omega_e}, \tag{1.1}$$

where ω_e is a power weight to be chosen based on efficiency considerations. If these power weights are the same across events, they may be dropped. The CL estimator is the one that maximizes the above function (or equivalently, its logarithmic transformation).

The events $A_e(\boldsymbol{m})$ can represent a combination of marginal and conditional events, though composite likelihoods are typically distinguished in one of two classes: the composite conditional likelihood (CCL) or the composite marginal likelihood (CML). In this paper, we will focus on the CML method because it has many immediate applications in the econometrics field, and is generally easier to specify and estimate. However, the CCL method may also be of value in specific econometric contexts (see Mardia *et al*., 2009 and Varin *et al*., 2011 for additional details).

1.3. The Composite Marginal Likelihood (CML) Inference Approach

In the CML method, the events $A_e(\boldsymbol{m})$ represent marginal events. The CML class of estimators subsumes the usual ordinary full-information likelihood estimator as a special case. For instance, consider the case of repeated unordered discrete choices from a specific individual. Let the individual's discrete choice at time t be denoted by the index d_t, and let this individual be observed to choose alternative m_t at choice occasion t $(t = 1, 2, 3, ..., T)$. Then, one may define the observed event for this individual as the sequence of observed choices across all the T choice occasions of the individual. Defined this way, the CML function contribution of this individual

becomes equivalent to the full-information maximum likelihood function contribution of the individual:[1]

$$L^{1}_{CML}(\boldsymbol{\theta},\boldsymbol{m}) = L\ (\boldsymbol{\theta},\boldsymbol{m}) = \text{Prob}\,(d_1 = m_1, d_2 = m_2, d_3 = m_3, ..., d_T = m_T)\,. \tag{1.2}$$

However, one may also define the events as the observed choices at each choice occasion for the individual. Defined this way, the CML function is:

$$L^{2}_{CML}(\boldsymbol{\theta},\boldsymbol{m}) = \text{Prob}\,(d_1 = m_1) \times \text{Prob}(d_2 = m_2) \times \text{Prob}(d_3 = m_3) \times ... \times \text{Prob}(d_T = m_T)\,. \tag{1.3}$$

This CML, of course, corresponds to the case of independence between each pair of observations from the same individual. As we will indicate later, the above CML estimator is consistent. However, this approach, in general, does not estimate the parameters representing the dependence effects across choices of the same individual (*i.e.,* only a subset of the vector $\boldsymbol{\theta}$ is estimable). A third approach to estimating the parameter vector $\boldsymbol{\theta}$ in the repeated unordered choice case is to define the events in the CML as the pairwise observations across all or a subset of the choice occasions of the individual. For presentation ease, assume that all pairs of observations are considered. This leads to a pairwise CML function contribution of individual q as follows:

$$L^{3}_{CML}(\boldsymbol{\theta},\boldsymbol{m}) = \prod_{t=1}^{T-1} \prod_{t'=t+1}^{T} \text{Prob}\,(d_t = m_t, d_{t'} = m_{t'})\,. \tag{1.4}$$

Almost all earlier research efforts employing the CML technique have used the pairwise approach, including Apanasovich *et al*. (2008), Varin and Vidoni (2009), Bhat and Sener (2009), Bhat *et al*. (2010a), Bhat and Sidharthan (2011), Vasdekis *et al.* (2012), Ferdous and Bhat (2013), and Feddag (2013). 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). However, the pairwise approach is a good balance between statistical and computational efficiency (besides, in almost all applications, the parameters characterizing error dependency are completely identified based on the pairwise approach). Importantly, the pairwise approach is able to explicitly recognize dependencies across choice occasions in the repeated choice case through the inter-temporal pairwise probabilities.

1.4. Asymptotic Properties of the CML Estimator with many independent replicates

The asymptotic properties of the CML estimator for the case with many independent replicates may be derived from the theory of unbiased estimating functions. For ease, we will first consider the case when we have Q independent observational units (also referred to as individuals in this paper) in a sample $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3, ..., \boldsymbol{Y}_{\boldsymbol{Q}}$, each $\boldsymbol{Y}_{\boldsymbol{q}}$ (q=1,2,…,Q) being a $\widetilde{H} \times 1$ vector. That is,

[1] In the discussion below, for presentation ease, we will ignore the power weight term ω_e. In some cases, such as in a panel case with varying number of observational occasions on each observation unit, the choice of ω_e can influence estimator asymptotic efficiency considerations. But it does not affect other asymptotic properties of the estimator.

$\boldsymbol{Y_q} = (Y_{q1}, Y_{q2}, ..., Y_{q\tilde{H}})$. $\tilde{H}$ in this context may refer to multiple observations of the same variable on the same observation unit (as in the previous section) or a single observation of multiple variables for the observation unit (for example, expenditures on groceries, transportation, and leisure activities for an individual). In either case, Q is large relative to $\tilde{H}$ (the case when Q is small is considered in the next section). We consider the case when observation is made directly on each of the continuous variables Y_{qh}, though the discussion in this section is easily modified to incorporate the case when observation is made on some truncated or censored form of Y_{qh} (such as in the case of a discrete choice variable). Let the observation on the random variable $\boldsymbol{Y_q}$ be $\boldsymbol{y_q} = (y_{q1}, y_{q2}, ..., y_{q\tilde{H}})$. Define $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y_Q})$. Also, we will consider a pairwise likelihood function as the CML estimator, though again the proof is generalizable in a straightforward manner to other types of CML estimators (such as using triplets or quadruplets rather than couplets in the CML). For the pairwise case, the estimator is obtained by maximizing (with respect to the unknown parameter vector $\boldsymbol{\theta}$, which is of dimension $\tilde{K}$) the logarithm of the following function:

$$\begin{aligned} L_{CML}(\boldsymbol{\theta}, \boldsymbol{y}) &= \prod_{q=1}^{Q} \prod_{h=1}^{\tilde{H}-1} \prod_{h'=h+1}^{\tilde{H}} \text{Prob}(Y_{qh} = y_{qh}, Y_{qh'} = y_{qh'}) \\ &= \prod_{q=1}^{Q} \prod_{h=1}^{\tilde{H}-1} \prod_{h'=h+1}^{\tilde{H}} f(y_{qh}, y_{qh'}) = \prod_{q=1}^{Q} \prod_{h=1}^{\tilde{H}-1} \prod_{h'=h+1}^{\tilde{H}} L_{qhh'}, \text{ where } L_{qhh'} = f(y_{qh}, y_{qh'}) \end{aligned} \tag{1.5}$$

Under usual regularity conditions (these are the usual conditions needed for likelihood objects to ensure that the logarithm of the CML function can be maximized by solving the corresponding score equations; the conditions are too numerous to mention here, but are listed in Molenberghs and Verbeke, 2005, page 191), the maximization of the logarithm of the CML function in the equation above is achieved by solving the composite score equations given by:

$$\boldsymbol{s_{CML}}(\boldsymbol{\theta}, \boldsymbol{y}) = \nabla \log L_{CML}(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=h+1}^{\tilde{H}} \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'}) = \boldsymbol{0}, \tag{1.6}$$

where $\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'}) = \dfrac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}}$. Since the equations $\boldsymbol{s_{CML}}(\boldsymbol{\theta}, \boldsymbol{y})$ are linear combinations of valid likelihood score functions $\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})$ associated with the event probabilities forming the composite log-likelihood function, they immediately satisfy the requirement of being unbiased. While this is stated in many papers and should be rather obvious, we provide a formal proof of the unbiasedness of the CML score equations (see also Yi *et al.*, 2011). In particular, we need to prove the following:

$$E[\boldsymbol{s_{CML}}(\boldsymbol{\theta}, \boldsymbol{y})] = E\left[\sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=h+1}^{\tilde{H}} \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'}) \right] = \sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=h+1}^{\tilde{H}} E\left[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})\right] = \boldsymbol{0}, \tag{1.7}$$

where the expectation above is taken with respect to the full distribution of $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_{\tilde{H}})$. The above equality will hold if $E[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})] = \mathbf{0}$ for all pairwise combinations h and h' for each q. To see that this is the case, we write:

$$E[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})] = \int_{\boldsymbol{y}_q} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} f(\boldsymbol{y}_q) d\boldsymbol{y}_q = \int_{y_{qd}} \int_{y_{qd'}} \int_{\boldsymbol{y}_{-qdd'}} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} f(y_{qh}, y_{qh'}, \boldsymbol{y}_{-qhh'}) dy_{qh} dy_{qh'} d\boldsymbol{y}_{-qhh'}, \quad (1.8)$$

where $\boldsymbol{y}_{-qhh'}$ represents the subvector of $\boldsymbol{y}_q$ with the elements y_{qh} and $y_{qh'}$ excluded. Continuing,

$$\begin{aligned}
E[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})] &= \int_{y_{qh}} \int_{y_{qh'}} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} \int_{\boldsymbol{y}_{-qhh'}} f(y_{qh}, y_{qh'}, \boldsymbol{y}_{-qhh'}) dy_{qh} dy_{qh'} d\boldsymbol{y}_{-qhh'} \\
&= \int_{y_{qh}} \int_{y_{qh'}} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} f(y_{qh}, y_{qh'}) dy_{qh} dy_{qh'} = \int_{y_{qh}} \int_{y_{qh'}} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} L_{qhh'} dy_{qh} dy_{qh'} \\
&= \int_{y_{qh}} \int_{y_{qh'}} \frac{1}{L_{qhh'}} \frac{\partial L_{qhh'}}{\partial \boldsymbol{\theta}} L_{qhh'} dy_{qh} dy_{qh'} = \int_{y_{qh}} \int_{y_{qh'}} \frac{\partial L_{qhh'}}{\partial \boldsymbol{\theta}} dy_{qh} dy_{qh'} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{y_{qh}} \int_{y_{qh'}} L_{qhh'} dy_{qh} dy_{qh'} = \frac{\partial}{\partial \boldsymbol{\theta}} (1) = \mathbf{0}
\end{aligned} \quad (1.9)$$

Next, consider the asymptotic properties of the CML estimator. To derive these, define the mean composite score function across observation units in the sample as follows: $\boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{y}) = \frac{1}{Q} \sum_{q=1}^{Q} \boldsymbol{s}_q(\boldsymbol{\theta}, \boldsymbol{y}_q)$, where $\boldsymbol{s}_q(\boldsymbol{\theta}, \boldsymbol{y}_q) = \sum_{h=1}^{\tilde{H}-1} \sum_{h'=h+1}^{\tilde{H}} \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})$. Then, $E[\boldsymbol{s}_q(\boldsymbol{\theta}, \boldsymbol{y}_q)] = \sum_{h=1}^{\tilde{H}-1} \sum_{h'=h+1}^{\tilde{H}} E[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})] = \mathbf{0}$ for all values of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_0$ be the true unknown parameter vector value, and consider the score function at this vector value and label it as $\boldsymbol{s}_q(\boldsymbol{\theta}_0, \boldsymbol{y}_q)$. Then, when drawing a sample from the population, the analyst is essentially drawing values of $\boldsymbol{s}_q(\boldsymbol{\theta}_0, \boldsymbol{y}_q)$ from its distribution in the population with zero mean and variance given by $\boldsymbol{J} = Var[\boldsymbol{s}_q(\boldsymbol{\theta}_0, \boldsymbol{y}_q)]$, and taking the mean across the sampled values of $\boldsymbol{s}_q(\boldsymbol{\theta}_0, \boldsymbol{y}_q)$ to obtain $\boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})$. Invoking the Central Limit Theorem (CLT), we have

$$\sqrt{Q}\, \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y}) \xrightarrow{d} MVN_{\tilde{K}}(\mathbf{0}, \boldsymbol{J}) \quad (1.10)$$

where $MVN_{\tilde{K}}(.,.)$ stands for the multivariate normal distribution of $\tilde{K}$ dimensions. Next, let $\hat{\boldsymbol{\theta}}_{CML}$ be the CML estimator, so that, by design of the CML estimator, $\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{CML}, \boldsymbol{y}) = \mathbf{0}$. Expanding

$\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}, \boldsymbol{y})$ around $\boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y})$ in a first-order Taylor series, we obtain $\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}, \boldsymbol{y}) = \boldsymbol{0} = \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y}) + \nabla \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y}) \left[\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}} - \boldsymbol{\theta}_{\boldsymbol{0}}\right]$, or equivalently,

$$\sqrt{Q}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}} - \boldsymbol{\theta}_{\boldsymbol{0}}\right] = \sqrt{Q}\left[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y})\right]^{-1} \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y}). \tag{1.11}$$

From the law of large numbers (LLN), we also have that $\nabla \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y})$, which is the sample mean of $\nabla \boldsymbol{s}_{\boldsymbol{q}}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y}_{\boldsymbol{q}})$, converges to the population mean for the quantity. That is,

$$\left[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y})\right] \xrightarrow{d} \boldsymbol{H} = E\left[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_{\boldsymbol{0}}, \boldsymbol{y})\right] \tag{1.12}$$

Using Equations (1.10) and (1.12) in Equation (1.11), applying Slutsky's theorem, and assuming non-singularity of $\boldsymbol{J}$ and $\boldsymbol{H}$, we finally arrive at the following limiting distribution:

$$\sqrt{Q}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}} - \boldsymbol{\theta}_{\boldsymbol{0}}\right] \xrightarrow{d} MVN_{\tilde{K}}(\boldsymbol{0}, \boldsymbol{G}^{-1}), \text{ where } \boldsymbol{G} = \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H} \tag{1.13}$$

where $\boldsymbol{G}$ is the Godambe (1960) information matrix. Thus, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ is centered on the true parameter vector $\boldsymbol{\theta}_{\boldsymbol{0}}$. Further, the variance of $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ reduces as the number of sample points Q increases. The net result is that $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ converges in probability to $\boldsymbol{\theta}_{\boldsymbol{0}}$ as $Q \to \infty$ (with $\tilde{H}$ fixed), leading to the consistency of the estimator. In addition, $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ is normally distributed, with its covariance matrix being $\boldsymbol{G}^{-1}/Q$. However, both $\boldsymbol{J}$ and $\boldsymbol{H}$, and therefore $\boldsymbol{G}$, are functions of the unknown parameter vector $\boldsymbol{\theta}_{\boldsymbol{0}}$. But $\boldsymbol{J}$ and $\boldsymbol{H}$ may be estimated in a straightforward manner at the CML estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ as follows:

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\frac{\partial \log L_{CML,q}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{CML,q}}{\partial \boldsymbol{\theta}'}\right)\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}, \text{ where } \log L_{CML,q} = \sum_{h=1}^{\tilde{H}-1}\sum_{h'=h+1}^{\tilde{H}} \log L_{qhh'}, \tag{1.14}$$

and

$$\begin{aligned}\hat{\boldsymbol{H}} &= -\frac{1}{Q}\sum_{q=1}^{Q}\left[\nabla \boldsymbol{s}_{\boldsymbol{q}}(\boldsymbol{\theta}, \boldsymbol{y}_{\boldsymbol{q}})\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} = -\frac{1}{Q}\sum_{q=1}^{Q}\sum_{h=1}^{\tilde{H}-1}\sum_{h'=1}^{\tilde{H}}\left[\nabla \boldsymbol{s}_{qdd'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \\ &= -\frac{1}{Q}\left[\sum_{q=1}^{Q}\sum_{h=1}^{\tilde{H}-1}\sum_{h'=1}^{\tilde{H}}\frac{\partial^2 \log L_{qhh'}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}\end{aligned} \tag{1.15}$$

If computation of the second derivative is time consuming, one can exploit the second Bartlett identity (Ferguson, 1996, page 120), which is valid for each observation unit's likelihood term in the composite likelihood. That is, using the condition that

$$\boldsymbol{J}_{\boldsymbol{q}} = Var\left[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_{\boldsymbol{0}}, y_{qh}, y_{qh'}\right] = -\boldsymbol{H}_{\boldsymbol{q}} = -E\left[-\nabla \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_{\boldsymbol{0}}, y_{qh}, y_{qh'}\right] = E\left[\nabla \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_{\boldsymbol{0}}, y_{qh}, y_{qh'}\right], \tag{1.16}$$

an alternative estimate for $\hat{\boldsymbol{H}}$ is as below:

$$\hat{\boldsymbol{H}} = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{h=1}^{\tilde{H}-1}\sum_{h'=1}^{\tilde{H}} Var\left[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta_0}, y_{qh}, y_{qh'}\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}$$
$$= \frac{1}{Q}\sum_{q=1}^{Q}\sum_{h=1}^{\tilde{H}-1}\sum_{h'=1}^{\tilde{H}}\left(\left[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta_0}, y_{qh}, y_{qh'}\right]\left[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta_0}, y_{qh}, y_{qh'}\right]'\right)_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \quad (1.17)$$
$$= \frac{1}{Q}\sum_{q=1}^{Q}\sum_{h=1}^{\tilde{H}-1}\sum_{h'=1}^{\tilde{H}}\left(\left[\frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}'}\right]\right)_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}$$

Finally, the covariance matrix of the CML estimator is given by $\frac{\hat{\boldsymbol{G}}^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}$.

The empirical estimates above can be imprecise when Q is not large enough. An alternative procedure to obtain the covariance matrix of the CML estimator is to use a jackknife approach as follows (see Zhao and Joe, 2005):

$$\text{Cov}(\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}) = \frac{Q-1}{Q}\sum_{q=1}^{Q}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}^{(-q)} - \hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}\right)\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}^{(-q)} - \hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}\right)', \quad (1.18)$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}^{(-q)}$ is the CML estimator with the qth observational unit dropped from the data. However, this can get time-consuming, and so an alternative would be to use a first-order approximation for $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}^{(-q)}$ with a single step of the Newton-Raphson algorithm with $\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}$ as the starting point.

1.5. Asymptotic Properties of the CML Estimator for the Case of Very Few or No Independent Replicates

Even in the case when the data include very few or no independent replicates (as would be the case with global social or spatial interactions across all observational units in a cross-sectional data in which the dimension of $\tilde{H}$ is equal to the number of observational units and Q=1), the CML estimator will retain the good properties of being consistent and asymptotically normal as long as the data is formed by pseudo-independent and overlapping subsets of observations (such as would be the case when the social interactions taper off relatively quickly with the social separation distance between observational units, or when spatial interactions rapidly fade with geographic distance based on an autocorrelation function decaying toward zero; see Cox and Reid, 2004 for a technical discussion).[2] The same situation holds in cases with temporal processes; the CML estimator will retain good properties as long as we are dealing with a stationary time series with short-range dependence (the reader is referred to Davis and Yau, 2011 and Wang *et al.*, 2013 for additional discussions of the asymptotic properties of the CML estimator for the case of time-series and spatial models, respectively).

[2] Otherwise, there may be no real solution to the CML function maximization and the usual asymptotic results will not hold.

The covariance matrix of the CML estimator needs estimates of $\boldsymbol{J}$ and $\boldsymbol{H}$. The "bread" matrix $\boldsymbol{H}$ can be estimated in a straightforward manner using the Hessian of the negative of $\log L_{CML}(\boldsymbol{\theta})$, evaluated at the CML estimate $\hat{\boldsymbol{\theta}}$. This is because the information identity remains valid for each pairwise term forming the composite marginal likelihood. But the estimation of the "vegetable" matrix $\boldsymbol{J}$ is more involved. Further details of the estimation of the CML estimator's covariance matrix for the case with spatial data are discussed in Section 2.3.

1.6. Relative Efficiency of the CML Estimator

The CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator, because information embedded in the higher dimension components of the full information estimator are ignored by the CML estimator. This can also be formally shown by starting from the CML unbiased estimating functions $E[\boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})] = \boldsymbol{0}$, which can be written as follows (we will continue to assume continuous observation on the variable vector of interest, so that $\boldsymbol{Y}$ is a continuous variable, though the presentation is equally valid for censored and truncated observations on $\boldsymbol{Y}$):

$$E[\boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})] = \boldsymbol{0} = \int_{\boldsymbol{y}} \frac{\partial \log L_{CML}}{\partial \boldsymbol{\theta}} f(\boldsymbol{y}) d\boldsymbol{y} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = \int_{\boldsymbol{y}} \frac{\partial \log L_{CML}}{\partial \boldsymbol{\theta}} L_{ML} d\boldsymbol{y} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \tag{1.19}$$

Take the derivative of the above function with respect to $\boldsymbol{\theta}$ to obtain the following:

$$\boldsymbol{0} = \int_{\boldsymbol{y}} \frac{\partial^2 \log L_{CML}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L_{ML} d\boldsymbol{y} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} + \int_{\boldsymbol{y}} \frac{\partial \log L_{CML}}{\partial \boldsymbol{\theta}} \frac{\partial \log L_{ML}}{\partial \boldsymbol{\theta}} L_{ML} d\boldsymbol{y} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \tag{1.20}$$

$$= E[\nabla \boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})] + E\left[\boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y}) \boldsymbol{s_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y})\right],$$

where $\boldsymbol{s_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y})$ is the score function of the full likelihood. From above, we get the following:

$$\boldsymbol{H} = -E\left[\nabla \boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})\right] = \text{Cov}[\boldsymbol{s_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y}), \boldsymbol{s'_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})], \text{ and}$$

$$\boldsymbol{G} = \text{Cov}[\boldsymbol{s_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y}), \boldsymbol{s'_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y})]\left[\text{Var}(\boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y}))\right]^{-1} \text{Cov}[\boldsymbol{s_{CML}}(\boldsymbol{\theta_0}, \boldsymbol{y}), \boldsymbol{s'_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y}),] \tag{1.21}$$

Then, using the multivariate version of the Cauchy-Schwartz inequality (Lindsay, 1988), we obtain the following:

$$\boldsymbol{IFISHER} = \text{Var}[\boldsymbol{s_{ML}}(\boldsymbol{\theta_0}, \boldsymbol{y})] \geq \boldsymbol{G}. \tag{1.22}$$

Thus, from a theoretical standpoint, the difference between the regular ML information matrix (*i.e.*, $\boldsymbol{IFISHER}$) and the Godambe information matrix (*i.e.*, $\boldsymbol{G}$) is positive definite, which implies that the difference between the asymptotic variances of the CML estimator and the ML estimator is positive semi-definite (see also Cox and Reid, 2004). However, many studies have found that the efficiency loss of the CML estimator (relative to the maximum likelihood (ML)

estimator) is negligible to small in applications. These studies are either based on precise analytic computations of the information matrix ***IFISHER*** and the Godambe matrix ***G*** to compare the asymptotic efficiencies from the ML and the CML methods, or based on empirical efficiency comparisons between the ML and CML methods for specific contexts by employing a simulation design with finite sample sizes. A brief overview of these studies is presented in the next section.

1.6.1. Comparison of ML and CML Estimator Efficiencies

Examples of studies that have used precise analytic computations to compare the asymptotic efficiency of the ML and CML estimators include Cox and Reid (2004), Hjort and Varin (2008), and Mardia *et al*. (2009). Cox and Reid (2004) derive ***IFISHER*** and ***G*** for some specific situations, including the case of a sample of independent and identically distributed vectors, each of which is multivariate normally distributed with an equi-correlated structure between elements. In the simple cases they examine, they show that the loss of efficiency between ***IFISHER*** and ***G*** is of the order of 15%. They also indicate that in the specific case of Cox's (1972) quadratic exponential distribution-based multivariate binary data model, the full likelihood function and a pairwise likelihood function for binary data generated using a probit link are equivalent, showing that the composite likelihood estimator can achieve the same efficiency as that of a full maximum likelihood estimator. Hjort and Varin (2008) also study the relationship between the ***IFISHER*** and ***G*** matrices, but for Markov chain models, while Mardia *et al.* (2007) and Mardia *et al.* (2009) examine efficiency considerations in the context of multivariate vectors with a distribution drawn from closed exponential families. These studies note special cases when the composite likelihood estimator is fully efficient, though all of these are rather simplified model settings.

Several papers have also analytically studied efficiency considerations in clustered data, especially the case when each cluster is of a different size (such as in the case of spatially clustered data from different spatial regions with different numbers of observational units within each spatial cluster, or longitudinal data on observational units with each observational unit contributing a different number of sample observations). In such situations, the unweighted CML function will give more weight to clusters that contribute more sample observations than those with fewer observations. To address this situation, a weighted CML function may be used. Thus, Le Cessie and Van Houwelingen (1994) suggest, in their binary data model context, that each cluster should contribute about equally to the CML function. This may be achieved by power-weighting each cluster's CML contribution by a factor that is the inverse of the number of choice occasions minus one. The net result is that the composite likelihood contribution of each cluster collapses to the likelihood contribution of the cluster under the case of independence within a cluster. In a general correlated panel binary data context, Kuk and Nott (2000) confirmed the above result for efficiently estimating parameters not associated with dependence within clusters for the case when the correlation is close to zero. However, their analysis suggested that the unweighted CML function remains superior for estimating the correlation (within cluster) parameter. In a relatively more recent paper, Joe and Lee (2009) theoretically

studied the issue of efficiency in the context of a simple random effect binary choice model. They indicate that the weights suggested by Le Cessie and Van Houwelingen (1994) and Kuk and Nott (2000) can provide poor efficiency even for non-dependence parameters when the correlation between pairs of the underlying latent variables for the "repeated binary choices over time" case they studied is moderate to high. Based on analytic and numeric analyses using a longitudinal binary choice model with an autoregressive correlation structure, they suggest that using a weight of $(T_q - 1)^{-1}[1 + 0.5(T_q - 1)]^{-1}$ for a cluster appears to do well in terms of efficiency for all parameters and across varying dependency levels (T_q is the number of observations contributed by unit or individual q). Further, the studies by Joe and Lee (2009) and Varin and Vidoni (2006), also in the context of clustered data, suggest that the inclusion of too distant pairings in the CML function can lead to a loss of efficiency.

A precise analytic computation of the asymptotic efficiencies of the CML and full maximum likelihood approaches, as just discussed, is possible only for relatively simple models with or without clustering. This, in turn, has led to the examination of the empirical efficiency of the CML approach using simulated data sets for more realistic model contexts. Examples include Renard *et al.* (2004), Fieuws and Verbeke (2006), and Eidsvik *et al.* (2013). These studies indicate that the CML estimator performs well relative to the ML estimator. For instance, Renard *et al.* (2004) examined the performance of CML and ML estimators in the context of a random coefficients binary choice model, and found an average loss of efficiency of about 20% in the CML parameter estimates relative to the ML parameter estimates. Fiews and Verbeke (2006) examined the performance of the CML and ML estimators in the context of a multivariate linear model based on mixing, where the mixing along each dimension involves a random coefficient vector followed by a specification of a general covariance structure across the random coefficients of different dimensions. They found that the average efficiency loss across all parameters was less than 1%, and the highest efficiency loss for any single parameter was of the order of only 5%. Similarly, in simulated experiments with a spatial Gaussian process model, Eidsvik *et al.* (2013) used a spatial blocking strategy to partition a large spatially correlated space of a Gaussian response variable to estimate the model using a CML technique. They too found rather small efficiency losses because of the use of the CML as opposed to the ML estimator. However, this is an area that needs much more attention both empirically and theoretically. Are there situations when the CML estimator's loss is less or high relative to the ML estimator, and are we able to come up with some generalizable results from a theoretical standpoint that apply not just to simple models but also more realistic models used in the field? In this regard, is there a "file drawer" problem where results are not being reported when the CML estimator in fact loses a lot of efficiency? Or is the current state of reporting among scholars in the field a true reflection of the CML estimator's loss in efficiency relative to the ML? So far, the CML appears to be remarkable in its ability to pin down parameters, but there needs to be much more exploration in this important area. This opens up an exciting new direction of research and experimentation.

1.6.2. Comparison of Maximum Simulated Likelihood (MSL) and CML Estimator Efficiencies

The use of the maximum likelihood estimator is feasible for many types of models. But the estimation of many other models that incorporate analytically intractable expressions in the likelihood function in the form of integrals, such as in mixed multinomial logit models or multinomial probit models or count models with certain forms of heterogeneity or large-dimensional multivariate dependency patterns (just to list a few), require an approach to empirically approximate the intractable expression. This is usually done using simulation techniques, leading to the MSL inference approach (see Train, 2009), though quadrature techniques are also sometimes used for cases with 1-3 dimensions of integrals in the likelihood function expression. When simulation methods have to be used to evaluate the likelihood function, there is also a loss in asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator. Specifically, 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 ML estimator obtained as the inverse of the cross-product of first derivatives is theoretically positive semi-definite for finite number of draws per observation. Consequently, given that we also know that the difference between the asymptotic covariance matrices of the CML and ML estimators is theoretically positive semi-definite, it is difficult to state from a theoretical standpoint whether the CML estimator efficiency will be higher or lower than the MSL estimator efficiency. However, in a simulation comparison of the CML and MSL methods for multivariate ordered response systems, Bhat *et al.* (2010b) found that the CML estimator's efficiency was almost as good as that of the MSL estimator, but with the benefits of a very substantial reduction in computational time and much superior convergence properties. As they state "….any reduction in the efficiency of the CML approach relative to the MSL approach is in the range of non-existent to small". Paleti and Bhat (2013) examined the case of panel ordered-response structures, including the pure random coefficients (RC) model with no autoregressive error component, as well as the more general case of random coefficients combined with an autoregressive error component. The ability of the MSL and CML approaches to recover the true parameters is examined using simulated datasets. The results indicated that the performances of the MSL approach (with 150 scrambled and randomized Halton draws) and the simulation-free CML approach were of about the same order in all panel structures in terms of the absolute percentage bias (APB) of the parameters and empirical efficiency. However, the simulation-free CML approach exhibited no convergence problems of the type that affected the MSL approach. At the same time, the CML approach was about 5-12 times faster than the MSL approach for the simple random coefficients panel structure, and about 100 times faster than the MSL approach when an autoregressive error component was added. Thus, the CML appears to lose relatively little by way of efficiency, while also offering a more stable and much faster estimation approach in the panel ordered-ordered-response context. Similar results of substantial computational

efficiency and little to no finite sample efficiency loss (and sometimes even efficiency gains) have been reported by Bhat and Sidharthan (2011) for cross-sectional and panel unordered-response multinomial probit models with random coefficients (though the Bhat and Sidharthan paper actually combines the CML method with a specific analytic approximation method to evaluate the multivariate normal cumulative distribution function).

Finally, the reader will note that there is always some simulation bias in the MSL method for finite number of simulation draws, and the consistency of the MSL method is guaranteed only when the number of simulation draws rises faster than the square root of the sample size (Bhat, 2001, McFadden and Train, 2000). The CML estimator, on the other hand, is unbiased and consistent under the usual regularity conditions, as discussed earlier in Section 1.4.

1.7. Robustness of Consistency of the CML Estimator

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. Another way to look at this is that the consistency of the CML approach is predicated only on the correctness of the assumed lower dimensional distribution, and not on the correctness of the entire multivariate distribution. On the other hand, the consistency of the full likelihood estimator is predicated on the correctness of the assumed full multivariate distribution. Thus, for example, Yi *et al.* (2011) examined the performance of the CML (pairwise) approach in the case of clustered longitudinal binary data with non-randomly missing data, and found that the approach appears quite robust to various alternative specifications for the missing data mechanism. Xu and Reid (2011) provided several specific examples of cases where the CML is consistent, while the full likelihood inference approach is not.

1.8. Model Selection in the CML Inference Approach

Procedures similar to those available with the maximum likelihood approach are also available for model selection with the CML approach. The statistical test for a single parameter may be pursued using the usual t-statistic based on the inverse of the Godambe information matrix. When the statistical test involves multiple parameters between two nested models, an appealing statistic, which is also similar to the likelihood ratio test in ordinary maximum likelihood estimation, is the composite likelihood ratio test (CLRT) statistic. Consider the null hypothesis $H_0 : \boldsymbol{\tau} = \boldsymbol{\tau}_{\mathbf{0}}$ against $H_1 : \boldsymbol{\tau} \neq \boldsymbol{\tau}_{\mathbf{0}}$, where τ is a subvector of $\boldsymbol{\theta}$ of dimension $\widetilde{d}$; *i.e.*, $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\alpha}')'$. The statistic takes the familiar form shown below:

$$CLRT = 2[\log L_{CML}(\hat{\boldsymbol{\theta}}) - \log L_{CML}(\hat{\boldsymbol{\theta}}_{\mathbf{R}})], \tag{1.23}$$

where $\hat{\boldsymbol{\theta}}_{\mathbf{R}}$ is the composite marginal likelihood estimator under the null hypothesis $(\boldsymbol{\tau}_{\mathbf{0}}', \hat{\boldsymbol{\alpha}}_{\boldsymbol{CML}}'(\boldsymbol{\tau}_{\mathbf{0}}))$. More informally speaking, $\hat{\boldsymbol{\theta}}$ is the CML estimator of the unrestricted model, and $\hat{\boldsymbol{\theta}}_{\mathbf{R}}$ is the CML estimator for the restricted model. The CLRT statistic does not have a standard chi-squared asymptotic distribution. This is because the CML function that is maximized does not correspond to the parametric model from which the data originates; rather, the CML may be viewed in this regard as a "mis-specification" of the true likelihood function because of the independence assumption among the likelihood objects forming the CML function (see Kent, 1982, Section 3). To write the asymptotic distribution of the CLRT statistic, first define $[\boldsymbol{G}_{\tau}(\boldsymbol{\theta})]^{-1}$ and $[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}$ as the $\widetilde{d} \times \widetilde{d}$ submatrices of $[\boldsymbol{G}(\boldsymbol{\theta})]^{-1}$ and $[\boldsymbol{H}(\boldsymbol{\theta})]^{-1}$, respectively, which correspond to the vector $\boldsymbol{\tau}$. Then, the CLRT has the following asymptotic distribution:

$$CLRT \sim \sum_{i=1}^{\widetilde{d}} \lambda_i \widetilde{W}_i^2, \tag{1.24}$$

where $\widetilde{W}_i^2$ for $i = 1, 2, \ldots, \widetilde{d}$ are independent χ_1^2 variates and $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d$ are the eigenvalues of the matrix $[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})][\boldsymbol{G}_{\tau}(\boldsymbol{\theta})]^{-1}$ evaluated under the null hypothesis (this result may be obtained based on the (profile) likelihood ratio test for a mis-specified model; see Kent, 1982, Theorem 3.1 and the proof therein). Unfortunately, the departure from the familiar asymptotic chi-squared distribution with $\widetilde{d}$ degrees of freedom for the traditional maximum likelihood procedure is annoying. Pace *et al.* (2011) have recently proposed a way out, indicating that the following adjusted *CLRT* statistic, *ADCLRT*, may be considered to be asymptotically chi-squared distributed with $\widetilde{d}$ degrees of freedom:

$$ADCLRT = \frac{[\boldsymbol{S}_{\tau}(\boldsymbol{\theta})]'[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}[\boldsymbol{G}_{\tau}(\boldsymbol{\theta})][\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_{\tau}(\boldsymbol{\theta})}{[\boldsymbol{S}_{\tau}(\boldsymbol{\theta})]'[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_{\tau}(\boldsymbol{\theta})} \times CLRT \tag{1.25}$$

where $\boldsymbol{S}_{\tau}(\boldsymbol{\theta})$ is the $\widetilde{d} \times 1$ submatrix of $\boldsymbol{S}(\boldsymbol{\theta}) = \left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$ corresponding to the vector $\boldsymbol{\tau}$, and all the matrices above are computed at $\hat{\boldsymbol{\theta}}_{\mathbf{R}}$. The denominator of the above expression is a quadratic approximation to *CLRT,* while the numerator is a score-type statistic with an asymptotic $\chi_{\widetilde{d}}^2$ null distribution. Thus, *ADCLRT* is also very close to being an asymptotic $\chi_{\widetilde{d}}^2$ distribution under the null.

Alternatively, one can resort to parametric bootstrapping to obtain the precise distribution of the *CLRT* statistic for any null hypothesis situation. Such a bootstrapping procedure is rendered simple in the CML approach, and can be used to compute the *p*-value of the null hypothesis test. The procedure is as follows:

1. Compute the observed *CLRT* value as in Equation (1.23) from the estimation sample. Let the estimation sample be denoted as $\tilde{\boldsymbol{y}}_{\boldsymbol{obs}}$, and the observed *CLRT* value as $CLRT(\tilde{\boldsymbol{y}}_{\boldsymbol{obs}})$.
2. Generate C sample data sets $\tilde{\boldsymbol{y}}_1, \tilde{\boldsymbol{y}}_2, \tilde{\boldsymbol{y}}_3, ..., \tilde{\boldsymbol{y}}_C$ using the CML convergent values under the null hypothesis
3. Compute the *CLRT* statistic of Equation (1.23) for each generated data set, and label it as $CLRT(\tilde{\boldsymbol{y}}_c)$.
4. Calculate the *p*-value of the test using the following expression:

$$p = \frac{1 + \sum_{c=1}^{C} I\{CLRT(\tilde{\boldsymbol{y}}_c) \geq CLRT(\tilde{\boldsymbol{y}}_{\boldsymbol{obs}})\}}{C+1}, \text{ where } I\{A\} = 1 \text{ if } A \text{ is true.} \quad (1.26)$$

The above bootstrapping approach has been used for model testing between nested models in Varin and Czado (2010), Bhat *et al.* (2010b), and Ferdous *et al.* (2010).

When the null hypothesis entails model selection between two competing non-nested models, the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005) may be used. The CLIC takes the following form[3]:

$$\log L^{*}_{CML}(\hat{\boldsymbol{\theta}}) = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \quad (1.27)$$

The model that provides a higher value of CLIC is preferred.

1.9. Positive-Definiteness of the Implied Multivariate Covariance Matrix

In cases where the CML approach is used as a vehicle to estimate the parameters in a higher dimensional multivariate covariance matrix, one has to ensure that the implied multivariate covariance matrix in the higher dimensional context is positive definite. For example, consider a multivariate ordered-response model context, and let the latent variables underlying the multivariate ordered-response model be multivariate normally distributed. This symmetric covariance (correlation) matrix $\boldsymbol{\Sigma}$ has to be positive definite (that is, all the eigenvalues of the matrix should be positive, or, equivalently, the determinant of the entire matrix and every principal submatrix of $\boldsymbol{\Sigma}$ should be positive). But the CML approach does not estimate the entire correlation matrix as one single entity. However, there are three ways that one can ensure the positive-definiteness of the $\boldsymbol{\Sigma}$ matrix. The first technique is to use Bhat and Srinivasan's (2005) strategy of reparameterizing the correlation matrix $\boldsymbol{\Sigma}$ through the Cholesky matrix, and then using these Cholesky-decomposed parameters as the ones to be estimated. That is, the Cholesky of an initial positive-definite specification of the correlation matrix is taken before starting the optimization routine to maximize the CML function. Then, within the optimization procedure, one can reconstruct the $\boldsymbol{\Sigma}$ matrix, and then pick off the appropriate elements of this matrix to

[3] This penalized log-composite likelihood is nothing but the generalization of the usual Akaike's Information Criterion (AIC). In fact, when the candidate model includes the true model in the usual maximum likelihood inference procedure, the information identity holds (*i.e.*, $\boldsymbol{H(\theta) = J(\theta)}$) and the CLIC in this case is exactly the AIC [$= \log L_{ML}(\hat{\boldsymbol{\theta}}) -$ (# of model parameters)].

construct the CML function 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. 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 (for example, by replacing the negative eigenvalue components in the matrix with a small positive value, or by adding a sufficiently high positive value to the diagonals of a matrix and normalizing to obtain a correlation matrix; see Rebonato and Jaeckel, 1999, Higham, 2002, and Schoettle and Werner, 2004 for detailed discussions of these and other adjusting schemes; a review of these techniques is beyond the scope of this paper). The values of this "nearest" valid correlation matrix can be translated to the pairwise correlation estimates, and the analyst can allow the iterations to proceed and hope that the final implied convergent correlation matrix is positive-definite.

1.10. The Maximum Approximate Composite Marginal Likelihood Approach

In many application cases, the probability of observing the lower dimensional event itself in a CML approach may entail multiple dimensions of integration. For instance, in the case of a multinomial probit model with I choice alternatives per individual (assume for ease in presentation that all individuals have all I choice alternatives), and a spatial dependence structure (across individuals) in the utilities of each alternative, the CML approach involves compounding the likelihood of the joint probability of the observed outcomes of pairs of individuals. However, this joint probability itself entails the evaluation of integration of a multivariate normal cumulative distribution (MVNCD) function of dimension equal to $2 \times (I-1)$. The evaluation of such a function cannot be pursued using quadrature techniques due to the curse of dimensionality when the dimension of integration exceeds two (see Bhat, 2003). In this case, the MVNCD function evaluation for each agent has to be evaluated using simulation or other analytic approximation techniques. Typically, the MVNCD function is approximated using simulation techniques through the use of the Geweke-Hajivassiliou-Keane (GHK) simulator or the Genz-Bretz (GB) simulator, which are among the most effective simulators for evaluating the MVNCD

function (see Bhat *et al*., 2010b for a detailed description of these simulators). Some other sparse grid-based techniques for simulating the multivariate normal probabilities have also been proposed by Heiss and Winschel (2008), Huguenin *et al*. (2009), and Heiss (2010). In addition, Bayesian simulation using Markov Chain Monte Carlo (MCMC) techniques (instead of MSL techniques) have been used in the literature (see Albert and Chib, 1993, McCulloch and Rossi, 2000, and Train, 2009). However, all these MSL and Bayesian techniques require extensive simulation, are time-consuming, are not very straightforward to implement, and create convergence assessment problems as the number of dimensions of integration increases. Besides, they do not possess the simulation-free appeal of the CML function in the first place.

To accommodate the situation when the CML function itself may involve the evaluation of MVNCD functions, Bhat (2011) proposed a combination of an *analytic approximation* method to evaluate the MVNCD function with the CML function, and labeled this as the Maximum Approximate Composite Marginal Likelihood (MACML) approach. While several analytic approximations have been reported in the literature for MVNCD functions (see, for example, Solow, 1990, Joe, 1995, Gassmann *et al.*, 2002, and Joe, 2008), the one Bhat proposes for his MACML approach is based on decomposition into a product of conditional probabilities. Similar to the CML approach that decomposes a large multidimensional problem into lower level dimensional components, the analytic approximation method also decomposes the MVNCD function to involve only the evaluation of lower dimensional univariate and bivariate normal cumulative distribution functions. Thus, there is a type of conceptual consistency in Bhat's proposal of combining the CML method with the MVNCD analytic approximation. The net result is that the approximation approach is fast and lends itself nicely to combination with the CML approach. Further, unlike Monte-Carlo simulation approaches, even two to three decimal places of accuracy in the analytic approximation is generally adequate to accurately and precisely recover the parameters and their covariance matrix estimates because of the smooth nature of the first and second derivatives of the approximated analytic log-likelihood function. The MVNCD approximation used by Bhat for discrete choice mode estimation itself appears to have been first proposed by Solow (1990) based on Switzer (1977), and then refined by Joe (1995). However, the focus of the earlier studies was on computing a single MVNCD function accurately rather than Bhat's use of the approximation for choice model estimation where multiple MVNCD function evaluations are needed.

To describe the MVNCD approximation, let $(W_1, W_2, W_3, ..., W_I)$ be a multivariate normally distributed random vector with zero means, variances of 1, and a correlation matrix $\boldsymbol{\Sigma}$. Then, interest centers on approximating the following orthant probability:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, ...,\ W_I < w_I). \tag{1.28}$$

The above joint probability may be written as the product of a bivariate marginal probability and univariate conditional probabilities as follows ($I \geq 3$):

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2) \times \prod_{i=3}^{I} \Pr(W_i < w_i \mid W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, ...,\ W_{i-1} < w_{i-1}). \quad (1.29)$$

Next, define the binary indicator $\widetilde{I}_i$ that takes the value 1 if $W_i < w_i$ and zero otherwise. Then $E(\widetilde{I}_i) = \Phi(w_i)$, where $\Phi(.)$ is the univariate normal standard cumulative distribution function. Also, we may write the following:

$$\begin{aligned} \text{Cov}(\widetilde{I}_i, \widetilde{I}_j) &= E(\widetilde{I}_i \widetilde{I}_j) - E(\widetilde{I}_i)E(\widetilde{I}_j) = \Phi_2(w_i, w_j, \rho_{ij}) - \Phi(w_i)\Phi(w_j),\ i \neq j \\ \text{Cov}(\widetilde{I}_i, \widetilde{I}_i) &= Var(\widetilde{I}_i) = \Phi(w_i) - \Phi^2(w_i) \\ &= \Phi(w_i)[1 - \Phi(w_i)], \end{aligned} \quad (1.30)$$

where ρ_{ij} is the ij^{th} element of the correlation matrix $\boldsymbol{\Sigma}$. With the above preliminaries, consider the following conditional probability:

$$\begin{aligned} &\Pr(W_i < w_i \mid W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, ...,\ W_{i-1} < w_{i-1}) \\ &= E(\widetilde{I}_i \mid \widetilde{I}_1 = 1,\ \widetilde{I}_2 = 1,\ \widetilde{I}_3 = 1, ...,\ \widetilde{I}_{i-1} = 1). \end{aligned} \quad (1.31)$$

The right side of the expression may be approximated by a linear regression model, with $\widetilde{I}_i$ being the "dependent" random variable and $\widetilde{\boldsymbol{I}}_{<i} = (\widetilde{I}_1, \widetilde{I}_2, ...\widetilde{I}_{i-1})$ being the independent random variable vector.[4] In deviation form, the linear regression for approximating Equation (1.31) may be written as:

$$\widetilde{I}_i - E(\widetilde{I}_i) = \boldsymbol{\alpha}'[\widetilde{\boldsymbol{I}}_{<i} - E(\widetilde{\boldsymbol{I}}_{<i})] + \widetilde{\eta}, \quad (1.32)$$

where $\boldsymbol{\alpha}$ is the least squares coefficient vector and $\widetilde{\eta}$ is a mean zero random term. In this form, the usual least squares estimate of $\boldsymbol{\alpha}$ is given by:

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{i,<i}, \text{ where} \quad (1.33)$$

[4] This first-order approximation can be continually improved by increasing the order of the approximation. For instance, a second-order approximation would approximate the right side of Equation (1.31) by the expectation from a linear regression model that has $\widetilde{I}_i$ as the "dependent" random variable and $\breve{\boldsymbol{I}}_{<i} = (\widetilde{I}_1, \widetilde{I}_2, ...\widetilde{I}_{i-1}, \widetilde{I}_{12}, \widetilde{I}_{13}, ...I_{1,i-1}, \widetilde{I}_{23}, \widetilde{I}_{24}, ...\widetilde{I}_{2,i-1}, ...I_{i-2,i-1})$ as the independent random variable vector, where $\widetilde{I}_{i'j'} = \widetilde{I}_{i'}\widetilde{I}_{j'}$. Essentially this adds second-order interactions in the independent random variable vector (see Joe, 1995). However, doing so entails trivariate and four-variate normal cumulative distribution function (CDF) evaluations (when I >4) as opposed to univariate and bivariate normal CDF evaluations in the first-order approximation, thus increasing computational burden. As discussed in Bhat (2011) and shown in Bhat and Sidharthan (2011), the first-order approximation is more than adequate (when combined with the CML approach) for estimation of MNP models. Thus, in the rest of this paper, we will use the term approximation to refer to the first-order approximation evaluation of the MVNCD function.

$$\boldsymbol{\Omega}_{<i} = \text{Cov}(\boldsymbol{I}_{<i}, \boldsymbol{I}_{<i}) = \begin{bmatrix} \text{Cov}(\widetilde{I}_1, \widetilde{I}_1) & \text{Cov}(\widetilde{I}_1, \widetilde{I}_2) & \text{Cov}(\widetilde{I}_1, \widetilde{I}_3) & \cdots & \text{Cov}(\widetilde{I}_1, \widetilde{I}_{i-1}) \\ \text{Cov}(\widetilde{I}_2, \widetilde{I}_1) & \text{Cov}(\widetilde{I}_2, \widetilde{I}_2) & \text{Cov}(\widetilde{I}_2, \widetilde{I}_3) & \cdots & \text{Cov}(\widetilde{I}_2, \widetilde{I}_{i-1}) \\ \text{Cov}(\widetilde{I}_3, \widetilde{I}_1) & \text{Cov}(I_3, I_2) & \text{Cov}(\widetilde{I}_3, \widetilde{I}_3) & \cdots & \text{Cov}(\widetilde{I}_3, \widetilde{I}_{i-1}) \\ \vdots & & & & \\ \text{Cov}(\widetilde{I}_{i-1}, \widetilde{I}_1) & \text{Cov}(\widetilde{I}_{i-1}, \widetilde{I}_2) & \text{Cov}(\widetilde{I}_{i-1}, \widetilde{I}_3) & \cdots & \text{Cov}(\widetilde{I}_{i-1}, \widetilde{I}_{i-1}) \end{bmatrix}, \text{ and} \tag{1.34}$$

$$\boldsymbol{\Omega}_{\boldsymbol{i},<\boldsymbol{i}} = \text{Cov}(\boldsymbol{I}_{<i}, \boldsymbol{I}_{\boldsymbol{i}}) = \begin{bmatrix} \text{Cov}(\widetilde{I}_i, \widetilde{I}_1) \\ \text{Cov}(\widetilde{I}_i, \widetilde{I}_2) \\ \text{Cov}(\widetilde{I}_i, \widetilde{I}_3) \\ \vdots \\ \text{Cov}(\widetilde{I}_i, \widetilde{I}_{i-1}) \end{bmatrix}.$$

Finally, putting the estimate of $\hat{\boldsymbol{\alpha}}$ back in Equation (1.32), and predicting the expected value of $\widetilde{I}_i$ conditional on $\widetilde{\boldsymbol{I}}_{<i} = \mathbf{1}$ (*i.e.*, $\widetilde{I}_1 = 1,\ \widetilde{I}_2 = 1,\ \widetilde{I}_{i-1} = 1$), we get the following approximation for Equation (1.31):

$$\begin{aligned} &\Pr(W_i < w_i \mid W_1 < w_1,\ W_2 < w_2, ...,\ W_{i-1} < w_{i-1}) \approx \\ &\Phi(w_i) + (\boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{\boldsymbol{i},<\boldsymbol{i}})'(1 - \Phi(w_1),\ 1 - \Phi(w_2) ... 1 - \Phi(w_{i-1}))' \end{aligned} \tag{1.35}$$

This conditional probability approximation can be plugged into Equation (1.29) to approximate the multivariate orthant probability in Equation (1.28). The resulting expression for the multivariate orthant probability comprises only univariate and bivariate standard normal cumulative distribution functions.

One remaining issue is that the decomposition of Equation (1.28) into conditional probabilities in Equation (1.29) is not unique. Further, different permutations (*i.e.*, orderings of the elements of the random vector $\boldsymbol{W} = (W_1, W_2, W_3, ..., W_I)$) for the decomposition into the conditional probability expression of Equation (1.29) will lead, in general, to different approximations. One approach to resolve this is to average across the $I!/2$ permutation approximations. However, as indicated by Joe (1995), the average over a few randomly selected permutations is typically adequate for the accurate computation of the multivariate orthant probability. In the case when the approximation is used for model estimation (where the integrand in each individual's log-likelihood contribution is a parameterized function of the $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ parameters), even a single permutation of the $\boldsymbol{W}$ vector per choice occasion may suffice, as several papers in the literature have now shown (see later chapters).

2. APPLICATION TO TRADITIONAL DISCRETE CHOICE MODELS

In this section, we will develop a blueprint (complete with matrix notation) for the use of the CML inference method to estimate traditional discrete choice models. The focus will be on two specific kinds of discrete choice models: Ordered-response models and unordered-response models. In the case when there are only two alternatives to choose from (the binary choice case), the ordered-response and the unordered-response formulations collapse to the same structure. But these formulations differ when extended to the multinomial (more than two alternatives) choice case. The next section provides a brief overview of ordered-response and unordered-response model systems. Section 2.2 then focuses on aspatial specifications within each type of discrete choice model, while Section 2.3 focuses on spatial specifications. Section 2.4 discusses applications of the CML method to count models. In each of Sections 2.2, 2.3, and 2.4, we provide a list of references of applications after presenting the formulation and CML estimation approach. Doing so allows us to present the model structure and estimation without unnecessary interspersing with references. The contents of the individual sections do inevitably draw quite substantially from the corresponding references of applications. Also, many codes to estimate the models presented are available at http://www.caee.utexas.edu/prof/bhat/CODES.htm (these codes are in the GAUSS matrix programming language).

2.1. Ordered and Unordered-Response Model Systems

Ordered-response models are used when analyzing discrete outcome data with a finite number of mutually exclusive categories 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). 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 (1975) and Winship and Mare (1984) for some early expositions of the ordered-response model formulation. The reader is also referred to Greene and Hensher (2010) for a comprehensive history and treatment of the ordered-response model structure. These 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.

Unordered-response models are used when analyzing discrete outcome data with a finite number of mutually exclusive categories that do not represent any kind of ordinality. Examples include mode choice data or brand choice data or college choice data. In general, unordered-response models will include valuations (by decision-makers) of attributes that are alternative-specific. Most unordered-response models in economics and other fields are based on the

concept of utility-maximizing. That is, the attributes and individual characteristics are assumed to be translated into a latent utility index for each alternative, and the individual chooses the alternative that maximizes utility. The reader is referred to Train (2009) for a good exposition of the unordered-response model formulation.

In general, the ordered-response formulation may be viewed as originating from a decision-rule that is based on the horizontal partitioning of a single latent variable, while the unordered-response formulation may be viewed as originating from a decision-rule that is based on the vertical comparison of multiple latent variables (one each for each alternative, that represents the composite utility of each alternative) to determine the maximum. A detailed theoretical comparison of the two alternatives is provided in Bhat and Pulugurta (1998).

2.2. Aspatial Formulations

2.2.1. Ordered-Response Models

The applications of the ordered response model structure are quite widespread. The aspatial formulations of this structure may take the form of a cross-sectional univariate ordered-response probit (CUOP), a cross-sectional multivariate ordered-response probit (CMOP), or a panel multivariate ordered-response probit (PMOP). Within each of these formulations, many different versions are possible. In the discussion below, we present each formulation in turn in a relatively general form.

2.2.1.1 The CUOP Model

Most applications of the ordered-response model structure are confined to the analysis of a single outcome at one point in time (that is, a cross-sectional analysis). Let q be an index for observation units or individuals (q = 1, 2,…, Q, where Q denotes the total number of individuals in the data set), and let k be the index for ordinal outcome category (k =1, 2,…, K). Let the actual observed discrete (ordinal) level for individual q be m_q (m_q may take one of the K values; *i.e.*, $m_q \in \{1, 2,\ldots, K\}$). In the usual ordered response framework notation, we may write the latent propensity (y_q^*) for the ordered-response variable as a function of relevant covariates and relate this latent propensity to the ordinal outcome categories through threshold bounds:

$$y_q^* = \boldsymbol{\beta}_q' \boldsymbol{x}_q + \varepsilon_q, y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{2.1}$$

where $\boldsymbol{x}_q$ is an ($L\times1$) vector of exogenous variables (not including a constant), $\boldsymbol{\beta}_q$ is a corresponding ($L\times1$) vector of individual-specific coefficients to be estimated, ε_q is an idiosyncratic random error term that we will assume in the presentation below is independent of the elements of the vectors $\boldsymbol{\beta}_q$ and $\boldsymbol{x}_q$, and $\psi_{q,k}$ is the individual-specific upper bound threshold for discrete level k ($\psi_{q,0} = -\infty$ and $\psi_{q,K} = \infty; -\infty < \psi_{q,1} < \psi_{q,2} < \ldots < \psi_{q,K-1} < \infty \; \forall q$ in the usual ordered response fashion). The ε_q terms are assumed independent and identical across

individuals. The typical assumption for ε_q is that it is either normally or logistically distributed, though non-parametric or mixtures-of-normal distributions may also be considered. In this paper, we will consider a normal distribution for ε_q, because this has substantial benefits in estimation when $\boldsymbol{\beta}_q$ is also considered to be multivariate normally distributed (or skew normally distributed, or mixtures of normal distributed). For identification reasons, the variance of ε_q is normalized to one.[5]

Next, consider that the individual-specific thresholds are parameterized as a non-linear function of a set of variables $\boldsymbol{z}_q$ (which does not include a constant), $\psi_{q,k} = f_k(\boldsymbol{z}_q)$. The non-linear nature of the functional form should ensure that (1) the thresholds satisfy the ordering condition (*i.e.*, $-\infty < \psi_{q1} < \psi_{q2} < \psi_{q,K-1} < \infty$), and (2) allows identification for any variables that are common in $\boldsymbol{x}_q$ and $\boldsymbol{z}_q$. There are several plausible reasons provided in the ordered-response literature to motivate such varying thresholds across observation units, all of which originate in the realization that the set of thresholds represents a dimension to introduce additional heterogeneity over and beyond the heterogeneity already embedded in the latent variable y_q^*. For instance, the threshold heterogeneity may be due to a different triggering mechanism (across individuals) for the translation (mapping) of the latent underlying y_q^* propensity variable to observed ordinal data or different perceptions (across respondents) of response categories in a survey. Such generalized threshold models are referred to by different names based on their motivating origins, but we will refer to them in the current paper as generalized ordered-response probit (GORP) models. Following Eluru *et al*. (2008), we parameterize the thresholds as:

$$\psi_{q,k} = \psi_{q,k-1} + \exp(\alpha_k + \boldsymbol{\gamma}_k' \boldsymbol{z}_q) \tag{2.2}$$

In the above equation, α_k is a scalar, and $\boldsymbol{\gamma}_k$ is a vector of coefficients associated with ordinal level $k = 1, 2, ..., K-1$. The above parameterization immediately guarantees the ordering condition on the thresholds for each and every individual, while also enabling the identification of parameters on variables that are common to the $\boldsymbol{x}_q$ and $\boldsymbol{z}_q$ vectors. For identification reasons, we adopt the normalization that $\psi_{q,1} = \exp(\alpha_1)$ for all q (equivalently, all elements of the vector $\boldsymbol{\gamma}_1$ are normalized to zero, which is innocuous as long as the vector $\boldsymbol{x}_q$ is included in the risk propensity equation).

Finally, to allow for unobserved response heterogeneity among observations, the parameter $\boldsymbol{\beta}_q$ is defined as a realization from a multivariate normal distribution with mean

[5] The exclusion of a constant in the vector $\boldsymbol{x}_q$ of Equation (2.1) is an innocuous normalization as long as all the intermediate thresholds (ψ_1 through ψ_{K-1}) are left free for estimation. Similarly, the use of the standard normal distribution rather than a non-standard normal distribution for the error term is also an innocuous normalization (see Zavoina and McKelvey, 1975; Greene and Hensher, 2010).

vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \boldsymbol{LL}'$, where $\boldsymbol{L}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Omega}$.[6] Then, we can write $\boldsymbol{\beta}_q = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_q$, where $\widetilde{\boldsymbol{\beta}}_q \sim MVN_L(0, \boldsymbol{\Omega})$ (MVN_L represents the multivariate normal distribution of dimension L). If this multivariate distribution becomes degenerate, then $\boldsymbol{\beta}_q = \boldsymbol{b}$ $\forall q$, and the Random Coefficients-Generalized Ordered Response Probit (RC-GORP) model collapses to the Generalized Ordered Response Probit (GORP) model. Further, in the GORP model, if all elements of $\boldsymbol{\gamma}_k$ are zero for all k, the result is the standard ordered-response probit (SORP) model.

The CUOP model of Equation (2.1) may be written as:

$$y_q^* = \boldsymbol{b}'\boldsymbol{x}_q + \widetilde{\boldsymbol{\beta}}_q \boldsymbol{x}_q + \varepsilon_q, y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}. \tag{2.3}$$

Then, the latent variable is univariate normally distributed as $y_q^* \sim N(B_q, \sigma_q^2)$, where

$$B_q = \boldsymbol{b}'\boldsymbol{x}_q \text{ and } \sigma_q^2 = \boldsymbol{x}_q \boldsymbol{\Omega} \boldsymbol{x}_q' + 1. \tag{2.4}$$

Estimation is straightforward in this case using the maximum likelihood method. The parameter vector to be estimated in the model is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \delta, \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$, where $\overline{\boldsymbol{\Omega}}$ is a column vector obtained by vertically stacking the upper triangle elements of the matrix $\boldsymbol{\Omega}$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_2', \boldsymbol{\gamma}_3', ..., \boldsymbol{\gamma}_{I-1}')'$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_{K-1})'$. The likelihood function $L(\boldsymbol{\theta})$ for the CUOP model takes the following form:

$$L(\boldsymbol{\theta}) = \prod_{q=1}^{Q} P(y_q = m_q) = \prod_{q=1}^{Q} \left\{ \left[\Phi\left(\frac{\psi_{q,m_q} - B_q}{\sigma_q^2} \right) \right] - \left[\Phi\left(\frac{\psi_{q,m_q-1} - B_q}{\sigma_q^2} \right) \right] \right\}, \tag{2.5}$$

where $\Phi(.)$ is the univariate cumulative standard normal distribution function. To ensure the positive definiteness of the covariance matrix $\boldsymbol{\Omega}$, the likelihood function is rewritten in terms of the Cholesky-decomposed matrix $\boldsymbol{L}$ of $\boldsymbol{\Omega}$. The maximum simulated likelihood approach then proceeds by optimizing with respect to the elements of $\boldsymbol{L}$ rather than $\boldsymbol{\Omega}$. Once convergence is achieved, the implied covariance matrix $\boldsymbol{\Omega}$ may be reconstructed from the estimated matrix $\boldsymbol{L}$.

The estimation of the CUOP model presented above is very straightforward, and there have been many applications of the model or its more restrictive variants. In addition, there is a sprinkling of applications associated with two and three correlated ordered-response outcomes. Studies of two correlated ordered-response outcomes include Scotti (2006), Mitchell and Weale (2007), Scott and Axhausen (2006), and LaMondia and Bhat (2011). The study by Scott and Kanaroglou (2002) represents an example of three correlated ordered-response outcomes. But the

[6] For ease of presentation, we will treat all elements of $\boldsymbol{\beta}_q$ as random, but this is not necessary; the researcher can fix some elements of $\boldsymbol{\beta}_q$ and let the remaining elements be random. Also, it should be noted that, while random coefficients on exogenous variables can be estimated with cross-sectional data, it is generally easier to estimate random coefficients with panel or repeated-choice data where the random coefficients on the exogenous variables are specified to be individual-specific and the overall residual error term is specified to be choice-occasion specific.

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, there are many instances when interest may be centered around analyzing more than three ordered-response outcomes simultaneously, such as in the case of the number of episodes of each of several activity purposes, 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 has been to apply numerical simulation techniques based on a maximum simulated likelihood (MSL) approach (for example, see Bhat and Zhao, 2002, Greene, 2009, and Greene and Hensher, 2010) or a Bayesian inference approach (for example, see Müller and Czado, 2005 and Girard and Parent, 2001). 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 (see Bhat *et al.* 2010a and Müller and Czado, 2005). As a consequence, another approach that has seen some (though very limited) use recently is the composite marginal likelihood (CML) approach, as discussed next.

References for the CUOP Model

There have been many applications of the cross-sectional generalized ordered-response model. The reader is referred to Greene and Hensher (2010) and Eluru *et al.* (2008).

2.2.1.2. The CMOP Model

In many cases, a whole set of ordinal variables may be inter-related due to unobserved factors. For instance, the injury severity levels sustained by the occupants of a vehicle in a specific crash may be inter-related due to unobserved crash factors (in addition to being related due to observed crash factors), as may be the injury severity level of all occupants across all vehicles involved in a crash. Similarly, the evaluation ratings of a student of a professor on multiple dimensions (such as “interest in student learning”, “course well communicated”, and “tests returned promptly) may also be correlated. The estimation of such multivariate ordered outcome models are discussed in this section.

As earlier, 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 k_i be the index for ordinal outcome category ($k_i = 1,2,...,K_i$).. Let the actual 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:

$$y_{qi}^{*} = \boldsymbol{\beta}_{\boldsymbol{qi}}'\boldsymbol{x}_{\boldsymbol{q}} + \varepsilon_{qi}, y_{qi} = k_i \text{ if } \psi_{q,k_i-1}^{i} < y_{qi}^{*} < \psi_{q,k_i}^{i}, \tag{2.6}$$

where all notations are as earlier except for the addition of the index i. Define $\boldsymbol{y}_{\boldsymbol{q}}^{*} = (y_{q1}^{*}, y_{q2}^{*},..., y_{qI}^{*})'$, $\widetilde{\mathbf{x}}_{\boldsymbol{q}} = \mathbf{IDEN}_I \otimes \boldsymbol{x}_{\boldsymbol{q}}'$ ($I \times IL$ matrix; $\mathbf{IDEN}_I$ is an identity matrix of size I), $\boldsymbol{\beta}_{qi} = \boldsymbol{b}_{\boldsymbol{i}} + \widetilde{\boldsymbol{\beta}}_{qi}$, $\widetilde{\boldsymbol{\beta}}_q = (\widetilde{\boldsymbol{\beta}}_{q1}', \widetilde{\boldsymbol{\beta}}_{q2}', ..., \widetilde{\boldsymbol{\beta}}_{qI}')'$ ($IL \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}_{\boldsymbol{I}}', \boldsymbol{b}_{\boldsymbol{2}}', ..., \boldsymbol{b}_{\boldsymbol{I}}')'$ ($IL \times I$ vector), $\boldsymbol{\psi}_{\boldsymbol{q}}^{\mathbf{up}} = (\psi_{q,m_{q1}}^{1}, \psi_{q,m_{q2}}^{2}, ..., \psi_{q,m_{qI}}^{I})$ ($I \times 1$ vector), $\boldsymbol{\psi}_{\boldsymbol{q}}^{\mathbf{low}} = (\psi_{q,m_{q1}-1}^{1}, \psi_{q,m_{q2}-1}^{2}, ..., \psi_{q,m_{qI}-1}^{I})$ ($I \times 1$) vector, and let $\widetilde{\boldsymbol{\beta}}_q \sim MVN_{I \times L}(0, \boldsymbol{\Omega})$. Also, let $\psi_{q,k}^{i} = \psi_{q,k-1}^{i} + \exp(\alpha_{ki} + \boldsymbol{\gamma}_{\boldsymbol{ki}}'\boldsymbol{z}_{\boldsymbol{qk}})$, and define $\boldsymbol{\gamma}_{\boldsymbol{i}} = (\boldsymbol{\gamma}_{\boldsymbol{2i}}', \boldsymbol{\gamma}_{\boldsymbol{3i}}', ..., \boldsymbol{\gamma}_{\boldsymbol{K_i\text{-}1,i}}')'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\boldsymbol{1}}', \boldsymbol{\gamma}_{\boldsymbol{2}}', ..., \boldsymbol{\gamma}_{\boldsymbol{I}}')'$, $\boldsymbol{\alpha}_{\boldsymbol{i}} = (\alpha_{1i}, \alpha_{2i}, ..., \alpha_{K_i-1})'$, and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{\boldsymbol{1}}', \boldsymbol{\alpha}_{\boldsymbol{2}}', ..., \boldsymbol{\alpha}_{\boldsymbol{I}}')'$. 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 $\varepsilon_q = (\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \ldots, \varepsilon_{qI})'$, and assume that ε_q is multivariate normal distributed with a mean vector of zeros and a correlation matrix as follows:

$$\varepsilon_q \sim N\left[\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1I}\\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2I}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho_{I1} & \rho_{I2} & \rho_{I3} & \cdots & 1\end{pmatrix}\right], \text{ or} \tag{2.7}$$

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

The off-diagonal terms of $\boldsymbol{\Sigma}$**,** along with the covariance matrix $\boldsymbol{\Omega}$, 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 covariances between pairs of observed ordered-response variables. Then, we can write: $\boldsymbol{y}_{\boldsymbol{q}}^{*} \sim MVN_I(\boldsymbol{B}_{\boldsymbol{q}}, \boldsymbol{\Xi}_{\boldsymbol{q}})$, where $\boldsymbol{B}_{\boldsymbol{q}} = \widetilde{\mathbf{x}}_q\boldsymbol{b}$ and $\boldsymbol{\Xi}_{\boldsymbol{q}} = \widetilde{\mathbf{x}}_q\boldsymbol{\Omega}\widetilde{\mathbf{x}}_q' + \boldsymbol{\Sigma}$. Let the vector of actual observed ordinal outcomes for individual q be stacked into an ($I \times 1$) vector $\boldsymbol{m}_q = (m_{q1}, m_{q2}, \ldots, m_{qI})'$. Also let $\boldsymbol{y}_q = (y_{q1}, y_{q2}, \ldots, y_{qI})'$. The parameter vector to be estimated in the CMOP model is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Sigma}}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$. The likelihood function for individual q takes the following form:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y}_q = \boldsymbol{m}_q) = \int_{D_{y_q^*}} f_I(\boldsymbol{y}_q^* \mid \boldsymbol{B}, \boldsymbol{\Xi}) d\boldsymbol{y}_q^*, \tag{2.8}$$

where $D_{y_q^*}$ is the integration domain defined as $D_{y_q^*} = \{\boldsymbol{y}_q^* : \boldsymbol{\psi}_q^{\text{low}} < \boldsymbol{y}_q^* < \boldsymbol{\psi}_q^{\text{low}}\}$, and $f_I(.)$ is the multivariate normal density function of dimension I . The likelihood function above involves I-dimensional rectangular integrals for each individual q.

As indicated earlier, models that require integration of more than three dimensions (I >3) in a multivariate ordered-response model are typically estimated using maximum simulation likelihood (MSL) approaches. 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. Yet another MSL method to approximate the MVNCD function in the likelihood functions of Equation (2.8) is based on the Genz-Bretz (GB) algorithm (see Bhat *et al*., 2010b for a discussion). Alternatively, Chen and Dey (2000), Herriges *et al.* (2008), Jeliazkov *et al.* (2008), and Hasegawa (2010) have considered a Bayesian estimation approach for the multivariate ordered response system through the use of standard Markov Chain Monte Carlo (MCMC) techniques. 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 (see Müller and Czado, 2005). In this regard, both the MSL and the Bayesian approaches are "brute force" simulation techniques that are not very straightforward to implement and can create numerical stability, convergence, and precision problems as the number of dimensions increase.

The CML estimation of the CMOP model, on the other hand, can be very effective and fast. In particular, the pairwise likelihood function for individual q is formed by the product of likelihood contributions of pairs of ordinal variables as follows:

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

$$= \left(\prod_{i=1}^{I-1} \prod_{g=j+1}^{I} \left[\begin{array}{l} \Phi_2(\varphi_{q,m_{qi}}^i, \varphi_{q,m_{qg}}^g, \rho_{qig}) \quad - \Phi_2(\varphi_{q,m_{qi}}^i, \varphi_{q,m_{qg}-1}^g, \rho_{qig}) \\ - \Phi_2(\varphi_{q,m_{qi}-1}^i, \varphi_{q,m_{qg}}^g, \rho_{qig}) + \Phi_2(\varphi_{q,m_{qi}-1}^i, \varphi_{q,m_{qg}-1}^g, \rho_{qig}) \end{array} \right] \right) \tag{2.9}$$

where $\Phi_2(.,.,\rho_{qig})$ is the standard bivariate normal cumulative distribution function with correlation ρ_{qig}, $\varphi_{q,m_{qi}}^i = \dfrac{\psi_{q,m_{qi}}^i - \boldsymbol{b}' \boldsymbol{x}_{qi}}{\sqrt{\text{Var}\left(y_{qi}^*\right)}}$, $\rho_{qjg} = \dfrac{\text{Cov}\left(y_{qi}^*, y_{qg}^*\right)}{\sqrt{\text{Var}\left(y_{qi}^*\right)}\sqrt{\text{Var}\left(y_{qg}^*\right)}}$, and the $\text{Var}\left(y_{qi}^*\right)$,

$\text{Var}(y^*_{qg})$, and $\text{Cov}(y^*_{qi}, y^*_{qg})$ terms are obtained by picking off the appropriate 2×2 sub-matrix of the larger covariance matrix $\boldsymbol{\Xi}_q$ of $(y^*_{q1}, y^*_{q2}, ..., y^*_{qI})$. The pairwise marginal likelihood function is $L^{CMOP}_{CML}(\boldsymbol{\theta}) = \prod_q L^{CMOP}_{CML,q}(\boldsymbol{\theta})$.

The asymptotic covariance matrix estimator is $\frac{\hat{\boldsymbol{G}}^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}$, with

$$\hat{\boldsymbol{H}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q}\sum_{i=1}^{I-1}\sum_{g=i+1}^{I}\frac{\partial^2 \log L^{CMOP}_{CML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{CML}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q}\sum_{i=1}^{I-1}\sum_{g=i+1}^{I}\frac{\partial^2 \log \Pr(y_{qi}=m_{qi}, y_{qg}=m_{qg})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{CML}}$$

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\sum_{i=1}^{I-1}\sum_{g=i+1}^{I}\frac{\partial \log \Pr(y_{qi}=m_{qi}, y_{qg}=m_{qg})}{\partial\boldsymbol{\theta}}\right)\left(\sum_{i=1}^{I-1}\sum_{g=i+1}^{I}\frac{\partial \log \Pr(y_{qi}=m_{qi}, y_{qg}=m_{qg})}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\theta}_{CML}} \quad (2.10)$$

An alternative estimator for $\hat{\boldsymbol{H}}$ is as below:

$$\hat{\boldsymbol{H}} = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{I-1}\sum_{g=i+1}^{I}\left(\left[\frac{\partial \log \Pr(y_{qi}=m_{qi}, y_{qg}=m_{qg})}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log \Pr(y_{qi}=m_{qi}, y_{qg}=m_{qg})}{\partial\boldsymbol{\theta}'}\right]\right)_{\hat{\theta}_{CML}} \quad (2.11)$$

One final issue. The covariance matrix $\boldsymbol{\Xi}$ has to be positive definite, which will be the case if the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are positive definite. The simplest way to ensure the positive-definiteness of these matrices is to use a Cholesky-decomposition and parameterize the CML function in terms of the Cholesky parameters (rather than the original covariance matrices). Also, the matrix $\boldsymbol{\Sigma}$ is a correlation matrix, which can be maintained by writing each diagonal element (say the aa^{th} element) of the lower triangular Cholesky matrix of $\boldsymbol{\Sigma}$ as $\sqrt{1-\sum_{j=1}^{a-1} l_{aj}^2}$, where the l_{aj} elements are the Cholesky factors that are estimated.

References for the CML Estimation of the CMOP Model

Archer, M., Paleti, R., Konduri, K.C., Pendyala, R.M., Bhat, C.R., 2013. Modeling the connection between activity-travel patterns and subjective well-being. *Transportation Research Record* 2382, 102-111.

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

Feddag, M.-L., 2013. Composite likelihood estimation for multivariate probit latent traits models. *Communications in Statistics - Theory and Methods* 42(14), 2551-2566.

Katsikatsou, M., Moustaki, I, Yang-Wallentin, F., and Jöreskog, K.G., 2012. Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis* 56(12), 4243-4258.

LaMondia, J.J., Bhat, C.R., 2011. A study of visitors' leisure travel behavior in the northwest territories of Canada, *Transportation Letters: The International Journal of Transportation Research* 3(1), 1-19.

Seraj, S., Sidharthan, R., Bhat, C.R., Pendyala, R.M., Goulias, K.G., 2012. Parental attitudes towards children walking and bicycling to school. *Transportation Research Record* 2323, 46-55.

2.2.1.3. The PMOP Model

As earlier, let q be an index for individuals (q = 1, 2, …, Q), and let t be an index for the t^{th} observation on individual q (t = 1, 2, …, T, where T denotes the total number of observations on individual q).[7] Let the observed discrete (ordinal) level for individual q at the t^{th} observation be m_{qt} (m_{qt} may take one of K values; *i.e.*, $m_{qt} \in \{1, 2,\ldots, K\}$). In the usual random-coefficients ordered response framework notation, we write the latent variable (y^*_{qt}) as a function of relevant covariates as:

$$y^*_{qt} = \boldsymbol{\beta}'_{\boldsymbol{q}} \boldsymbol{x}_{\boldsymbol{qt}} + \varepsilon_{qt}, y_{qt} = k \text{ if } \psi_{q,t,k-1} < y^*_{qj} < \psi_{q,t,k}, \quad (2.12)$$

where $\boldsymbol{x}_{\boldsymbol{qt}}$ is a ($L\times1$)-vector of exogenous variables (including a constant now), $\boldsymbol{\beta}_{\boldsymbol{q}}$ is an individual-specific ($L\times1$)-vector of coefficients to be estimated that is a function of unobserved individual attributes, ε_{qt} is a standard normal error term uncorrelated across individuals q and across observations of the same individual, and $\psi_{q,t,k}$ is the upper bound threshold for ordinal discrete level k (k=1,2,…,K) for individual q at choice occasion t. The thresholds are written as $\psi_{q,t,k} = \psi_{q,t,k-1} + \exp(\alpha_k + \boldsymbol{\gamma}'_{\boldsymbol{k}} \boldsymbol{z}_{qt})$ for k=2,3,…,K-1, with $\psi_{q,t,0} < \psi_{q,t,1} < \psi_{q,t,2} \ldots < \psi_{q,t,K-1} < \psi_{q,t,K}$; $\psi_{q,t,0} = -\infty, \psi_{q,t,1} = 0, \psi_{q,t,K} = +\infty$. Assume that the $\boldsymbol{\beta}_{\boldsymbol{q}}$ vector in Equation (2.12) is a time-invariant realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \boldsymbol{LL}'$, where $\boldsymbol{L}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Omega}$.[8] Also, assume that the ε_{qt} term, which captures the idiosyncratic effect of all omitted variables for individual q at the t^{th} choice occasion, is independent of the elements of the

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

[8] More general autoregressive structures can also be considered for ε_{qt} and $\boldsymbol{\beta}_{\boldsymbol{q}}$ to accommodate fading and time-varying covariance effects in the latent variables y^*_{qt} (see Bhat, 2011 and Paleti and Bhat, 2013). This does not complicate the econometrics of the CML estimation method, but can lead to substantial number of additional parameters and may be asking too much from typical estimation data sets. In this paper, we present the case of independent ε_{qt} across choice occasions and time-invariant random coefficients.

$\boldsymbol{\beta}_q$ and $\boldsymbol{x}_{qt}$ vectors. Define $\boldsymbol{y}_q = (y_{q1}, y_{q2}, ..., y_{qT})'$ ($T \times 1$ matrix), $\boldsymbol{\varepsilon}_q = (\varepsilon_{q1}, \varepsilon_{q2}, ..., \varepsilon_{qT})'$ ($T \times 1$ matrix) $\boldsymbol{y}_q^* = (y_{q1}^*, y_{q2}^*, ..., y_{qT}^*)'$ ($T \times 1$ matrix), $\boldsymbol{x}_q = (\boldsymbol{x}_{q1}, \boldsymbol{x}_{q2}, ..., \boldsymbol{x}_{qT})'$ ($T \times L$ matrix), $\boldsymbol{\psi}_q^{\mathbf{up}} = (\psi_{q,1,m_{q1}}, \psi_{q,2,m_{q2}}, ..., \psi_{q,T,m_{qT}})$ ($T \times 1$ vector), $\boldsymbol{\psi}_q^{\mathbf{low}} = (\psi_{q,1,m_{qi}-1}, \psi_{q,2,m_{q2}-1}, ..., \psi_{q,T,m_{qJ}-1})$ ($T \times 1$) vector. Also, let the vector of actual observed ordinal outcomes for individual q be stacked into a ($T \times 1$) vector $\boldsymbol{m}_q = (m_{q1}, m_{q2}, ..., m_{qT})'$. Then, we may write $\boldsymbol{y}_q^* \sim MVN_T(\boldsymbol{B}_q, \boldsymbol{\Xi}_q)$, where $\boldsymbol{B}_q = \boldsymbol{x}_q \boldsymbol{b}$ and $\boldsymbol{\Xi}_q = (\boldsymbol{x}_q \boldsymbol{\Omega} \boldsymbol{x}_q' + \mathbf{IDEN}_T)$, and the parameter vector to be estimated in the PMOP model is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$, where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_2', \boldsymbol{\gamma}_3', ..., \boldsymbol{\gamma}_{K-1}')'$ and $\boldsymbol{\alpha} = (\alpha_2', \alpha_3', ..., \alpha_{K-1}')'$. The likelihood function for individual q takes the following form:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y}_q = \boldsymbol{m}_q) = \int_{D_{y_q^*}} f_T(\boldsymbol{y}_q^* \mid \boldsymbol{B}_q, \boldsymbol{\Xi}_q) d\boldsymbol{y}_q^*, \tag{2.13}$$

where $D_{\boldsymbol{y}^*}$ is the integration domain defined as $D_{\mathrm{y}_q^*} = \{\boldsymbol{y}_q^* : \boldsymbol{\psi}_q^{\mathbf{low}} < \boldsymbol{y}_q^* < \boldsymbol{\psi}_q^{\mathbf{low}}\}$, and $f_T(.)$ is the multivariate normal density function of dimension T. The likelihood function above involves T-dimensional rectangular integrals for each individual q. The above model is labeled as a mixed autoregressive ordinal probit model by Varin and Czado (2010), who 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 rectangular 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. On the other hand, the CML approach is easy to apply in such situations, through a pairwise marginal likelihood approach that takes the following form:

$$L_{CML,q}^{PMOP}(\boldsymbol{\theta}) = \left(\prod_{t=1}^{T-1} \prod_{g=t+1}^{T} \left[\Pr\left(y_{qt} = m_{qt}, y_{qg} = m_{qg}\right) \right] \right)$$

$$= \left(\prod_{t=1}^{T-1} \prod_{g=t+1}^{T} \begin{bmatrix} \Phi_2(\varphi_{q,t,m_{qt}}, \varphi_{q,g,m_{qg}}, \rho_{qtg}) & -\Phi_2(\varphi_{q,t,m_{qt}}, \varphi_{q,g,m_{qg}-1}, \rho_{qtg}) \\ -\Phi_2(\varphi_{q,t,m_{qt}-1}, \varphi_{q,g,m_{qg}}, \rho_{qtg}) + \Phi_2(\varphi_{q,t,m_{qt}-1}, \varphi_{q,g,m_{qg}-1}, \rho_{qtg}) \end{bmatrix} \right) \tag{2.14}$$

where $\varphi_{q,t,m_{qt}} = \dfrac{\psi_{q,t,m_{qt}} - \boldsymbol{b}' \boldsymbol{x}_{qt}}{\sqrt{\mathrm{Var}\left(y_{qt}^*\right)}}$ and $\rho_{qtg} = \dfrac{\mathrm{Cov}\left(y_{qt}^*, y_{qg}^*\right)}{\sqrt{\mathrm{Var}\left(y_{qt}^*\right)}\sqrt{\mathrm{Var}\left(y_{qg}^*\right)}}$

In the above expression, the $\mathrm{Var}\left(y_{qt}^*\right)$, $\mathrm{Var}\left(y_{qg}^*\right)$, and $\mathrm{Cov}\left(y_{qt}^*, y_{qg}^*\right)$ terms are obtained by picking off the appropriate (2×2)-sub-matrix of the larger covariance matrix $\boldsymbol{\Xi}_q$ of $(y_{q1}^*, y_{q2}^*, ..., y_{qT}^*)$. The pairwise marginal likelihood function is $L_{CML}^{PMOP}(\boldsymbol{\theta}) = \prod_q L_{CML,q}^{PMOP}(\boldsymbol{\theta})$. The covariance matrix of the estimator can be obtained exactly as in the CMOP case.

The analysis above assumes the presence of a balanced panel; that is, it assumes the same number of choice instances per individual. In the case when the number of choice instances varies across individuals, Joe and Lee (2009) proposed placing a power weight for individual q as $w_q = (T_q - 1)^{-1}[1 + 0.5(T_q - 1)]^{-1}$ (where the number of observations from individual q is T_q) and constructing the marginal likelihood contribution of individual q as:

$$L_{CML_q}^{PMOP}(\boldsymbol{\theta}) = \left(\prod_{t=1}^{T-1} \prod_{g=t+1}^{J} \left[\begin{array}{l} \Phi_2(\varphi_{q,t,m_{qj}}, \varphi_{q,g,m_{qg}}, \rho_{qtg}) \quad - \Phi_2(\varphi_{q,t,m_{qj}}, \varphi_{q,g,m_{qg}-1}, \rho_{qtg}) \\ - \Phi_2(\varphi_{q,t,m_{qj}-1}, \varphi_{q,g,m_{qg}}, \rho_{qtg}) + \Phi_2(\varphi_{q,t,m_{qj}-1}, \varphi_{q,g,m_{qg}-1}, \rho_{qtg}) \end{array} \right] \right)^{w_q} \tag{2.15}$$

References for the CML Estimation of the PMOP Model

Paleti, R., Bhat, C.R., 2013. The composite marginal likelihood (CML) estimation of panel ordered-response models. *Journal of Choice Modelling* 7, 24-43.

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

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

Vasdekis, V.G.S., Cagnone, S., Moustaki, I., 2012. A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika* 77(3), 425-441.

2.2.2. Unordered-Response Models

In the class of unordered-response models, the "workhorse" multinomial logit model introduced by Luce and Suppes (1965) and McFadden (1974) has been used extensively in practice for econometric discrete choice analysis, and has a very simple and elegant structure. However, it is also saddled with the familiar independence from irrelevant alternatives (IIA) property – that is, the ratio of the choice probabilities of two alternatives is independent of the characteristics of other alternatives in the choice set. This has led to several extensions of the MNL model through the relaxation of the independent and identically distributed (IID) error distribution (across alternatives) assumption. Two common model forms of non-IID error distribution include the generalized extreme-value (GEV) class of models proposed by McFadden (1978) and the multinomial probit (MNP) model that allows relatively flexible error covariance structures (up to certain limits of identifiability; see Train, 2009, Chapter 5). Both of these non-IID kernel structures (or even the IID versions of the GEV and the MNP models, which lead to the MNL and the independent MNP models, respectively) can further be combined with continuous mixing error structures. While many different continuous distributions can be used to accommodate these additional structures, it is most common to adopt a normal distribution. For instance, when introducing random coefficients, it is typical to use the multivariate normal distribution for the mixing coefficients, almost to the point that the terms mixed logit or mixed

GEV or mixed probit are oftentimes used synonymously with normal mixing (see Fiebig *et al*., 2010, Dube *et al*., 2002).[9]

In the context of the normal error distributions just discussed, the use of a GEV kernel structure leads to a mixing of the normal distribution with a GEV kernel, while the use of an MNP kernel leads once again to an MNP model. Both structures have been widely used in the past, with the choice between a GEV kernel or an MNP kernel really being a matter of "which is easier to use in a given situation" (Ruud, 2007). In recent years, the mixing of the normal with the GEV kernel has been the model form of choice in the economics and transportation fields, mainly due to the relative ease with which the probability expressions in this structure can be simulated (see Bhat *et al*., 2008 and Train, 2009 for detailed discussions). On the other hand, the use of an MNP kernel has not seen as much use in recent years, because the simulation estimation is generally more difficult. In any case, while there have been several approaches proposed to simulate these models with a GEV or an MNP kernel, most of these involve pseudo-Monte Carlo or quasi-Monte Carlo simulations in combination with a quasi-Newton optimization routine in a maximum simulated likelihood (MSL) inference approach (see Bhat, 2001, 2003). As has been discussed earlier, in such an inference approach, consistency, efficiency, and asymptotic normality of the estimator is critically predicated on the condition that the number of simulation draws rises faster than the square root of the number of individuals in the estimation sample. Unfortunately, for many practical situations, the computational cost to ensure good asymptotic estimator properties can be prohibitive and literally infeasible (in the context of the computation resources available and the time available for estimation) as the number of dimensions of integration increases.

The Maximum Approximate Composite Marginal Likelihood (MACML) inference approach proposed by Bhat (2011), on the other hand, allows the estimation of models with both GEV and MNP kernels using simple, computationally very efficient, and simulation-free estimation methods. In the MACML inference approach, models with the MNP kernel, when combined with additional normal random components, are much easier to estimate because of the conjugate addition property of the normal distribution (which puts the structure resulting from the addition of normal components to the MNP kernel back into an MNP form). On the other hand, the MACML estimation of models obtained by superimposing normal error components over a GEV kernel requires a normal scale mixture representation for the extreme value error terms, and adds an additional layer of computational effort (see Bhat, 2011). Given that the use of a GEV kernel or an MNP kernel is simply a matter of convenience, we will henceforth focus in this paper on the MNP kernel within the unordered-response model structure.

[9] It has been well known that using non-normal distributions can lead to convergence/computational problems, and it is not uncommon to see researchers consider non-normal distributions only to eventually revert to the use of a normal distribution (see, for example, Bartels *et al*., 2006 and Small *et al*., 2005). However, one appealing approach is to use a multivariate skew-normal (MSN) distribution for the response surface, as proposed by Bhat and Sidharthan (2012).

The aspatial formulations of the unordered-response structure may take the form of a cross-sectional multinomial probit (CMNP), or a cross-sectional multivariate multinomial probit (CMMNP), or a panel multinomial probit (PMNP).

2.2.2.1. The CMNP Model

In the discussion below, we will assume that the number of choice alternatives in the choice set is the same across all individuals. The case of different numbers of choice alternatives per individual poses no complication, since the only change in such a case is that the dimensionality of the multivariate normal cumulative distribution (MVNCD) function changes from one individual to the next.

Consider the following specification of utility for individual q and alternative i:

$$U_{qi} = \boldsymbol{\beta}_{\boldsymbol{q}}' \boldsymbol{x}_{\boldsymbol{qi}} + \xi_{qi};\ \boldsymbol{\beta}_{\boldsymbol{q}} = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}},\ \widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}} \sim MVN_L(\mathbf{0}, \boldsymbol{\Omega}), \tag{2.16}$$

where $\boldsymbol{x}_{\boldsymbol{qi}}$ is an $(L \times 1)$-column vector of exogenous attributes (including a constant for each alternative, except one of the alternatives), and $\boldsymbol{\beta}_{\boldsymbol{q}}$ is an individual-specific $(L \times 1)$-column vector of corresponding coefficients that varies across individuals based on unobserved individual attributes. Assume that the $\boldsymbol{\beta}_{\boldsymbol{q}}$ vector is a realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \boldsymbol{LL}'$. We also assume that ξ_{qi} is independent and identically normally distributed across q, but allow a general covariance structure across alternatives for individual q. Specifically, let $\boldsymbol{\xi}_{\boldsymbol{q}} = (\xi_{q1}, \xi_{q2}, ..., \xi_{qI})'$ ($I \times 1$ vector). Then, we assume $\boldsymbol{\xi}_{\boldsymbol{q}} \sim MVN_I(0, \boldsymbol{\Lambda})$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Lambda}$ or identifiability. Specifically, only utility differentials matter in discrete choice models. Taking the utility differentials with respect to the first alternative, only the elements of the covariance matrix $\boldsymbol{\Lambda}_{\mathbf{1}}$ of $\widetilde{\xi}_{qi1} = \xi_{qi} - \xi_{q1}$ $(i \neq 1)$ are estimable. However, the MACML inference approach proposed here, like the traditional GHK simulator, takes the difference in utilities against the chosen alternative during estimation. Thus, if individual q is observed to choose alternative m_q, the covariance matrix $\boldsymbol{\Lambda}_{\boldsymbol{m_q}}$ is desired for the individual. However, even though different differenced covariance matrices are used for different individuals, they must originate in the same matrix $\boldsymbol{\Lambda}$. To achieve this consistency, $\boldsymbol{\Lambda}$ is constructed from $\boldsymbol{\Lambda}_{\mathbf{1}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. An additional normalization needs to be imposed on $\boldsymbol{\Lambda}$ because the scale is also not identified. For this, we normalize the element of $\boldsymbol{\Lambda}$ in the second row and second column to the value of one. Note that these normalizations are innocuous and are needed for identification. The $\boldsymbol{\Lambda}$ matrix so constructed is fully general. Also, in MNP models, identification is tenuous when only individual-specific covariates are used (see Keane, 1992 and Munkin and Trivedi, 2008). In particular, exclusion restrictions are needed in the form of at least one individual characteristic being excluded from

each alternative's utility in addition to being excluded from a base alternative (but appearing in some other utilities). But these exclusion restrictions are not needed when there are alternative-specific variables.

The model above may be written in a more compact form by defining the following vectors and matrices: $\mathbf{U}_q = (U_{q1}, U_{q2}, ..., U_{qI})'$ ($I \times 1$ vector), $\boldsymbol{x_q} = (\boldsymbol{x_{q1}}, \boldsymbol{x_{q2}}, \boldsymbol{x_{q3}}, ..., \boldsymbol{x_{qI}})'$ ($I \times L$ matrix), $\mathbf{V}_q = \boldsymbol{x_q b}$ ($I \times 1$ vector), $\widetilde{\boldsymbol{\Omega}}_q = \boldsymbol{x_q \Omega x_q'}$ ($I \times I$ matrix), and $\widetilde{\boldsymbol{\Xi}}_q = \widetilde{\boldsymbol{\Omega}}_q + \boldsymbol{\Lambda}$ ($I \times I$ matrix). Then, we may write, in matrix notation, $\mathbf{U}_q = \mathbf{V}_q + \boldsymbol{\xi}_q$ and $\mathbf{U}_q \sim MVN_I(\mathbf{V}_q, \widetilde{\boldsymbol{\Xi}}_q)$. Also, let $\mathbf{u}_q = (u_{q1}, u_{q2}, \ldots u_{qI})'$ ($i \neq m_q$) be an (I–1)×1 vector, where m_q is the actual observed choice of individual q, and $u_{qi} = U_{qi} - U_{qm_q}$ ($i \neq m_q$). Then, $\mathbf{u}_q < \mathbf{0}_{I-1}$, because alternative m_q is the chosen alternative by individual q.

To develop the likelihood function, define $\mathbf{M}_q$ as an identity matrix of size I-1 with an extra column of '-1' values added at the m_q^{th} column (thus, $\mathbf{M}_q$ is a matrix of dimension $(I\text{ - }1) \times (I)$). Then, $\mathbf{u}_q$ is distributed as follows: $\mathbf{u}_q \sim MVN_{I-1}(\boldsymbol{B_q}, \boldsymbol{\Xi_q})$, where $\boldsymbol{B_q} = \mathbf{M}_q \mathbf{V}_q$ and $\boldsymbol{\Xi_q} = \mathbf{M}_q \widetilde{\boldsymbol{\Xi}}_q \mathbf{M}_q'$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Lambda}}')'$. Let $\boldsymbol{\omega}_{\boldsymbol{\Xi_q}}$ be the diagonal matrix of standard deviations of $\boldsymbol{\Xi_q}$. Using the usual notations as described earlier, the likelihood contribution of individual q is as below:

$$L_q(\boldsymbol{\theta}) = \Phi_{I-1}(\boldsymbol{\omega}_{\boldsymbol{\Xi_q}}^{-1}(-\boldsymbol{B_q}), \boldsymbol{\Xi_q}^*), \tag{2.17}$$

where $\boldsymbol{\Xi_q}^* = \boldsymbol{\omega}_{\boldsymbol{\Xi_q}}^{-1} \boldsymbol{\Xi_q} \boldsymbol{\omega}_{\boldsymbol{\Xi_q}}^{-1}$.

The MVNCD approximation discussed earlier is computationally efficient and straightforward to implement when maximizing the likelihood function of Equation (2.17).[10] As such, the MVNCD approximation can be used for any value of K and any value of I, as long as there is data support for the estimation of parameters. The positive-definiteness of $\boldsymbol{\Sigma}$ can be ensured by using a Cholesky-decomposition of the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$, and estimating these Cholesky-decomposed parameters. Note that, to obtain the Cholesky factor for $\boldsymbol{\Lambda}$, we first obtain the Cholesky factor for $\boldsymbol{\Lambda}_1$, and then add a column of zeros as the first column and a row of zeros as the first row to the Cholesky factor of $\boldsymbol{\Lambda}_1$. The covariance matrix in this CMOP case is obtained using the usual Fisher information matrix, since the full (approximate) likelihood is being maximized.

Bhat and Sidharthan (2011) apply the MACML estimation approach for estimating the CMNP model with five random coefficients and five alternatives, and compare the performance

[10]As indicated earlier, the CML class of estimators subsumes the usual ordinary full-information likelihood estimator as a special case. It is this characteristic of the CML approach that leads us to the label MACML for the estimation approach proposed here. Specifically, even in cross-sectional MNP contexts, when our approach involves only the approximation of the MVNCD function in the maximum likelihood function, the MACML label is appropriate since the maximum likelihood function is a special case of the CML function.

of the MSL and MACML approaches (though, in their simulations, they constrain $\mathbf{\Lambda}$ to be an identity matrix multiplied by 0.5). They conclude that the MACML approach recovers parameters much more accurately than the MSL approach, while also being about 50 times faster than the MSL approach. They also note that as the number of random coefficients and/or alternatives in the unordered-response model increases, one can expect even higher computational efficiency factors for the MACML over the MSL approach.

References for the CML Estimation of the CMNP Model

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

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

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

2.2.2.2. The CMMNP Model

Let there be G nominal (unordered multinomial response) variables for an individual, and let g be the index for variables (g = 1, 2, 3,…, G). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 3$) and let i_g be the corresponding index (i_g = 1, 2, 3,…, I_g). Note that I_g may vary across individuals. Also, it is possible that some nominal variables do not apply for some individuals, in which case G itself is a function of the individual q. However, for presentation ease, we assume that all the G nominal variables are relevant for each individual, and that all the alternatives I_g are available for each variable g.

Consider the g^{th} variable and assume that the individual q chooses the alternative m_{qg}. Also, assume the usual random utility structure for each alternative i_g.

$$U_{qgi_g} = \boldsymbol{\beta}'_{qg} \boldsymbol{x}_{qgi_g} + \xi_{qgi_g}, \tag{2.18}$$

where $\boldsymbol{x}_{qgi_g}$ is a (L_g×1)-column vector of exogenous attributes, $\boldsymbol{\beta}_{qg}$ is a column vector of corresponding coefficients, and ξ_{qgi_g} is a normal error term. Assume that the $\boldsymbol{\beta}_{qg}$ vector is a realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}_g$ and covariance matrix $\mathbf{\Omega}_g = \boldsymbol{L}_g \boldsymbol{L}'_g$, where $\boldsymbol{L}_g$ is the lower-triangular Cholesky factor of $\mathbf{\Omega}_g$. While one can allow covariance among the $\boldsymbol{\beta}_{qg}$ vectors across the coefficients of the different unordered-response variables for each individual, this specification will be profligate in the parameters to be estimated. So, we will assume that the $\boldsymbol{\beta}_{qg}$ vectors are independent across the unordered-response

dimensions for each individual. We also assume that ξ_{qgi_g} is independent and identically normally distributed across individuals q, but allow a general covariance structure across alternatives for individual q. Specifically, let $\boldsymbol{\xi_{qg}} = (\xi_{qg1}, \xi_{qg2}, ..., \xi_{qgI_g})'$ ($I_g \times 1$ vector). Then, we assume $\boldsymbol{\xi_{qg}} \sim MVN_I(0, \boldsymbol{\Lambda_g})$. Let $u^*_{qgi_g m_{qg}} = U_{qgi_g} - U_{qgm_{qg}} (i_g \neq m_{qg})$, where m_{qg} is the chosen alternative for the gth unordered-response variable by individual q, and stack the latent utility differentials into a vector $\mathbf{u}^*_{\boldsymbol{qg}} = \left[\left(u^*_{qg1m_{qg}}, u^*_{qg2m_{qg}}, ..., u^*_{qgI_g m_{qg}} \right)'; i_g \neq m_{qg} \right]$ [$(I_g - 1) \times 1$ vector]. Let $\boldsymbol{x_{qg}} = (\boldsymbol{x_{qg1}}, \boldsymbol{x_{qg2}}, \boldsymbol{x_{qg3}}, ..., \boldsymbol{x_{qgI_q}})'$ ($I_g \times L$ matrix), $\mathbf{V}_{qg} = \boldsymbol{x_{qg}} \boldsymbol{b_g}$ ($I_g \times 1$ vector) and $\widetilde{\boldsymbol{\Omega}}_{qg} = \boldsymbol{x_{qg}} \boldsymbol{\Omega}_g \boldsymbol{x'_{qg}}$ ($I_g \times I_g$ matrix). Define $\mathbf{M}_{qg}$ as an identity matrix of size $I_g - 1$, with an extra column of '-1' values added at the m^{th}_{qg} column. Also, construct the matrices $\boldsymbol{B}_{qg} = \mathbf{M}_{qg} \mathbf{V}_{qg}$, $\breve{\boldsymbol{\Omega}}_{qg} = \mathbf{M}_{qg} \widetilde{\boldsymbol{\Omega}}_{qg} \mathbf{M}'_{qg}$, and $\breve{\boldsymbol{\Lambda}}_{qg} = \mathbf{M}_{qg} \boldsymbol{\Lambda}_g \mathbf{M}'_{qg}$.

When there are G unordered-response variables, consider the stacked $\left[\sum_{g=1}^{G} (I_g - 1) \right] \times 1 -$ vector $\mathbf{u}^*_{\boldsymbol{q}} = \left[\left(\mathbf{u}^{*\prime}_{\boldsymbol{q1}}, \mathbf{u}^{*\prime}_{\boldsymbol{q2}}, ..., \mathbf{u}^{*\prime}_{\boldsymbol{qG}} \right)' \right]$, each of whose element vectors is formed by differencing utilities of alternatives from the chosen alternative m_{qg} for the g^{th} variable. Also, form a block diagonal covariance matrix $\ddot{\boldsymbol{\Omega}}_q$ of size $\left[\sum_{g=1}^{G} (I_g - 1) \right] \times \left[\sum_{g=1}^{G} (I_g - 1) \right]$, each block diagonal holding the matrix $\breve{\boldsymbol{\Omega}}_{qg}$, and the following matrix of the same size as $\ddot{\boldsymbol{\Omega}}_q$:

$$\ddot{\boldsymbol{\Lambda}}_{\mathbf{q}} = \begin{bmatrix} \breve{\boldsymbol{\Lambda}}_{\mathbf{q1}} & \breve{\boldsymbol{\Lambda}}_{\mathbf{q12}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{q1G}} \\ \breve{\boldsymbol{\Lambda}}_{\mathbf{q21}} & \breve{\boldsymbol{\Lambda}}_{\mathbf{q2}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{q2G}} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \breve{\boldsymbol{\Lambda}}_{\mathbf{qG1}} & \breve{\boldsymbol{\Lambda}}_{\mathbf{qG2}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{qG}} \end{bmatrix} \tag{2.19}$$

The off-diagonal elements in $\ddot{\boldsymbol{\Lambda}}_{\mathbf{q}}$ capture the dependencies across the utility differentials of different variables, the differential being taken with respect to the chosen alternative for each variable. It must be ensured that $\ddot{\boldsymbol{\Lambda}}_{\mathbf{q}}$ across individuals is derived from a common covariance matrix $\boldsymbol{\Lambda}$ for the original $\left(\sum_{g=1}^{G} I_g \right)$-error term vector $\boldsymbol{\xi_q} = (\boldsymbol{\xi'_{q1}}, \boldsymbol{\xi'_{q2}}, ..., \boldsymbol{\xi'_{qG}})'$. Appropriate identification considerations will have to be placed on the elements of $\boldsymbol{\Lambda}$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b'_1}, \boldsymbol{b'_2}, ... \boldsymbol{b'_G}, \overline{\boldsymbol{\Omega}}'_1, \overline{\boldsymbol{\Omega}}'_2, ..., \overline{\boldsymbol{\Omega}}'_G, \overline{\boldsymbol{\Lambda}}')'$. Using the notations as described earlier,

and defining $\boldsymbol{B}'_q = (\boldsymbol{B}'_{q1}, \boldsymbol{B}'_{q2}, ..., \boldsymbol{B}'_{qG})'$ and $\boldsymbol{\Xi}_q = \ddot{\boldsymbol{\Omega}}_q + \ddot{\boldsymbol{\Lambda}}_q$, the likelihood contribution of individual q is as below:

$$L_q(\boldsymbol{\theta}) = \Phi_{\tilde{I}}(\boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1}(\text{-}\boldsymbol{B}'_q), \boldsymbol{\Xi}_q^{\,*}), \tag{2.20}$$

where $\boldsymbol{\Xi}_q^{\,*} = \boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1} \boldsymbol{\Xi}_q \boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1}$ and $\tilde{I} = \sum_{g=1}^{G}(I_g - 1)$

The above likelihood function involves the evaluation of a $\sum_{g=1}^{G}(I_g - 1)$-dimensional integral for each individual, which can be very expensive if there are several variables and/or if each variable can take a large number of values. But, once again the Maximum Approximated Composite Marginal Likelihood (MACML) approach of Bhat (2011) can be used gainfully in this context, in which the MACML function only involves the computation of univariate and bivariate cumulative distributive functions. Specifically, consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the G nominal variables) of the joint pairwise probability of the chosen alternatives m_{qg} for the g^{th} variable and m_{ql} for the l^{th} variable for individual q.

$$L_{CML,q}^{CMMNP}(\boldsymbol{\theta}) = \prod_{g=1}^{G-1}\prod_{l=g+1}^{G} \Pr(d_{qg} = m_{qg}, d_{ql} = m_{ql}), \tag{2.21}$$

where d_{qg} is an index for the individual's choice for the g^{th} variable. One can also write:

$$\Pr(d_{qg} = m_{qg}, d_{ql} = m_{ql}) = \Phi_{\ddot{I}}(\boldsymbol{\omega}_{\tilde{\boldsymbol{\Xi}}_{qgl}}^{-1}(\text{-}\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}_{qgl}^{*}), \tag{2.22}$$

where $\ddot{I} = I_g + I_l - 2$ (I_g is the number of alternatives for the g^{th} variable), $\tilde{\boldsymbol{B}}_{qgl} = \boldsymbol{\Delta}_{qgl}\boldsymbol{B}_q$, $\tilde{\boldsymbol{\Xi}}_{qgl} = \boldsymbol{\Delta}_{qgl}\boldsymbol{\Xi}_{qgl}\boldsymbol{\Delta}'_{qgl}$, $\tilde{\boldsymbol{\Xi}}_{qgl}^{*} = \boldsymbol{\omega}_{\tilde{\boldsymbol{\Xi}}_{qgl}}^{-1}\tilde{\boldsymbol{\Xi}}_{qgl}\boldsymbol{\omega}_{\tilde{\boldsymbol{\Xi}}_{qgl}}^{-1}$, and $\boldsymbol{\Delta}_{qgl}$ is a $\ddot{I} * \tilde{I}$-selection matrix with an identity matrix of size $(I_g - 1)$ occupying the first $(I_g - 1)$ rows and the $\left[\sum_{j=1}^{g-1}(I_j - 1) + 1\right]^{th}$ through $\left[\sum_{j=1}^{g}(I_j - 1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j - 1) = 0$), and another identity matrix of size $(I_l - 1)$ occupying the last $(I_l - 1)$ rows and the $\left[\sum_{j=1}^{l-1}(I_j - 1) + 1\right]^{th}$ through $\left[\sum_{j=1}^{l}(I_j - 1)\right]^{th}$ columns. The net result is that the pairwise likelihood function now only needs the evaluation of a $\ddot{I}$-dimensional cumulative normal distribution function (rather than the $\tilde{I}$-dimensional cumulative distribution function in the maximum likelihood function). This can lead to substantial computation efficiency, and can be evaluated using the MVNCD

approximation of the MACML procedure. The MACML estimator $\hat{\boldsymbol{\theta}}_{MACML}$, obtained by maximizing the logarithm of the function $L^{CMMNP}_{MACML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q} L^{CMMNP}_{MACML,q}(\boldsymbol{\theta}),\ \textit{where}\ L^{CMMNP}_{MACML,q}(\boldsymbol{\theta}) = \prod_{g=1}^{G-1}\prod_{l=g+1}^{G} \Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qgl}})$ (with the MVNCD approximation), is asymptotically normal distributed with mean $\boldsymbol{\theta}$ and covariance matrix that can be estimated as:

$$\frac{\hat{\boldsymbol{G}}^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}, \tag{2.23}$$

$$\text{with } \hat{\boldsymbol{H}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q}\sum_{g=1}^{G-1}\sum_{l=g+1}^{G} \frac{\partial^2 \log\left[\Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{qgl})\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{MACML}}$$

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\sum_{g=1}^{G-1}\sum_{l=g+1}^{G} \frac{\partial \log\left[\Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{qgl})\right]}{\partial\boldsymbol{\theta}}\right)\left(\sum_{g=1}^{G-1}\sum_{l=g+1}^{G} \frac{\partial \log\left[\Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{qgl})\right]}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\boldsymbol{\theta}}_{MACML}} \tag{2.24}$$

An alternative estimator for $\hat{\boldsymbol{H}}$ is as below:

$$\hat{\boldsymbol{H}} = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{g=1}^{G-1}\sum_{l=g+1}^{G}\left(\left[\frac{\partial \log\left[\Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{qgl})\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log\left[\Phi_{\tilde{I}}(\boldsymbol{\omega}^{-1}_{\tilde{\boldsymbol{\Xi}}_q}(-\tilde{\boldsymbol{B}}_{qgl}), \tilde{\boldsymbol{\Xi}}^{*}_{qgl})\right]}{\partial\boldsymbol{\theta}'}\right]\right)_{\hat{\boldsymbol{\theta}}_{MACML}} \tag{2.25}$$

There are two important issues that need to be dealt with during estimation, each of which is discussed in turn below.

Identification

The estimated model needs to be theoretically identified. Suppose one considers utility differences with respect to the first alternative for each of the *G* variables. Then, the analyst can restrict the variance term of the top left diagonal of the covariance matrix (say $\breve{\boldsymbol{\Lambda}}^{*}_{g}$) of error differences $\left[\left(\xi_{qg2}-\xi_{qg1}\right),\left(\xi_{qg3}-\xi_{qg1}\right),\ldots\left(\xi_{qgI_g}-\xi_{qg1}\right)\right]'$ to 1 to account for scale invariance. However, note that the matrix $\breve{\boldsymbol{\Lambda}}^{*}_{g}$ is different from the matrix $\breve{\boldsymbol{\Lambda}}_{g}$, which corresponds to the covariance of utility differences taken with respect to the chosen alternative for the individual. Next, create a matrix of dimension $\left[\sum_{g=1}^{G}(I_g-1)\right]\times\left[\sum_{g=1}^{G}(I_g-1)\right]$ similar to that of $\breve{\boldsymbol{\Lambda}}_{g}$ in Equation (2.19), except that the matrix is expressed in terms of utility differences with respect to the first alternative for each nominal variable:

$$\ddot{\breve{\mathbf{\Lambda}}}^{*} = \begin{bmatrix} \breve{\mathbf{\Lambda}}_{\mathbf{1}}^{*} & \breve{\mathbf{\Lambda}}_{\mathbf{12}}^{*} & . & . & . & \breve{\mathbf{\Lambda}}_{\mathbf{1G}}^{*} \\ \breve{\mathbf{\Lambda}}_{\mathbf{21}}^{*} & \breve{\mathbf{\Lambda}}_{\mathbf{2}}^{*} & . & . & . & \breve{\mathbf{\Lambda}}_{\mathbf{2G}}^{*} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \breve{\mathbf{\Lambda}}_{\mathbf{G1}}^{*} & \breve{\mathbf{\Lambda}}_{\mathbf{G2}}^{*} & . & . & . & \breve{\mathbf{\Lambda}}_{\mathbf{G}}^{*} \end{bmatrix} \tag{2.26}$$

In the general case, this allows the estimation of $\sum_{g=1}^{G}\left(\frac{I_g * (I_g - 1)}{2} - 1\right)$ variance terms across all the G variables (originating from $\left(\frac{I_g * (I_g - 1)}{2} - 1\right)$ terms embedded in each $\breve{\mathbf{\Lambda}}_{\mathbf{g}}^{*}$ matrix; g=1,2,…G), and $\sum_{g=1}^{G-1}\sum_{l=g+1}^{G}(I_g - 1)\times(I_l - 1)$ covariance terms in the off-diagonal matrices of the $\ddot{\breve{\mathbf{\Lambda}}}^{*}$ matrix characterizing the dependence between the latent utility differentials (with respect to the first alternative) across the variables (originating from $(I_g - 1)\times(I_l - 1)$ estimable covariance terms within each off-diagonal matrix $\breve{\mathbf{\Lambda}}_{\mathbf{gl}}^{*}$ in $\ddot{\breve{\mathbf{\Lambda}}}^{*}$).

To construct the general covariance matrix $\mathbf{\Lambda}$ for the original $\left(\sum_{g=1}^{G} I_g\right)$-error term vector $\boldsymbol{\xi}_{\boldsymbol{q}}$, while also ensuring all parameters are identifiable, zero row and column vectors are inserted for the first alternatives of each unordered dependent variable in $\ddot{\breve{\mathbf{\Lambda}}}^{*}$. To do so, define a matrix $\mathbf{D}$ of size $\left[\left(\sum_{g=1}^{G} I_g\right)\right]\times\left[\left(\sum_{g=1}^{G}(I_g - 1)\right)\right]$. The first I_1 rows and $(I_1 - 1)$ columns correspond to the first variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a first row of zeros into this first I_1 rows and $(I_1 - 1)$ columns of $\mathbf{D}$. The rest of the columns for the first I_1 rows and the rest of the rows for the first $(I_1 - 1)$ columns take a value of zero. Next, rows $(I_1 + 1)$ through $(I_1 + I_2)$ and columns (I_1) through $(I_1 + I_2 - 2)$ correspond to the second variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a first row of zeros into this position. Continue this for all G nominal variables. Thus, for the case with two nominal variables, one nominal variable with 3 alternatives and the second with four alternatives, the matrix **D** takes the form shown below:

$$\left[\begin{array}{cc:ccc} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]_{7*5} \tag{2.27}$$

Then, the general covariance matrix may be developed as $\mathbf{\Lambda} = \mathbf{D}\ddot{\mathbf{\Lambda}}^{*}\mathbf{D}'$. All parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on utility differences and, at the same time, it provides a consistent means to obtain the covariance matrix $\ddot{\mathbf{\Lambda}}_{\mathbf{q}}$ that is needed for estimation (and is with respect to each individual's chosen alternative for each variable). Specifically, define a matrix $\widetilde{\mathbf{M}}_q$ of size $\left[\left(\sum_{g=1}^{G}(I_g - 1)\right)\right] \times \left[\left(\sum_{g=1}^{G} I_g\right)\right]$. The first $(I_1 - 1)$ rows and I_1 columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. The rest of the columns for the first $(I_1 - 1)$ rows and the rest of the rows for the first I_1 columns take a value of zero. Next, rows (I_1) through $(I_1 + I_2 - 2)$ and columns $(I_1 + 1)$ through $(I_1 + I_2)$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Continue this procedure for all *G* nominal variables. With the matrix $\widetilde{\mathbf{M}}_q$ as defined, the covariance matrix $\ddot{\mathbf{\Lambda}}_q$ for any individual is given by $\ddot{\mathbf{\Lambda}}_q = \widetilde{\mathbf{M}}_q \mathbf{\Lambda} \widetilde{\mathbf{M}}_q'$.

Positive Definiteness

The matrices $\ddot{\mathbf{\Lambda}}_q$ and $\ddot{\mathbf{\Omega}}_q$ have to be positive definite. The simplest way to guarantee the positive definiteness of $\ddot{\mathbf{\Lambda}}_q$ is to ensure that $\ddot{\mathbf{\Lambda}}^{*}$ is positive definite. To do so, the Cholesky matrix of $\ddot{\mathbf{\Lambda}}^{*}$ may be used as the matrix of parameters to be estimated. However, note that the top diagonal element of each $\ddot{\mathbf{\Lambda}}_g^{*}$ is normalized to one for identification, and this restriction should be recognized when using the Cholesky factor of $\ddot{\mathbf{\Lambda}}^{*}$. This can be achieved by appropriately parameterizing the diagonal elements of the Cholesky decomposition matrix. Thus, consider the lower triangular Cholesky matrix $\ddot{\boldsymbol{L}}^{*}$ of the same size as $\ddot{\mathbf{\Lambda}}^{*}$. Whenever a diagonal element (say

the kk^{th} element) of $\ddot{\boldsymbol{\Lambda}}^*$ is to be normalized to one, the first element in the corresponding row of $\ddot{\boldsymbol{L}}^*$ is written as $\sqrt{1-\sum_{j=2}^{k} l_{kj}^2}$, where the l_{kj} elements are the Cholesky factors that are to be estimated. With this parameterization, $\ddot{\boldsymbol{\Lambda}}^*$ obtained as $\ddot{\boldsymbol{L}}^*\ddot{\boldsymbol{L}}^{*'}$ is positive definite and adheres to the scaling conditions. Using this, one constructs $\boldsymbol{\Lambda}$, and subsequently obtains $\ddot{\boldsymbol{\Lambda}}_q$ as discussed earlier. The resulting $\ddot{\boldsymbol{\Lambda}}_q$ is positive definite. The positive definiteness of $\ddot{\boldsymbol{\Omega}}_q$ is ensured by writing $\boldsymbol{\Omega}_{\mathbf{g}} = \boldsymbol{L}_{\boldsymbol{g}}\boldsymbol{L}_{\boldsymbol{g}}'$.

References for the CML Estimation of the CMMNP Model

Bhat, C.R., Paleti, R., Pendyala, R.M., Lorenzini, K., Konduri, K.C., 2013. Accommodating immigration status and self selection effects in a joint model of household auto ownership and residential location choice. *Transportation Research Record* 2382, 142-150.

Feddag, M.-L., 2013. Composite likelihood estimation for multivariate probit latent traits models. *Communications in Statistics - Theory and Methods* 42(14), 2551-2566.

Kortum, K., Paleti, R., Bhat, C.R., Pendyala, R.M., 2012. Joint model of residential relocation choice and underlying causal factors, *Transportation Research Record*, 2303, 28-37.

2.2.2.3. The Panel MNP (PMNP) Model

Consider the following model with '*t*' now being an index for choice occasion:

$$U_{qit} = \boldsymbol{\beta}_{\boldsymbol{q}}'\boldsymbol{x}_{\boldsymbol{qti}} + \xi_{qti},\ \boldsymbol{\beta}_{\boldsymbol{q}} \sim MVN(\boldsymbol{b},\boldsymbol{\Omega}),\ q = 1,2,...,Q,\ \ t = 1,2,...,T,\ i = 1,2,...,I. \tag{2.28}$$

For ease, we assume that all alternatives are available at each choice instance of each individual, and that we have a balanced panel (that is, we have the same number of choice instances from each individual). The first assumption is innocuous and helps in presentation. The relaxation of the second assumption only requires a different weight per individual, exactly as discussed earlier for the ordered-response case. $\boldsymbol{x}_{\boldsymbol{qti}}$ is a $(L\times 1)$-column vector of exogenous attributes whose first (*I-1*) elements correspond to alternative specific constants for (I-1) alternatives (with one of the alternatives being the base alternative) and the remaining variables being the non-constant variables. $\boldsymbol{\beta}_{\boldsymbol{q}}$ is an individual-specific $(L\times 1)$-column vector of corresponding coefficients that varies across individuals based on unobserved individual attributes. Assume that the $\boldsymbol{\beta}_{\boldsymbol{q}}$ vector is a realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega} = \boldsymbol{L}\boldsymbol{L}'$, where $\boldsymbol{L}$ is the lower-triangular Cholesky factor of $\boldsymbol{\Omega}$. Thus, as in the case of the panel ordered-response model, the coefficients $\boldsymbol{\beta}_{\boldsymbol{q}}$ are considered constant over choice situations of a given decision maker. We also assume that ξ_{qti} is independent and identically normally distributed across *individuals and choice occasions*, but allow a general

covariance structure across alternatives for each choice instance of each individual. Specifically, let $\boldsymbol{\xi_{qt}} = (\xi_{qt1}, \xi_{qt2}, ...\xi_{qtI})'$ ($I \times 1$ vector). Then, we assume $\boldsymbol{\xi_{qt}} \sim MVN_I(0, \boldsymbol{\Lambda})$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Lambda}$ for identifiability. To do so, we follow the same exact procedure as in the CMNP model. Specifically, only utility differentials matter at each choice occasion. Taking the utility differentials with respect to the first alternative, only the elements of the covariance matrix $\boldsymbol{\Lambda_1}$ of $\widetilde{\xi}_{qti1} = \xi_{qti} - \xi_{qt1}$ $(i \neq 1)$ are estimable, and $\boldsymbol{\Lambda}$ is constructed from $\boldsymbol{\Lambda_1}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. We also normalize the element of $\boldsymbol{\Lambda}$ in the second row and second column to the value of one.

Define the following vectors and matrices: $\mathbf{U}_{qt} = (U_{qt1}, U_{qt2}, ..., U_{qtI})'$ ($I \times 1$ vector), $\mathbf{U}_q = (\mathbf{U}_{q1}, \mathbf{U}_{q2}, ..., \mathbf{U}_{qI})'$ ($TI \times 1$ vector), $\boldsymbol{\xi_q} = (\boldsymbol{\xi'_{q1}}, \boldsymbol{\xi'_{q2}}, ...\boldsymbol{\xi'_{qT}})'$ ($TI \times 1$ vector), $\boldsymbol{x_{qt}} = (\boldsymbol{x_{qt1}}, \boldsymbol{x_{qt2}}, \boldsymbol{x_{qt3}}, ..., \boldsymbol{x_{qtI}})'$ ($I \times L$ matrix), $\boldsymbol{x}_q = (\boldsymbol{x}'_{q1}, \boldsymbol{x}'_{q2}, ..., \boldsymbol{x}'_{qT})'$ ($TI \times L$ matrix), $\mathbf{V}_q = \boldsymbol{x}_q \boldsymbol{b}$ ($TI \times 1$ vector), $\widetilde{\boldsymbol{\Omega}}_q = \boldsymbol{x}_q \boldsymbol{\Omega} \boldsymbol{x}'_q$ ($TI \times TI$ matrix), and $\widetilde{\boldsymbol{\Xi}}_{\boldsymbol{q}} = \widetilde{\boldsymbol{\Omega}}_q + (\mathbf{IDEN}_T \otimes \boldsymbol{\Lambda})$ ($TI \times TI$ matrix). Then, we may write, in matrix notation, $\mathbf{U}_{\boldsymbol{q}} = \mathbf{V}_{\boldsymbol{q}} + \boldsymbol{\xi}_{\boldsymbol{q}}$ and $\mathbf{U}_{\boldsymbol{q}} \sim MVN_{TI}(\mathbf{V}_{\boldsymbol{q}}, \widetilde{\boldsymbol{\Xi}}_{\boldsymbol{q}})$. Let the individual q choose alternative m_{qt} at the tth choice occasion. To develop the likelihood function, define $\mathbf{M}_q$ as an $[T \times (I-1)] \times [TI]$ block-diagonal matrix, each block diagonal being of size $(I\text{ - }1) \times (I)$) and containing the matrix $\mathbf{M}_{qt}$. $\mathbf{M}_{qt}$ itself is an identity matrix of size (I-1) with an extra column of '-1' values added at the m_{qt}^{th} column. Let $\boldsymbol{B}_q = \mathbf{M}_q \mathbf{V}_q$ and $\boldsymbol{\Xi}_{\boldsymbol{q}} = \mathbf{M}_q \widetilde{\boldsymbol{\Xi}}_{\boldsymbol{q}} \mathbf{M}'_q$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Lambda}}')'$. The likelihood contribution of individual q is as below:

$$L_q(\boldsymbol{\theta}) = \Phi_{\widetilde{J}}(\boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1}(-\boldsymbol{B}_{\boldsymbol{q}}), \boldsymbol{\Xi}_{\boldsymbol{q}}^{\ *}), \tag{2.29}$$

where $\widetilde{J} = T \times (I-1)$, and $\boldsymbol{\Xi}_{\boldsymbol{q}}^{\ *} = \boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1} \boldsymbol{\Xi}_{\boldsymbol{q}} \boldsymbol{\omega}_{\boldsymbol{\Xi}_q}^{-1}$.

The simulation approaches for evaluating the panel likelihood function involve integration of dimension $[T \times (I-1)]$. Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the T choice occasions) of the joint pairwise probability of the chosen alternatives m_{qt} for the t^{th} choice occasion and m_{qg} for the g^{th} choice occasion for individual q.

$$L_{CML,q}^{PMNP}(\boldsymbol{\theta}) = \prod_{t=1}^{T-1} \prod_{g=t+1}^{T} \Pr(d_{qt} = m_{qt}, d_{qg} = m_{qg}), \tag{2.30}$$

where d_{qt} is an index for the individual's choice on the tth choice occasion. One can also write:

$$\Pr(d_{qt}=m_{qt}, d_{qg}=m_{qg}) = \Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qgl}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}}), \tag{2.31}$$

where $\ddot{J}=2(I-1)$, $\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}=\boldsymbol{\Delta}_{qtg}\boldsymbol{B}_{\boldsymbol{q}}$, $\tilde{\boldsymbol{\Xi}}_{qtg}=\boldsymbol{\Delta}_{qtg}\boldsymbol{\Xi}_{qtg}\boldsymbol{\Delta}'_{qtg}$, $\tilde{\boldsymbol{\Xi}}^{*}_{qtg}=\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}\tilde{\boldsymbol{\Xi}}_{qtg}\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}$, and $\boldsymbol{\Delta}_{qtg}$ is a $\ddot{J}*\tilde{J}$-selection matrix with an identity matrix of size $(I-1)$ occupying the first $(I-1)$ rows and the $[(t-1)\times(I-1)+1]^{th}$ through $[t\times(I-1)]^{th}$ columns, and another identity matrix of size $(I-1)$ occupying the last $(I-1)$ rows and the $[(g-1)\times(I-1)+1]^{th}$ through $[g\times(I-1)]^{th}$ columns. The pairwise likelihood function now only needs the evaluation of a $\ddot{J}$-dimensional cumulative normal distribution function (rather than the $\tilde{I}$-dimensional cumulative distribution function in the maximum likelihood function). The MACML estimator $\hat{\boldsymbol{\theta}}_{MACML}$ is obtained by maximizing the logarithm of the function $L^{PMNP}_{MACML}(\boldsymbol{\theta})=\prod_{q=1}^{Q}L^{PMNP}_{MACML,q}(\boldsymbol{\theta})$, where $L^{PMNP}_{MACML,q}(\boldsymbol{\theta})=\prod_{t=1}^{T-1}\prod_{g=t+1}^{T}\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})$ (with the MVNCD approximation). The covariance matrix is estimated as:

$$\frac{\hat{\boldsymbol{G}}^{-1}}{Q}=\frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q},$$

$$\text{with } \hat{\boldsymbol{H}}=-\frac{1}{Q}\left[\sum_{q=1}^{Q}\sum_{t=1}^{T-1}\sum_{g=t+1}^{T}\frac{\partial^2\log\left[\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{MACML}}$$

$$\hat{\boldsymbol{J}}=\frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\sum_{t=1}^{T-1}\sum_{g=t+1}^{T}\frac{\partial\log\left[\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})\right]}{\partial\boldsymbol{\theta}}\right)\left(\sum_{t=1}^{T-1}\sum_{g=t+1}^{T}\frac{\partial\log\left[\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})\right]}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\boldsymbol{\theta}}_{MACML}} \tag{2.32}$$

An alternative estimator for $\hat{\boldsymbol{H}}$ is as below:

$$\hat{\boldsymbol{H}}=\frac{1}{Q}\sum_{q=1}^{Q}\sum_{t=1}^{T-1}\sum_{g=1}^{T}\left(\left[\frac{\partial\log\left[\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial\log\left[\Phi_{\ddot{J}}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qtg}}(\text{-}\tilde{\boldsymbol{B}}_{\boldsymbol{qtg}}), \tilde{\boldsymbol{\Xi}}^{*}_{\boldsymbol{qtg}})\right]}{\partial\boldsymbol{\theta}'}\right]\right)_{\hat{\boldsymbol{\theta}}_{MACML}} \tag{2.33}$$

References for the CML Estimation of the PMNP Model

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

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

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

2.3. Spatial Formulations

In the past decade, there has been increasing interest and attention on recognizing and explicitly accommodating spatial (and social) dependence among decision-makers (or other observation units) in urban and regional modeling, agricultural and natural resource economics, public economics, geography, marketing, sociology, political science, and epidemiology. The reader is referred to a special issue of *Regional Science and Urban Economics* entitled "Advances in spatial econometrics" (edited by Arbia and Kelejian, 2010) and another special issue of the *Journal of Regional Science* entitled "Introduction: Whither spatial econometrics?" (edited by Patridge *et al*., 2012) for a collection of recent papers on spatial dependence, and to Elhorst (2010), Anselin (2010), Ferdous and Bhat (2013), and Bhat *et al*. (2014a) for overviews of recent developments in the spatial econometrics field. Within the past few years, there has particularly been an explosion in studies that recognize and accommodate spatial dependency in discrete choice models. The typical way this is achieved is by applying spatial structures developed in the context of continuous dependent variables to the linear (latent) propensity variables underlying discrete choice dependent variables (see reviews of this literature in Fleming, 2004, Franzese and Hays, 2008, LeSage and Pace, 2009, Hays *et al*. 2010, Brady and Irwin, 2011, and Sidharthan and Bhat, 2012). The two dominant techniques, both based on simulation methods, for the estimation of such spatial discrete models are the frequentist recursive importance sampling (RIS) estimator (which is a generalization of the more familiar Geweke-Hajivassiliou-Keane or GHK simulator; see Beron and Vijverberg, 2004) and the Bayesian Markov Chain Monte Carlo (MCMC)-based estimator (see LeSage and Pace, 2009). However, both of these methods are confronted with multi-dimensional normal integration of the order of the number of observational units in ordered-response models, and are cumbersome to implement in typical empirical contexts with even moderate estimation sample sizes (see Bhat, 2011 and Franzese *et al*., 2010). The RIS and MCMC methods become even more difficult (to almost infeasible) to implement in a spatial unordered multinomial choice context because the likelihood function entails a multidimensional integral of the order of the number of observational units factored up by the number of alternatives minus one (in the case of multi-period data, the integral dimension gets factored up further by the number of time periods of observation). Recently, Bhat and colleagues have suggested a composite marginal likelihood (CML) inference approach for estimating spatial binary/ordered-response probit models, and the maximum approximate composite marginal likelihood (MACML) inference approach for estimating spatial unordered-response multinomial probit (MNP) models. These methods are easy to implement, require no simulation, and involve only univariate and bivariate cumulative normal distribution function evaluations, regardless of the number of alternatives, or the number of choice occasions per observation unit, or the number of observation units, or the nature of social/spatial dependence structures.

In the spatial analysis literature, the two workhorse specifications to capture spatial dependencies are the spatial lag and the spatial error specifications (Anselin, 1988). The spatial

lag specification, in reduced form, allows spatial dependence through both spatial spillover effects (observed exogenous variables at one location having an influence on the dependent variable at that location and neighboring locations) as well as spatial error correlation effects (unobserved exogenous variables at one location having an influence on the dependent variable at that location and neighboring locations). The spatial error specification, on the other hand, assumes that spatial dependence is only due to spatial error correlation effects and not due to spatial spillover effects. The spatial error specification is somewhat simpler in formulation and estimation than the spatial lag model. But, as emphasized by McMillen (2010), the use of a parametric spatial error structure is "troublesome because it requires the researcher to specify the actual structure of the errors", while it is much easier to justify a parametric spatial lag structure when accommodating spatial dependence. Beck *et al.* (2006) also find theoretical and conceptual issues with the spatial error model and refer to it as being "odd", because the formulation rests on the "hard to defend" position that "space matters in the error process but not in the substantive portion of the model". As they point out, the implication is that if a new independent variable is added to a spatial error model "so that we move it from the error to the substantive portion of the model", the variable magically ceases to have a spatial impact on neighboring observations. Of course, the spatial lag and spatial error specifications can be combined together in a Kelejian-Prucha specification (see Elhorst, 2010), or the spatial lag could be combined with spatially lagged exogenous variable effects in a Spatial Durbin specification (see Bhat *et al*., 2014a). In all of these cases, the spatial dependence leads also to spatial heteroscedasticity in the random error terms.

In this paper, we will assume the spatial lag structure as the specification of spatial dependency. However, it is very straightforward to extend our approach to other dependency specifications. Indeed, there is no conceptual difficulty in doing so, nor is there much impact on coding or computational burden. The focus on the spatial lag structure is simply for uniformity and notational ease. In addition to the spatial lag-based and resulting heteroscedasticity effect, it is also likely that there is heterogeneity (*i.e.*, differences in relationships between the dependent variable of interest and the independent variables across decision-makers or spatial units (see, Fotheringham and Brunsdon, 1999, Bhat and Zhao, 2002, Bhat and Guo, 2004). When combined with the spatial lag effect, the unobserved heterogeneity effects get correlated over decision agents based on the spatial (or social) proximity of the agents' locations, which is then referred to as spatial drift (see Bradlow *et al*., 2005 for a discussion). But such spatial drift effects have been largely ignored thus far in the literature (but see Bhat *et al*., 2014a). We explicitly incorporate such drift effects in the models discussed below. All notations from previous sections carry over to the sections below.

2.3.1 Spatial Ordered Response Models

2.3.1.1 The Spatial CUOP Model

The spatial CUOP (SCUOP) is an extension of the aspatial CUOP model from Section 2.2.1.1, and may be written as follows:

$$y_q^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'}^* + \boldsymbol{\beta}_q' \boldsymbol{x}_q + \varepsilon_q, \; y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{2.34}$$

where the $w_{qq'}$ terms are the elements of an exogenously defined distance-based spatial (or social) weight matrix $\mathbf{W}$ corresponding to individuals q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$), and δ $(0 < \delta < 1)$ is the spatial autoregressive parameter. The weights $w_{qq'}$ can take the form of a discrete function such as a contiguity specification ($w_{qq'}$=1 if the individuals q and q' are adjacent and 0 otherwise) or a specification based on a distance threshold ($w_{qq'} = c_{qq'} / \sum_{q'} c_{qq'}$, where $c_{qq'}$ is a dummy variable taking the value 1 if the individual q' is within the distance threshold and 0 otherwise). It can also take a continuous form such as those based on the inverse of distance $d_{qq'}$ and its power functions $\left(w_{qq'} = (1/d_{qq'}^n) \left[\sum_{q'} 1/d_{qq'}^n \right]^{-1} \right)$ $(n > 0)$, the inverse of exponential distance, and the shared edge length $\widetilde{d}_{qq'}$ between individuals (or observation units) $w_{qq'} = \widetilde{c}_{qq'} \widetilde{d}_{qq'} / \left(\sum_{q'} \widetilde{c}_{qq'} \widetilde{d}_{qq'}, \right)$ (where $\widetilde{c}_{qq'}$ is a dummy variable taking the value 1 if q and q' are adjoining based on some pre-specified spatial criteria, and 0 otherwise). All of these functional forms for the weight matrix may be tested empirically.

The latent propensity representation of Equation (2.34) can be written equivalently in vector notation as:

$$\boldsymbol{y}^* = \delta \mathbf{W} \boldsymbol{y}^* + \mathbf{x}\boldsymbol{b} + \widetilde{\mathbf{x}} \widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{2.35}$$

where $\boldsymbol{y}^* = (y_1^*, y_2^*, ..., y_Q^*)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_Q)'$ are $(Q \times 1)$ vectors, $\mathbf{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_Q)'$ is a $(Q \times L)$ matrix of exogenous variables for all Q individuals, $\widetilde{\boldsymbol{x}}$ is a $(Q \times QL)$ block-diagonal matrix with each block-diagonal of size $(1 \times L)$ being occupied by the vector $\boldsymbol{x}_q'$ $(q = 1,2,..., Q)$, and $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}_1', \widetilde{\boldsymbol{\beta}}_2', ..., \widetilde{\boldsymbol{\beta}}_Q')'$ is a $(QL \times 1)$ vector. Through simple matrix algebra manipulation, Equation (2.35) may be re-written as:

$$\boldsymbol{y}^* = \mathbf{S}\left(\mathbf{x}\boldsymbol{b} + \widetilde{\mathbf{x}} \widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon} \right), \tag{2.36}$$

where $\mathbf{S} = \left[\mathbf{IDEN}_Q - \delta \mathbf{W} \right]^{-1}$ is a $(Q \times Q)$ matrix. The vector $\boldsymbol{y}^*$ is multivariate normally distributed as $\boldsymbol{y}^* \sim MVN_Q(\boldsymbol{B}, \boldsymbol{\Xi})$, where

$$\boldsymbol{B} = \mathbf{S}\mathbf{x}\boldsymbol{b} \text{ and } \boldsymbol{\Xi} = \mathbf{S}\left[\widetilde{\mathbf{x}} \left(\mathbf{IDEN}_Q \otimes \boldsymbol{\Omega} \right) \widetilde{\mathbf{x}}' + \mathbf{IDEN}_Q \right] \mathbf{S}'. \tag{2.37}$$

The likelihood function $L(\boldsymbol{\theta})$ for the SCUOP model takes the following form:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y} = \boldsymbol{m}) = \int_{D_{y^*}} f_Q(\boldsymbol{y}^* \mid \boldsymbol{B}, \boldsymbol{\Xi}) d\boldsymbol{y}^*, \tag{2.38}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_Q)'$, $\boldsymbol{m} = (m_1, m_2, ..., m_Q)'$ is the corresponding ($Q \times 1$) vector of the actual observed ordinal levels, $D_{\boldsymbol{y}^*}$ is the integration domain defined as $D_{\mathrm{y}^*} = \{\boldsymbol{y}^* : \psi_{q,m_q-1} < y_q^* < \psi_{q,m_q}, \forall\, q = 1, 2, ..., Q\}$, and $f_Q(.)$ is the multivariate normal density function of dimension Q.

The rectangular integral in the likelihood function is of dimension Q, which can become problematic from a computational standpoint. Further, the use of traditional numerical simulation techniques can lead to convergence problems during estimation even for moderately sized Q (Bhat *et al.*, 2010a; Müller and Czado, 2005). The alternative is to use the composite marginal likelihood (CML) approach. Using a pairwise CML method, the function to be maximized is:

$$\begin{aligned} &L_{CML}^{\text{SCUOP}}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} L_{qq'}, \text{ where } L_{qq'} = P\left([\boldsymbol{y}]_q = [\boldsymbol{m}]_q, [\boldsymbol{y}]_{q'} = [\boldsymbol{m}]_{q'}\right). \text{ That is,} \\ &L_{qq'} = \left[\Phi_2(\varphi_q, \varphi_{q'}, v_{qq'}) - \Phi_2(\varphi_q, \mu_{q'}, v_{qq'}) - \Phi_2(\mu_q, \varphi_{q'}, v_{qq'}) + \Phi_2(\mu_q, \mu_{q'}, v_{qq'})\right] \end{aligned} \tag{2.39}$$

where $\varphi_q = \dfrac{\psi_{q,m_q} - [\boldsymbol{B}]_q}{\sqrt{[\boldsymbol{\Sigma}]_{qq}}}$, $\mu_q = \dfrac{\psi_{q,m_q-1} - [\boldsymbol{B}]_q}{\sqrt{[\boldsymbol{\Sigma}]_{qq}}}$, $v_{qq'} = \dfrac{[\boldsymbol{\Sigma}]_{qq'}}{\sqrt{[\boldsymbol{\Sigma}]_{qq}}\sqrt{[\boldsymbol{\Sigma}]_{q'q'}}}$.

In the above expression, $[\boldsymbol{B}]_q$ represents the q^{th} element of the column vector $\boldsymbol{B}$, while $[\boldsymbol{\Sigma}]_{qq'}$ represents the qq'^{th} element of the matrix $\boldsymbol{\Sigma}$.

The pairwise marginal likelihood function of Equation (2.39) comprises $Q(Q-1)/2$ pairs of bivariate probability computations, which can itself become quite time consuming. However, previous studies (Varin and Vidoni, 2009, Bhat *et al.*, 2010a, Varin and Czado, 2010) have shown that spatial dependency drops quickly with inter-observation distance. Therefore, there is no need to retain all observation pairs because the pairs formed from the closest observations provide much more information than pairs far from one another. The "optimal" distance for including pairings can be based on minimizing the trace of the asymptotic covariance matrix. Thus, the analyst can start with a low value of the distance threshold (leading to a low number of pairwise terms in the CML function) and then continually increase the distance threshold up to a point where the gains from increasing the distance threshold is very small or even drops. To be specific, for a given threshold, construct a $Q \times Q$ matrix $\widetilde{\mathbf{R}}$ with its q^{th} column filled with a $Q \times 1$ vector of zeros and ones as follows: if the observational unit q' is not within the specified threshold distance of unit q, the q'^{th} row has a value of zero; otherwise, the q'^{th} row has a value of one. By construction, the q^{th} row of the q^{th} column has a value of one. Let $\left[\widetilde{\mathbf{R}}\right]_{qq'}$ be the qq^{th} element of the matrix $\widetilde{\mathbf{R}}$, and let $\widetilde{W} = \sum_{q=1}^{Q-1} \sum_{q'=q+1}^{Q} \left[\widetilde{\mathbf{R}}\right]_{qq'}$. Define a set $\widetilde{\mathbf{C}}_{\boldsymbol{q}}$ of

all individuals (observation units) that have a value of '1' in the vector $\left[\widetilde{\mathbf{R}}\right]_q$, where $\left[\widetilde{\mathbf{R}}\right]_q$ is the *qth* column of the vector $\widetilde{\mathbf{R}}$. Then, the CML function is as follows:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1} \prod_{\substack{q'=q+1 \\ q' \in \widetilde{C}_q}}^{Q} L_{qq'} \,. \tag{2.40}$$

The covariance matrix of the CML estimator is $\frac{\left[\hat{\boldsymbol{G}}\right]^{-1}}{\widetilde{W}} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{\widetilde{W}}$, where

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}} \left[\sum_{q=1}^{Q-1} \sum_{\substack{q'=q+1 \\ q'' \in \widetilde{C}_q}}^{Q} \frac{\partial^2 \log L_{qq'}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\hat{\theta}_{CML}} \text{, or alternatively,} \tag{2.41}$$

$$\hat{\boldsymbol{H}} = \frac{1}{\widetilde{W}} \sum_{q=1}^{Q-1} \sum_{\substack{q'=q+1 \\ q' \in \widetilde{C}_q}}^{Q} \left(\left[\frac{\partial \log L_{qq'}}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial \log L_{qq'}}{\partial \boldsymbol{\theta}'} \right] \right)_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \tag{2.42}$$

However, the estimation of the "vegetable" matrix $\boldsymbol{J}$ is more difficult in this case. One cannot empirically estimate $\boldsymbol{J}$ as the sampling variance of the individual contributions to the composite score function (as was possible when there were Q independent contributions) because if the underlying spatial dependence in observation units. But a windows resampling procedure (see Heagerty and Lumley, 2000) may be used to estimate $\boldsymbol{J}$. This procedure entails the construction of suitable overlapping subgroups of the sample that may be viewed as independent replicated observations. Then, $\boldsymbol{J}$ may be estimated empirically. While there are several ways to implement this, Bhat (2011) suggests overlaying the spatial region under consideration with a square grid providing a total of $\widetilde{Q}$ internal and external nodes. Then, select the observational unit closest to each of the $\widetilde{Q}$ grid nodes to obtain $\widetilde{Q}$ observational units from the original Q observational units ($\tilde{q} = 1, 2, 3, \ldots, \widetilde{Q}$). Let $\widetilde{\boldsymbol{R}}_{\tilde{q}}$ be the $Q \times 1$ matrix representing the $\tilde{q}^{th}$ column vector of the matrix $\widetilde{\boldsymbol{R}}$, let $\widetilde{\mathbf{C}}_{\tilde{q}}$ be the set of all individuals (observation units) that have a value of '1' in the vector $\widetilde{\boldsymbol{R}}_{\tilde{q}}$, and let $\boldsymbol{y}_{\tilde{q}}$ be the sub-vector of $\boldsymbol{y}$ with values of '1' in the rows of $\widetilde{\boldsymbol{R}}_{\tilde{q}}$. Let $N_{\tilde{q}}$ be the sum (across rows) of the vector $\widetilde{\boldsymbol{R}}_{\tilde{q}}$ (that is, $N_{\tilde{q}}$ is the cardinality of $\widetilde{\mathbf{C}}_{\tilde{q}}$), so that the dimension of $\boldsymbol{y}_{\tilde{q}}$ is $N_{\tilde{q}} \times 1$. Let $l_{\tilde{q}}$ be the index of all elements in the vector $\boldsymbol{y}_{\tilde{q}}$, so that $l_{\tilde{q}}$ =1,2,… $N_{\tilde{q}}$. Next, define $\ddot{C}_{\tilde{q}} = \left[N_{\tilde{q}}(N_{\tilde{q}} - 1)\right]/2$. Then, the **J** matrix maybe empirically estimated as:

$$\hat{\mathbf{J}} = \frac{1}{\widetilde{Q}} \left[\sum_{\tilde{q}=1}^{\widetilde{Q}} \left[\frac{1}{\ddot{C}_{\tilde{q}}} \left(\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}-1} \sum_{l'=l_{\tilde{q}}+1}^{N_{\tilde{q}}} \frac{\partial \log L_{l_{\tilde{q}} l'}}{\partial \boldsymbol{\theta}} \right] \left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}-1} \sum_{l'=l_{\tilde{q}}+1}^{N_{\tilde{q}}} \frac{\partial \log L_{l_{\tilde{q}} l'}}{\partial \boldsymbol{\theta}} \right]' \right) \right] \right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} . \tag{2.43}$$

To ensure the constraints on the autoregressive term δ, the analyst can parameterize $\delta = 1/[1+\exp(\widetilde{\delta})]$. Once estimated, the $\widetilde{\delta}$ estimate can be translated back to estimates of an estimate of δ.

References for the CML Estimation of the Spatial CUOP (or SCUOP) Model

Ferdous, N., Pendyala, R.M., Bhat, C.R., Konduri, K.C., 2011. Modeling the influence of family, social context, and spatial proximity on use of nonmotorized transport mode, *Transportation Research Record* 2230, 111-120.

Spissu, E., Eluru, N., Sener, I.N., Bhat, C.R., Meloni, I., 2010. Cross-clustered model of frequency of home-based work participation in traditionally off-work hours. *Transportation Research Record* 2157, 138-146.

Whalen, K.E., Paez, A., Bhat, C., Moniruzzaman, M., Paleti, R., 2012. *T*-communities and sense of community in a university town: evidence from a student sample using a spatial ordered-response model. *Urban Studies* 49(6), 1357-1376.

2.3.1.2 The Spatial CMOP Model

We start with Equation (2.6) of the aspatial CMOP model in Section 2.2.1.2, and now add a spatial lag formulation:

$$y^*_{qi} = \delta_i \sum_{q'=1}^{Q} w_{qq'} y^*_{q'i} + \boldsymbol{\beta}'_{\boldsymbol{qi}} \boldsymbol{x_q} + \varepsilon_{qi}, y_{qi} = k_i \text{ if } \psi^i_{q,k_i-1} < y^*_{qi} < \psi^i_{q,k_i}. \tag{2.44}$$

Define $\boldsymbol{y^*_q} = (y^*_{q1}, y^*_{q2}, ..., y^*_{qI})'$ ($I \times 1$ vector), , $\boldsymbol{y^*} = [(\boldsymbol{y^*_1})', (\boldsymbol{y^*_2})', (\boldsymbol{y^*_3})', ..., (\boldsymbol{y^*_Q})']'$ ($QI \times 1$ vector), $\boldsymbol{y_q} = (y_{q1}, y_{q2}, ..., y_{qI})'$ ($I \times 1$ vector), $\boldsymbol{y} = [(\boldsymbol{y_1})', (\boldsymbol{y_2})', (\boldsymbol{y_3})', ..., (\boldsymbol{y_Q})']'$ ($QI \times 1$ vector), $\boldsymbol{m_q} = (m_{q1}, m_{q2}, ..., m_{qI})'$ ($I \times 1$ vector), $\boldsymbol{m} = (m_1, m_2, ..., m_Q)'$ ($QI \times 1$ vector), $\widetilde{\mathbf{x}}_{\boldsymbol{q}} = \mathbf{IDEN}_I \otimes \boldsymbol{x}'_{\boldsymbol{q}}$ ($I \times IL$ matrix; $\mathbf{IDEN}_I$ is an identity matrix of size I), $\widetilde{\mathbf{x}} = (\widetilde{\mathbf{x}}'_1, \widetilde{\mathbf{x}}'_2, \widetilde{\mathbf{x}}'_3, ... \widetilde{\mathbf{x}}'_{\boldsymbol{Q}})'$ ($QI \times IL$ matrix), $\boldsymbol{\beta}_{qi} = \boldsymbol{b_i} + \widetilde{\boldsymbol{\beta}}_{qi}$, $\widetilde{\boldsymbol{\beta}}_q = (\widetilde{\boldsymbol{\beta}}'_{q1}, \widetilde{\boldsymbol{\beta}}'_{q2} ..., \widetilde{\boldsymbol{\beta}}'_{qI})'$ ($IL \times 1$ vector), $\widetilde{\boldsymbol{\beta}}_q \sim MVN_{I \times L}(0, \boldsymbol{\Omega})$ (the $\widetilde{\boldsymbol{\beta}}_q$ random coefficients are independent across individuals), $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}'_1, \widetilde{\boldsymbol{\beta}}'_2 ..., \widetilde{\boldsymbol{\beta}}'_Q)'$ (Q$IL \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_{\boldsymbol{I}})'$ ($IL \times 1$ vector), $\boldsymbol{\varepsilon_q} = (\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \ldots, \varepsilon_{qI})'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \boldsymbol{\varepsilon}'_3, \ldots, \boldsymbol{\varepsilon}'_{\boldsymbol{Q}})'$ ($QI \times 1$ vector), $\boldsymbol{\delta} = (\delta_1, \delta_3, \delta_3, ..., \delta_I)'$ ($I \times 1$ vector), and $\widetilde{\boldsymbol{\delta}} = \mathbf{1}_Q \otimes \boldsymbol{\delta}$ ($QI \times 1$ vector; $\mathbf{1}_Q$ is a vector of size Q with all elements equal to 1). Also, define the following matrix:

$$\widetilde{\widetilde{\mathbf{x}}} = \begin{bmatrix} \widetilde{\mathbf{x}}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{x}}_2 & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widetilde{\mathbf{x}}_3 & \mathbf{0} \cdots \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \ldots \widetilde{\mathbf{x}}_{\mathbf{Q}} \end{bmatrix} (QI \times QIL \text{ matrix}). \tag{2.45}$$

Collect all the weights $w_{qq'}$ into a row-normalized spatial weight matrix $\mathbf{W}$. All other notations from Section 2.2.1.2 are carried over to this section, including the multivariate standard normal distribution specification for $\boldsymbol{\varepsilon}_q$ with mean zero and correlation matrix $\boldsymbol{\Sigma}$ (see Equation 2.7). With these definitions, Equation (2.44) may be re-written in matrix form as:

$$\boldsymbol{y}^* = [\widetilde{\boldsymbol{\delta}}.*(\mathbf{W} \otimes \mathbf{IDEN}_I)]\boldsymbol{y}^* + \widetilde{\mathbf{x}}\boldsymbol{b} + \ddot{\mathbf{x}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{2.46}$$

where the operation '.*' in the equation above is used to refer to the element by element multiplication. After further matrix manipulation, we obtain:

$$\boldsymbol{y}^* = \mathbf{S}\widetilde{\mathbf{x}}\boldsymbol{b} + \mathbf{S}(\ddot{\mathbf{x}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}), \text{ where } \mathbf{S} = \left[\mathbf{IDEN}_{QI} - \widetilde{\boldsymbol{\delta}}.*(\mathbf{W} \otimes \mathbf{IDEN}_I)\right]^{-1}. \tag{2.47}$$

The expected value and variance of $\boldsymbol{y}^*$ may be obtained from the above equation after developing the covariance matrix for the error vector $\mathbf{S}(\ddot{\mathbf{x}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon})$. This may be written as $\boldsymbol{\Xi} = \mathbf{S}\left[\ddot{\mathbf{x}}\left(\mathbf{IDEN}_Q \otimes \boldsymbol{\Omega}\right)\ddot{\mathbf{x}}' + \mathbf{IDEN}_Q \otimes \boldsymbol{\Sigma}\right]\mathbf{S}'$. Then, we obtain $\boldsymbol{y}^* \sim MVN_{QI}(\boldsymbol{B}, \boldsymbol{\Xi})$, where $\boldsymbol{B} = \mathbf{S}\widetilde{\mathbf{x}}\boldsymbol{b}$.

The parameter vector to be estimated in the SCMOP model is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Sigma}}', \boldsymbol{\gamma}', \boldsymbol{\alpha}', \boldsymbol{\delta}')'$. *Let* $\boldsymbol{\psi}_{\mathbf{q}}^{\mathbf{up}} = (\psi_{q,m_{q1}}^1, \psi_{q,m_{q2}}^2, ..., \psi_{q,m_{qI}}^I)$ $(I \times 1$ vector), $\boldsymbol{\psi}_{\mathbf{q}}^{\mathbf{low}} = (\psi_{q,m_{q1}-1}^1, \psi_{q,m_{q2}-1}^2, ..., \psi_{q,m_{qI}-1}^I)$ $(I \times 1)$ vector, $\boldsymbol{\psi}^{\mathbf{up}} = (\boldsymbol{\psi}_{\mathbf{1}}^{\mathbf{up}}, \boldsymbol{\psi}_{\mathbf{2}}^{\mathbf{up}}, ..., \boldsymbol{\psi}_{\mathbf{Q}}^{\mathbf{up}})$ $(QI \times 1$ vector), and $\boldsymbol{\psi}^{\mathbf{low}} = (\boldsymbol{\psi}_{\mathbf{1}}^{\mathbf{low}}, \boldsymbol{\psi}_{\mathbf{2}}^{\mathbf{low}}, ..., \boldsymbol{\psi}_{\mathbf{Q}}^{\mathbf{low}})$ $(QI \times 1$ vector). The likelihood function for the SCMOP model is:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y} = \boldsymbol{m}) = \int_{D_{y^*}} f_{QI}(\boldsymbol{y}^* \mid \boldsymbol{B}, \boldsymbol{\Xi}) d\boldsymbol{y}^*, \tag{2.48}$$

where $D_{y^*} = \{\boldsymbol{y}^* : \boldsymbol{\psi}^{\mathbf{low}} < \boldsymbol{y}^* < \boldsymbol{\psi}^{\mathbf{up}}$, and $f_{QI}(.)$ is the multivariate normal density function of dimension *QI*. The dimensionality of the rectangular integral in the likelihood function is *QI*, which is very difficult to evaluate using existing estimation methods. The alternative is to use the pairwise composite marginal likelihood (CML) approach:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{q'=q}^{Q} \prod_{i=1}^{I} \prod_{i'=i}^{I} L_{qq'ii'} \right) \text{ with } q' \neq q \text{ when } i = i', \text{ where}$$

$$L_{qq'ii'} = \begin{bmatrix} \Phi_2(\widetilde{\varphi}_q^i, \widetilde{\varphi}_{q'}^{i'}, \nu_{qq'ii'}) - \Phi_2(\widetilde{\varphi}_q^i, \widetilde{\vartheta}_{q'}^{i'}, \nu_{qq'ii'}) \\ -\Phi_2(\widetilde{\vartheta}_q^i, \widetilde{\varphi}_{q'}^{i'}, \nu_{qq'ii'}) + \Phi_2(\widetilde{\vartheta}_q^i, \widetilde{\vartheta}_{q'}^{i'}, \nu_{qq'ii'}) \end{bmatrix}, \tag{2.49}$$

$$\widetilde{\varphi}_q^i = \frac{\psi_{q,m_{qi}}^i - [\boldsymbol{B}]_{(q-1)\times I+i}}{\sqrt{[\boldsymbol{\Xi}]_{(q-1)\times I+i,(q-1)\times I+i}}}, \widetilde{\vartheta}_q^i = \frac{\psi_{q,m_{qi}-1}^i - [\boldsymbol{B}]_{(q-1)\times I+i}}{\sqrt{[\boldsymbol{\Xi}]_{(q-1)\times I+i,(q-1)\times I+i}}}, \text{ and}$$

$$\nu_{qq'ii'} = \frac{[\boldsymbol{\Xi}]_{(q-1)\times I+i,(q'-1)\times I+i'}}{\sqrt{[\boldsymbol{\Xi}]_{(q-1)\times I+i,(q-1)\times I+i}}\sqrt{[\boldsymbol{\Xi}]_{(q'-1)\times I+i',(q'-1)\times I+i'}}}.$$

The CML estimator is obtained by maximizing the logarithm of the function in Equation (2.49).

The number of pairings in the CML function above is $[QI(QI-1)]/2$. But again the number of pairings can be reduced by determining the "optimal" distance for including pairings across individuals based on minimizing the trace of the asymptotic covariance matrix (as discussed in the previous section).[11] To do so, define a set $\widetilde{C}_q$ as in the previous section that includes the set of individuals *q'* (including *q*) that are within a specified threshold distance of individual *q*. Then, the CML function reduces to the following expression:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{\substack{q'=q \\ q' \in \widetilde{C}_q}}^{Q} \prod_{i=1}^{I} \prod_{i'=i}^{I} L_{qq'ii'} \right) \text{ with } q' \neq q \text{ when } i = i'. \tag{2.50}$$

Let $\widetilde{W}$ be the total number of pairings used in the above CML function (after considering the distance threshold). The covariance matrix of the CML estimator is $\frac{[\hat{\boldsymbol{G}}]^{-1}}{\widetilde{W}} = \frac{[\hat{\boldsymbol{H}}^{-1}][\hat{\boldsymbol{J}}][\hat{\boldsymbol{H}}^{-1}]'}{\widetilde{W}}$, where

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}} \left[\sum_{q=1}^{Q} \sum_{\substack{q'=q \\ q' \in \widetilde{C}_q}}^{Q} \sum_{i=1}^{I} \sum_{i'=i}^{I} \frac{\partial^2 \log L_{qq'ii'}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\hat{\boldsymbol{\theta}}_{CML}} \quad q' \neq q \text{ when } i = i', \tag{2.51}$$

or alternatively,

$$\hat{\boldsymbol{H}} = \frac{1}{\widetilde{W}} \left[\sum_{q=1}^{Q} \sum_{\substack{q'=q \\ q' \in \widetilde{C}_q}}^{Q} \sum_{i=1}^{I} \sum_{i'=i}^{I} \left(\left[\frac{\partial \log L_{qq'ii'}}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial \log L_{qq'ii'}}{\partial \boldsymbol{\theta}'} \right] \right) \right]_{\hat{\boldsymbol{\theta}}_{CML}} \quad q' \neq q \text{ when } i = i'. \tag{2.52}$$

The sandwich matrix, $\hat{\boldsymbol{J}}$, may be computed by selecting $\widetilde{Q}$ $(\tilde{q} = 1,2,...,\widetilde{Q})$ observational units from the original *Q* observational units as discussed in the earlier section. Let $\widetilde{C}_{\tilde{q}}$ be the set of individuals (observation units) within the specified threshold distance, and let $N_{\tilde{q}}$ be the cardinality of $\widetilde{C}_{\tilde{q}}$. Let $l_{\tilde{q}}$ be an index so that $l_{\tilde{q}} = 1,2,\ldots N_{\tilde{q}}$. Next, define $\vec{C}_{\tilde{q}} = [(N_{\tilde{q}} I)((N_{\tilde{q}} I) - 1)]/2$. Then, the **J** matrix maybe empirically estimated as:

[11] Technically, one can consider a threshold distance separately for each ordinal variable, so that the individual pairings within each variable are based on this variable-specific threshold distance and the individual-variable pairings across variables are based on different thresholds across variables. But this gets cumbersome, and so we will retain a single threshold distance across all ordinal variables.

$$\hat{\mathbf{J}} = \frac{1}{\widetilde{Q}}\left[\sum_{\tilde{q}=1}^{\widetilde{Q}}\left[\frac{1}{\vec{\tilde{C}}_{\tilde{q}}}\left(\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{i=1}^{I}\sum_{i'=i}^{I}\frac{\partial \log L_{l_{\tilde{q}}l'ii'}}{\partial \boldsymbol{\theta}}\right]\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{i=1}^{I}\sum_{i'=i}^{I}\frac{\partial \log L_{l_{\tilde{q}}l'ii'}}{\partial \boldsymbol{\theta}'}\right]\right)\right]\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \tag{2.53}$$

There is another way that the analyst can consider cutting down the number of pairings even after using a threshold distance as a cut-off. That is by ignoring the pairings among different individuals (observation units) across the I ordinal variables. This will reduce the number of pairings quite substantially, while also retaining the pairings across individuals for each ordinal variable (that enables the estimation of the parameters of the vector $\boldsymbol{\delta}$) and the pairings across ordinal variables within the same individual (that enables the estimation of the parameters in the correlation matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\varepsilon}_{\boldsymbol{q}}$). The CML is:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q-1}\prod_{\substack{q'=q+1 \\ q'\in\tilde{C}_q}}^{Q}\prod_{i=1}^{I} L_{qq'i}\right)\left(\prod_{q=1}^{Q}\prod_{i=1}^{I-1}\prod_{i'=i+1}^{I} L_{qii'}\right) \tag{2.54}$$

The number of pairings $\widetilde{W}$ in the CML function above is much smaller than the CML function in Equation (2.50). The elements of the covariance matrix may be estimated as follows:

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q-1}\sum_{\substack{q'=q+1 \\ q'\in\tilde{C}_q}}^{Q}\sum_{i=1}^{I}\frac{\partial^2 \log L_{qq'i}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \sum_{q=1}^{Q}\sum_{i=1}^{I-1}\sum_{i'=i+1}^{I}\frac{\partial^2 \log L_{qii'}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{CML}}, \tag{2.55}$$

or alternatively,

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q-1}\sum_{\substack{q'=q+1 \\ q'\in\tilde{C}_q}}^{Q}\sum_{i=1}^{I}\left(\left[\frac{\partial \log L_{qq'i}}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qq'i}}{\partial\boldsymbol{\theta}'}\right]\right) + \sum_{q=1}^{Q}\sum_{i=1}^{I-1}\sum_{i'=i+1}^{I}\left(\left[\frac{\partial \log L_{qii'}}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qii'}}{\partial\boldsymbol{\theta}'}\right]\right)\right]_{\hat{\boldsymbol{\theta}}_{CML}}. \tag{2.56}$$

For estimating the $\hat{\mathbf{J}}$ matrix define $\widetilde{\mathbf{C}}_{\tilde{\boldsymbol{q}}}$ and $N_{\tilde{q}}$ be defined as earlier and let $\vec{\tilde{C}}_{\tilde{q}} = \left[N_{\tilde{q}}(N_{\tilde{q}}-1)/2\right]I + \left[I(I-1)/2\right]Q$ and $\vec{\tilde{\boldsymbol{S}}}_{\tilde{\boldsymbol{q}}} = \left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{i=1}^{I}\frac{\partial \log L_{l_{\tilde{q}}l'i}}{\partial\boldsymbol{\theta}} + \sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{i=1}^{I-1}\sum_{i'=i+1}^{I}\frac{\partial \log L_{l_{\tilde{q}}l'ii'}}{\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}$.

Then,

$$\hat{\mathbf{J}} = \frac{1}{\widetilde{Q}}\left[\sum_{\tilde{q}=1}^{\widetilde{Q}}\left[\frac{1}{\vec{\tilde{C}}_{\tilde{q}}}\left(\vec{\tilde{\boldsymbol{S}}}_{\tilde{\boldsymbol{q}}}\vec{\tilde{\boldsymbol{S}}}'_{\tilde{\boldsymbol{q}}}\right)\right]\right]. \tag{2.57}$$

The positive-definiteness of the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are ensured as discussed in Sections 2.3.1.1 and 2.2.1.2.

References for the CML Estimation of the Spatial CMOP (or SCMOP) Model

No known applications. But the spatial cross-sectional multivariate count model of Narayanamoorthy *et al.* (2013) is very similar to the SCMOP model."

2.3.1.3. The Spatial PMOP (SPMOP) Model

All notations from Section 2.2.1.3 are carried over. To include spatial dependency in the PMOP model, rewrite Equation (2.12) as follows:

$$y_{qt}^{*} = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q't}^{*} + \boldsymbol{\beta}_{\boldsymbol{q}}' \boldsymbol{x}_{\boldsymbol{qt}} + \varepsilon_{qt},\ y_{qt} = k \text{ if } \psi_{q,t,k-1} < y_{qj}^{*} < \psi_{q,t,k}, \tag{2.58}$$

Define $\boldsymbol{y}_q = (y_{q1}, y_{q2}, ..., y_{qT})'$ ($T \times 1$ matrix), $\boldsymbol{\varepsilon}_{\boldsymbol{q}} = (\varepsilon_{q1}, \varepsilon_{q2}, ..., \varepsilon_{qT})'$ ($T \times 1$ matrix), $\boldsymbol{y}_{\boldsymbol{q}}^{*} = (y_{q1}^{*}, y_{q2}^{*}, ..., y_{qT}^{*})'$ ($T \times 1$ matrix), $\boldsymbol{x}_{\boldsymbol{q}} = (\boldsymbol{x}_{\boldsymbol{q1}}, \boldsymbol{x}_{\boldsymbol{q2}}, ..., \boldsymbol{x}_{\boldsymbol{qT}})'$ ($T \times L$ matrix), $\boldsymbol{\psi}_{\mathbf{q}}^{\mathbf{up}} = (\psi_{q,1,m_{qi}}, \psi_{q,2,m_{q2}}, ..., \psi_{q,T,m_{qJ}})$ ($T \times 1$ vector), $\boldsymbol{\psi}_{\mathbf{q}}^{\mathbf{low}} = (\psi_{q,1,m_{qi}-1}, \psi_{q,2,m_{q2}-1}, ..., \psi_{q,T,m_{qJ}-1})$ ($T \times 1$) vector. Also, let the vector of actual observed ordinal outcomes for individual q be stacked into a ($T \times 1$) vector $\boldsymbol{m}_q = (m_{q1}, m_{q2}, ..., m_{qT})'$. To write the equation system in (2.58) compactly, we next define several additional vectors and matrices. Let $\boldsymbol{y}^{*} = [(\boldsymbol{y}_1^{*})', (\boldsymbol{y}_2^{*})', (\boldsymbol{y}_3^{*})', ..., (\boldsymbol{y}_Q^{*})']'$ ($QT \times 1$ vector), $\boldsymbol{y} = [(\boldsymbol{y}_1)', (\boldsymbol{y}_2)', (\boldsymbol{y}_3)', ..., (\boldsymbol{y}_Q)']'$ ($QT \times 1$ vector), $\boldsymbol{m} = (\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_{\boldsymbol{Q}})'$ ($QT \times 1$ vector), $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3', ...\mathbf{x}_{\boldsymbol{Q}}')'$ ($QT \times L$ matrix), $\boldsymbol{\beta}_q = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_q$, $\widetilde{\boldsymbol{\beta}}_q \sim MVN_L(0, \boldsymbol{\Omega})$ (the $\widetilde{\boldsymbol{\beta}}_q$ random coefficients are independent across individuals), $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}_1', \widetilde{\boldsymbol{\beta}}_2', ..., \widetilde{\boldsymbol{\beta}}_Q')'$ ($QL \times 1$ vector), $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \boldsymbol{\varepsilon}_3', \ldots, \boldsymbol{\varepsilon}_{\boldsymbol{Q}}')'$ ($QT \times 1$ vector),

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{x}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{x}_{\mathbf{Q}} \end{bmatrix} \ (QT \times QL \text{ block diagonal matrix}), \tag{2.59}$$

Also, collect all the weights $w_{qq'}$ into a spatial weight matrix **W**. The vector $\widetilde{\boldsymbol{\beta}}$ above has a mean vector of zero and a covariance matrix $\mathbf{IDEN}_Q \otimes \boldsymbol{\Omega}$ (of size $QT \times QT$), while the vector $\boldsymbol{\varepsilon}$ has a mean vector of zero and a covariance matrix $\mathbf{IDEN}_{QT}$.

Using the vector and the matrix notations defined above, Equation (2.58) may be re-written compactly as:

$$\boldsymbol{y}^{*} = [\delta(\mathbf{W} \otimes \mathbf{IDEN}_T)]\boldsymbol{y}^{*} + \mathbf{x}\boldsymbol{b} + \tilde{\mathbf{x}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{2.60}$$

After further matrix manipulation, we obtain:

$$\boldsymbol{y}^* = \mathbf{S}\mathbf{x}\boldsymbol{b} + \mathbf{S}(\tilde{\mathbf{x}}\tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}), \text{ where } \mathbf{S} = \left[\mathbf{IDEN}_{QT} - \delta(\mathbf{W} \otimes \mathbf{IDEN}_{T})\right]^{-1}. \tag{2.61}$$

Next, we obtain $\boldsymbol{y}^* \sim MVN_{QI}(\boldsymbol{B}, \Xi)$, where

$$\boldsymbol{B} = \mathbf{S}\mathbf{x}\boldsymbol{b} \text{ and } \Xi = \mathbf{S}\left[\tilde{\mathbf{x}}\left(\mathbf{IDEN}_{Q} \otimes \boldsymbol{\Omega}\right)\tilde{\mathbf{x}}' + \mathbf{IDEN}_{QT}\right]\mathbf{S}' \tag{2.62}$$

The parameter vector to be estimated in the SPMOP model is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \boldsymbol{\gamma}', \boldsymbol{\alpha}', \delta)'$. Let $\boldsymbol{\psi}^{\mathbf{up}} = (\boldsymbol{\psi}_{\mathbf{1}}^{\mathbf{up}}, \boldsymbol{\psi}_{\mathbf{2}}^{\mathbf{up}}, ..., \boldsymbol{\psi}_{\mathbf{Q}}^{\mathbf{up}})$ ($QT \times 1$ vector), and $\boldsymbol{\psi}^{\mathbf{low}} = (\boldsymbol{\psi}_{\mathbf{1}}^{\mathbf{low}}, \boldsymbol{\psi}_{\mathbf{2}}^{\mathbf{low}}, ..., \boldsymbol{\psi}_{\mathbf{Q}}^{\mathbf{low}})$ ($QT \times 1$ vector). The likelihood function for the SPMOP model is:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y} = \boldsymbol{m}) = \int_{D_{y^*}} f_{QT}(\boldsymbol{y}^* \mid \boldsymbol{B}, \Xi) d\boldsymbol{y}^*, \tag{2.63}$$

where $D_{y^*} = \{\boldsymbol{y}^* : \boldsymbol{\psi}^{\mathbf{low}} < \boldsymbol{y}^* < \boldsymbol{\psi}^{\mathbf{up}}$, and $f_{QT}(.)$ is the multivariate normal density function of dimension QT. The much simpler pairwise composite marginal likelihood (CML) function is:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{q'=q}^{Q} \prod_{t=1}^{T} \prod_{t'=t}^{T} L_{qq'tt'} \right) \text{ with } q' \neq q \text{ when } t = t', \text{ where}$$

$$L_{qq'tt'} = \begin{bmatrix} \Phi_2(\widetilde{\varphi}_{qt}, \widetilde{\varphi}_{q't'}, \nu_{qq'tt'}) - \Phi_2(\widetilde{\varphi}_{qt}, \widetilde{\vartheta}_{q't'}, \nu_{qq'tt'}) \\ -\Phi_2(\widetilde{\vartheta}_{qt}, \widetilde{\varphi}_{q't'}, \nu_{qq'tt'}) + \Phi_2(\widetilde{\vartheta}_{qt}, \widetilde{\vartheta}_{q't'}, \nu_{qq'tt'}) \end{bmatrix},$$

$$\widetilde{\varphi}_{qt} = \frac{\psi_{q,t,m_{qt}} - [\boldsymbol{B}]_{(q-1)\times T+t}}{\sqrt{[\Xi]_{(q-1)\times T+t,(q-1)\times T+t}}}, \widetilde{\vartheta}_{qt} = \frac{\psi_{q,t,m_{qt}-1} - [\boldsymbol{B}]_{(q-1)\times T+t}}{\sqrt{[\Xi]_{(q-1)\times T+t,(q-1)\times I+t}}}, \text{ and} \tag{2.64}$$

$$\nu_{qq'tt'} = \frac{[\Xi]_{(q-1)\times T+t,(q'-1)\times T+t'}}{\sqrt{[\Xi]_{(q-1)\times T+t,(q-1)\times T+t}}\sqrt{[\Xi]_{(q'-1)\times T+t',(q'-1)\times T+t'}}}.$$

To reduce the number of pairings, define a set $\widetilde{C}_q$ as in the previous section that includes the set of individuals q' (including q) that are within a specified threshold distance of individual q. Then, the CML function reduces to the following expression:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{\substack{q'=q \\ q' \in \widetilde{C}_q}}^{Q} \prod_{t=1}^{T} \prod_{t'=i}^{T} L_{qq'tt'} \right) \text{ with } q' \neq q \text{ when } t = t'. \tag{2.65}$$

Let $\widetilde{W}$ be the total number of pairings used in the above CML function (after considering the distance threshold). The covariance matrix of the CML estimator is $\frac{[\hat{\boldsymbol{G}}]^{-1}}{\widetilde{W}} = \frac{[\hat{\boldsymbol{H}}^{-1}][\hat{\boldsymbol{J}}][\hat{\boldsymbol{H}}^{-1}]'}{\widetilde{W}}$, where

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q}\sum_{\substack{q'=q \\ q'\in\widetilde{C}_q}}^{Q}\sum_{t=1}^{T}\sum_{t'=t}^{T}\frac{\partial^2 \log L_{qq'tt'}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{CML}} \quad q' \neq q \text{ when } t = t', \tag{2.66}$$

or alternatively,

$$\hat{\boldsymbol{H}} = \frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q}\sum_{\substack{q'=q \\ q'\in\widetilde{C}_q}}^{Q}\sum_{t=1}^{T}\sum_{t'=t}^{T}\left(\left[\frac{\partial \log L_{qq'tt'}}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qq'tt'}}{\partial\boldsymbol{\theta}'}\right]\right)\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \quad q' \neq q \text{ when } t = t'. \tag{2.67}$$

Defining $\widetilde{C}_{\tilde{q}}$, $N_{\tilde{q}}$, and $\ddot{C}_{\tilde{q}} = \left[\left(N_{\tilde{q}}I\right)\left(\left(N_{\tilde{q}}I\right)-1\right)\right]/2$ as in the previous section, the **J** matrix maybe empirically estimated as:

$$\hat{\mathbf{J}} = \frac{1}{\widetilde{Q}}\left[\sum_{\tilde{q}=1}^{\widetilde{Q}}\left[\frac{1}{\ddot{C}_{\tilde{q}}}\left(\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{t=1}^{T}\sum_{t'=t}^{T}\frac{\partial \log L_{l_{\tilde{q}}l'tt'}}{\partial\boldsymbol{\theta}}\right]\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{t=1}^{T}\sum_{t'=t}^{T}\frac{\partial \log L_{l_{\tilde{q}}l'tt'}}{\partial\boldsymbol{\theta}'}\right]\right)\right]\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}} \tag{2.68}$$

One can also ignore the pairings among different individuals (observation units) across the *T* time periods. The CML then is:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q-1}\prod_{\substack{q'=q+1 \\ q'\in\widetilde{C}_q}}^{Q}\prod_{t=1}^{T}L_{qq't}\right)\left(\prod_{q=1}^{Q}\prod_{t=1}^{T-1}\prod_{t'=t+1}^{T}L_{qtt'}\right) \tag{2.69}$$

The elements of the covariance matrix in this case may be estimated as follows:

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q-1}\sum_{\substack{q'=q+1 \\ q'\in\widetilde{C}_q}}^{Q}\sum_{t=1}^{T}\frac{\partial^2 \log L_{qq't}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \sum_{q=1}^{Q}\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\frac{\partial^2 \log L_{qtt'}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{CML}}, \tag{2.70}$$

or alternatively,

$$\hat{\boldsymbol{H}} = -\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q-1}\sum_{\substack{q'=q+1 \\ q'\in\widetilde{C}_q}}^{Q}\sum_{t=1}^{T}\left(\left[\frac{\partial \log L_{qq't}}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qq't}}{\partial\boldsymbol{\theta}'}\right]\right) + \sum_{q=1}^{Q}\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\left(\left[\frac{\partial \log L_{qtt'}}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log L_{qtt'}}{\partial\boldsymbol{\theta}'}\right]\right)\right]_{\hat{\boldsymbol{\theta}}_{CML}}, \tag{2.71}$$

For estimating the $\hat{\mathbf{J}}$ matrix, define $\widetilde{C}_{\tilde{q}}$ and $N_{\tilde{q}}$ as earlier and let $\ddot{C}_{\tilde{q}} = \left[N_{\tilde{q}}(N_{\tilde{q}}-1)/2\right]I + \left[I(I-1)/2\right]Q$ and $\ddot{\boldsymbol{S}}_{\tilde{\boldsymbol{q}}} = \left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{l'=l_{\tilde{q}}}^{N_{\tilde{q}}}\sum_{t=1}^{T}\frac{\partial \log L_{l_{\tilde{q}}l't}}{\partial\boldsymbol{\theta}} + \sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}}\sum_{i=1}^{T-1}\sum_{t'=t+1}^{T}\frac{\partial \log L_{l_{\tilde{q}}l'tt'}}{\partial\boldsymbol{\theta}'}\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}$.

Then,

$$\hat{\mathbf{J}} = \frac{1}{\tilde{\tilde{Q}}}\left[\sum_{\tilde{q}=1}^{\tilde{Q}}\left[\frac{1}{\tilde{\tilde{C}}_{\tilde{q}}}\left(\tilde{\tilde{\boldsymbol{S}}}_{\tilde{\boldsymbol{q}}}\tilde{\tilde{\boldsymbol{S}}}'_{\tilde{\boldsymbol{q}}}\right)\right]\right]. \tag{2.72}$$

References for the CML Estimation of the Spatial PMOP (SPMOP) Model

Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis and Prevention* 52, 188-203.

Ferdous, N., Bhat, C.R., 2013. A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. *Journal of Geographical Systems* 15(1), 1-29.

Paleti, R., Bhat, C.R., Pendyala, R.M., Goulias, K.G., 2013. Modeling of household vehicle type choice accommodating spatial dependence effects. *Transportation Research Record* 2343, 86-94.

2.3.2. Unordered-Response Models

2.3.2.1. The Spatial CMNP (SCMNP) Model

The formulation in this case is similar to the aspatial case in Section 2.2.2.1, with the exception that a spatial lag term is included. Of course, this also completely changes the model structure from the aspatial case.

$$U_{qi} = \delta\sum_{q'} w_{qq'}U_{q'i} + \boldsymbol{\beta}'_{\boldsymbol{q}}\boldsymbol{x}_{\boldsymbol{qi}} + \xi_{qi}; \boldsymbol{\beta}_{\boldsymbol{q}} = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}, \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}} \sim MVN_L(\mathbf{0},\boldsymbol{\Omega});\ |\delta| < 1, \tag{2.73}$$

where all notations are the same as in Section 2.2.2.1.[12] Let $\boldsymbol{\xi}_{\boldsymbol{q}} = (\xi_{q1},\xi_{q2},...,\xi_{qI})'$ ($I \times 1$ vector). Then, we assume $\boldsymbol{\xi}_{\boldsymbol{q}} \sim MVN_I(0,\boldsymbol{\Lambda})$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Lambda}$ for identifiability, as discussed in Section 2.2.2.1. The model above may be written in a more compact form by defining the following vectors and matrices: $\mathbf{U}_q = (U_{q1},U_{q2},...,U_{qI})'$ ($I \times 1$ vector), $\mathbf{U} = (\mathbf{U}_1,\mathbf{U}_2,...,\mathbf{U}_Q)'$ ($QI \times 1$ vector), $\boldsymbol{\xi} = (\boldsymbol{\xi}'_1,\boldsymbol{\xi}'_2,...,\boldsymbol{\xi}'_Q)'$ ($QI \times 1$ vector), $\boldsymbol{x}_{\boldsymbol{q}} = (\boldsymbol{x}_{\boldsymbol{q1}},\boldsymbol{x}_{\boldsymbol{q2}},\boldsymbol{x}_{\boldsymbol{q3}},...,\boldsymbol{x}_{\boldsymbol{qI}})'$ ($I \times L$ matrix), $\mathbf{x} = (\boldsymbol{x}'_1,\boldsymbol{x}'_2,...,\boldsymbol{x}'_Q)'$ ($QI \times L$ matrix), and $\tilde{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{\beta}}'_1,\tilde{\boldsymbol{\beta}}'_2,...,\tilde{\boldsymbol{\beta}}'_Q\right)'$ ($QL \times 1$ vector). Also, define the following matrix:

[12] One can allow the spatial lag dependence parameter δ to vary across alternatives i. However, due to identification considerations, one of the alternatives should be used as the base (with a zero dependence parameter). But doing so while also allowing the dependence parameters to vary across the remaining alternatives creates exchangeability problems, since the model estimation results will not be independent of the decision of which alternative is considered as the base. Hence, we prefer the specification that restricts the dependence parameter to be the same across alternatives i.

$$\vec{\mathbf{x}} = \begin{bmatrix} \boldsymbol{x}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}_2 & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{x}_3 & \mathbf{0} \cdots \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \ldots \boldsymbol{x}_Q \end{bmatrix} \ (QI \times QL \text{ matrix}), \tag{2.74}$$

Then, we can write Equation (2.73) in matrix form as:

$$\mathbf{U} = \mathbf{S}\left[\mathbf{x}\boldsymbol{b} + \vec{\mathbf{x}}\tilde{\boldsymbol{\beta}} + \boldsymbol{\xi}\right], \tag{2.75}$$

where $\mathbf{S} = \left[\mathbf{IDEN}_{QI} - (\delta \mathbf{W} \otimes \mathbf{IDEN}_I)\right]^{-1}$ ($QI \times QI$ matrix), and $\mathbf{W}$ is the $(Q \times Q)$ weight matrix with the weights $w_{qq'}$ as its elements. Also, $\mathbf{U} \sim MVN_{QI}(\mathbf{V}, \tilde{\boldsymbol{\Xi}})$, where $\mathbf{V} = \mathbf{S}\mathbf{x}\boldsymbol{b}$ and $\tilde{\boldsymbol{\Xi}} = \mathbf{S}\left[\vec{\mathbf{x}}(\mathbf{IDEN}_Q \otimes \boldsymbol{\Omega})\vec{\mathbf{x}}' + (\mathbf{IDEN}_Q \otimes \boldsymbol{\Lambda})\right]\mathbf{S}'$. Let $\mathbf{u}_q = (u_{q1}, u_{q2}, \ldots u_{qI})'$ $(i \neq m_q)$ be an $(I–1)\times 1$ vector for individual q, where m_q is the actual observed choice of individual q and $u_{qi} = U_{qi} - U_{qm_q}$ $(i \neq m_q)$. Stack the $\mathbf{u}_q$ vectors across individuals (observation units): $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_Q')'$ $[Q(I-1)\times 1 \text{ Vector}]$. The distribution of $\mathbf{u}$ may be derived from the distribution of $\mathbf{U}$ by defining a $[Q \times (I-1)] \times [QI]$ block diagonal matrix $\mathbf{M}$, with each block diagonal having $(I-1)$ rows and I columns corresponding to each individual q. This $(I-1)\times I$ matrix for individual q corresponds to an $(I-1)$ identity matrix with an extra column of '-1' values added as the m_q th column. For instance, consider the case of I = 4 and Q = 2. Let individual 1 be observed to choose alternative 2 and individual 2 be observed to choose alternative 1. Then $\mathbf{M}$ takes the form below.

$$\mathbf{M} = \left[\begin{array}{cccc|cccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right]. \tag{2.76}$$

With the above construction for matrix $\mathbf{M}$, it is easy to see that $\mathbf{u} \sim MVN_{Q(I-1)}(\boldsymbol{B}, \boldsymbol{\Xi})$, where $\boldsymbol{B} = \mathbf{M}\mathbf{V}$ and $\boldsymbol{\Xi} = \mathbf{M}\tilde{\boldsymbol{\Xi}}\mathbf{M}'$. The likelihood of the observed sample (*i.e.*, individual 1 choosing alternative m_1, individual 2 choosing alternative m_2, …, individual Q choosing alternative m_Q) may then be written succinctly as $\text{Prob}[\mathbf{u} < \mathbf{0}_{Q(I-1)}]$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Lambda}}', \delta)'$. Using the usual notations, the likelihood function is:

$$L(\boldsymbol{\theta}) = \Phi_{Q(I-1)}(\boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}(-\mathbf{B}\ ), \boldsymbol{\Xi}^{*}), \tag{2.77}$$

where $\boldsymbol{\Xi}^{*} = \boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Xi}\boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}$. This is cumbersome and impractical (if not infeasible) for most realistically-sized sample sizes. However, one can use the MACML technique. To do so, write the pairwise CML function corresponding to the full likelihood of Equation (2.77) as:

$$L_{CML}^{SCMNP}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1}\prod_{q'=q+1}^{Q} L_{qq'}, \text{ where } L_{qq'} = \Pr(d_q = m_q, d_{q'} = m_{q'}), \tag{2.78}$$

d_q is an index for the individual's choice, and

$$\Pr(d_q = m_q, d_{q'} = m_{q'}) = \Phi_{\bar{J}}(\boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'}}^{-1}(\text{-}\ddot{\boldsymbol{B}}_{qq'}), \ddot{\boldsymbol{\Xi}}_{qq'}^{*}), \tag{2.79}$$

where $\ddot{J} = 2(I-1)$, $\ddot{\boldsymbol{B}}_{qq'} = \boldsymbol{\Delta}_{qq'}\boldsymbol{B}$, $\ddot{\boldsymbol{\Xi}}_{\mathbf{qq'}} = \boldsymbol{\Delta}_{qq'}\boldsymbol{\Xi}\boldsymbol{\Delta}_{qq'}^{'}$, $\ddot{\boldsymbol{\Xi}}_{\mathbf{qq'}}^{*} = \boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'}}^{-1}\ddot{\boldsymbol{\Xi}}_{\mathbf{qq'}}\boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'}}^{-1}$, and $\boldsymbol{\Delta}_{qq'}$ is a $\ddot{J} \times Q(I-1)$-selection matrix with an identity matrix of size $(I-1)$ occupying the first $(I-1)$ rows and the $\left[(q-1)\times(I-1)+1\right]^{th}$ through $\left[q\times(I-1)\right]^{th}$ columns, and another identity matrix of size $(I-1)$ occupying the last $(I-1)$ rows and the $\left[(q'-1)\times(I-1)+1\right]^{th}$ through $\left[q'\times(I-1)\right]^{th}$ columns.

The number of pairings in the CML expression of Equation (2.78) can be reduced as explained in Section 2.3.1.1. Specifically, define a set $\widetilde{C}_q$ as in the previous section that includes the set of individuals *q'* (including *q*) that are within a specified threshold distance of individual q. Then, the CML function reduces to the following expression:

$$L_{CML}^{SCMNP}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1}\prod_{\substack{q'=q+1 \\ q'\in\widetilde{C}_q}}^{Q} L_{qq'}. \tag{2.80}$$

The expressions to obtain the covariance matrix are exactly the same as in Section 2.3.1.1, with $L_{qq'} = \Phi_{\bar{J}}(\boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'}}^{-1}(\text{-}\ddot{\boldsymbol{B}}_{qq'}), \ddot{\boldsymbol{\Xi}}_{qq'}^{*})$.

References for the CML Estimation of the Spatial MNP (SMNP) Model

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

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

Sener, I.N., Bhat, C.R., 2012. Flexible spatial dependence structures for unordered multinomial choice models: formulation and application to teenagers' activity participation. *Transportation* 39(3), 657-683.

Sidharthan, R., Bhat, C.R, Pendyala, R.M., Goulias, K.G., 2011. Model for children's school travel mode choice: accounting for effects of spatial and social interaction. *Transportation Research Record* 2213, 78-86.

2.3.2.2. The Spatial CMMNP Model

Rewrite Equation (2.18) from Section 2.2.2.2 to include spatial dependency in the utility that individual q attributes to alternative i_g (i_g=1,2,..., I_g) for the g^{th} variable.

$$U_{qgi_g} = \delta_g \sum_{q'=1}^{Q} w_{qq'} U_{qgi_g} + \boldsymbol{\beta}'_{\boldsymbol{qg}} \boldsymbol{x}_{\boldsymbol{qgi_g}} + \xi_{qgi_g}, \tag{2.81}$$

with all notations as earlier. $\boldsymbol{x}_{\boldsymbol{qgi_g}}$ is an $L_g \times 1$-column vector of exogenous attributes, $\boldsymbol{\beta}_{\boldsymbol{qg}} \sim MVN_{L_q}(\boldsymbol{b_g}, \boldsymbol{\Omega_g})$, and $\boldsymbol{\xi}_{\boldsymbol{qg}} \sim MVN_I(0, \boldsymbol{\Lambda_g})$ ($\boldsymbol{\xi}_{\boldsymbol{qg}} = (\xi_{qg1}, \xi_{qg2}, ...\xi_{qgI_g})'$ ($I_g \times 1$ vector)). As in Section 2.2.2.2, we will assume that the $\boldsymbol{\beta}_{\boldsymbol{qg}} (= \boldsymbol{b_g} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{qg}})$ vectors are independent across the unordered-response dimensions for each individual. We also assume that ξ_{qgi_g} is independent and identically normally distributed across individuals q. Let m_{qg} be the actual chosen alternative for the gth unordered-response variable by individual q. Define the following:

$\mathbf{U}_{qg} = (U_{qg1}, U_{qg2}, ..., U_{qgI_g})'$ ($I_g \times 1$ vector), $\mathbf{U}_q = (\mathbf{U}'_{q1}, \mathbf{U}'_{q2}, ...\mathbf{U}'_{qG})'$ $\widetilde{G} \times 1$ vector $\left(\widetilde{G} = \left(\sum_{g=1}^{G} I_g\right)\right)$, $\boldsymbol{\xi}_{\boldsymbol{q}} = (\boldsymbol{\xi}_{\boldsymbol{q1}}, \boldsymbol{\xi}_{\boldsymbol{q2}}, ...\boldsymbol{\xi}_{\boldsymbol{qG}})'$ ($\widetilde{G} \times 1$ vector), $\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2, ...\mathbf{U}'_Q)'$ ($Q\widetilde{G} \times 1$ vector),

$\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_Q)'$ ($Q\widetilde{G} \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_G)'$ ($\widetilde{L} \times 1$ vector) $\left(\widetilde{L} = \left(\sum_{g=1}^{G} L_g\right)\right)$,

$\boldsymbol{x}_{qg} = (\boldsymbol{x}_{qg1}, \boldsymbol{x}_{qg2}, ..., \boldsymbol{x}_{qgI_g})'$ ($I_g \times L_g$ matrix),

$$\boldsymbol{x}_q = \begin{bmatrix} \boldsymbol{x}_{q1} & 0 & 0 & 0. \cdots 0 \\ 0 & \boldsymbol{x}_{q2} & 0 & 0 \cdots 0 \\ 0 & 0 & \boldsymbol{x}_{q3} & 0 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \boldsymbol{x}_{qG} \end{bmatrix} (\widetilde{G} \times \widetilde{L} \text{ matrix}), \quad \tilde{\mathbf{x}} = \begin{bmatrix} \boldsymbol{x}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}_2 & \mathbf{0} & \mathbf{0} \cdots \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{x}_3 & \mathbf{0} \cdots \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \ldots \boldsymbol{x}_{\mathbf{Q}} \end{bmatrix} (Q\widetilde{G} \times Q\widetilde{L} \text{ matrix}),$$

$\mathbf{x} = (\boldsymbol{x}'_1, \boldsymbol{x}'_2, ..., \boldsymbol{x}'_Q)'$ ($Q\widetilde{G} \times \widetilde{L}$ matrix), $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}} = \left(\widetilde{\boldsymbol{\beta}}'_{\boldsymbol{q1}}, \widetilde{\boldsymbol{\beta}}'_{\boldsymbol{q2}}, ..., \widetilde{\boldsymbol{\beta}}'_{\boldsymbol{qG}}\right)'$ ($\widetilde{L} \times 1$ vector), and $\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\boldsymbol{\beta}}'_1, \widetilde{\boldsymbol{\beta}}'_2, ..., \widetilde{\boldsymbol{\beta}}'_Q\right)'$ ($Q\widetilde{L} \times 1$ vector). Let $\boldsymbol{\xi}_{\boldsymbol{q}} \sim MVN_{\widetilde{G}}(\mathbf{0}_{\widetilde{\mathbf{G}}}, \boldsymbol{\Lambda})$, where the covariance matrix $\boldsymbol{\Lambda}$ is to be constructed as discussed in Section 2.2.2.2. Then, $\boldsymbol{\xi} \sim MVN_{Q\widetilde{G}}(\mathbf{0}_{\mathbf{Q}\widetilde{\mathbf{G}}}, \mathbf{IDEN}_{\mathbf{Q}} \otimes \boldsymbol{\Lambda})$. Also, define $\widetilde{\boldsymbol{\Omega}}_{\mathbf{qg}} = \mathbf{x}_{\mathbf{qg}} \boldsymbol{\Omega}_{\mathbf{g}} \mathbf{x}'_{\mathbf{qg}}$ ($I_g \times I_g$ matrix), and the following matrices:

$$\widetilde{\mathbf{\Omega}}_{\mathbf{q}} = \begin{bmatrix} \widetilde{\mathbf{\Omega}}_{\mathbf{q1}} & 0 & 0 & 0.\cdots 0 \\ 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{q2}} & 0 & 0\cdots 0 \\ 0 & 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{q3}} & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \widetilde{\mathbf{\Omega}}_{\mathbf{qG}} \end{bmatrix} (\widetilde{G}\times\widetilde{G} \text{ matrix}), \widetilde{\mathbf{\Omega}} = \begin{bmatrix} \widetilde{\mathbf{\Omega}}_{\mathbf{1}} & 0 & 0 & 0.\cdots 0 \\ 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{2}} & 0 & 0\cdots 0 \\ 0 & 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{3}} & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \widetilde{\mathbf{\Omega}}_{\mathbf{Q}} \end{bmatrix}, (Q\widetilde{G}\times Q\widetilde{G} \text{ matrix}),$$

and

$$\widetilde{\boldsymbol{\delta}} = \begin{bmatrix} \delta_1 & 0 & 0 & 0.\cdots 0 \\ 0 & \delta_2 & 0 & 0\cdots 0 \\ 0 & 0 & \delta_3 & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \delta_G \end{bmatrix} (\widetilde{G}\times\widetilde{G} \text{ matrix}),$$

Equation (2.81) may then be written in matrix form as:

$$\mathbf{U} = \mathbf{S}\left[\mathbf{x}\boldsymbol{b} + \tilde{\mathbf{x}}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\xi}\right], \tag{2.82}$$

where $\mathbf{S} = \left[\mathbf{IDEN}_{\mathbf{Q\tilde{G}}} - (\mathbf{1}_{\mathbf{QQ}} \otimes \widetilde{\boldsymbol{\delta}}).*(\mathbf{W} \otimes \mathbf{IDEN}_{\tilde{\mathbf{G}}})\right]$, $\mathbf{W}$ is the $(Q\times Q)$ weight matrix with the weights $w_{qq'}$, and ".*" refers to the element-by-element multiplication of the two matrices involved. Also, $\mathbf{U} \sim MVN_{Q\tilde{G}}(\mathbf{V}, \widetilde{\mathbf{\Xi}})$, where $\mathbf{V} = \mathbf{S}\mathbf{x}\boldsymbol{b}$ and $\widetilde{\mathbf{\Xi}} = \mathbf{S}\left[\widetilde{\mathbf{\Omega}} + (\mathbf{IDEN}_{\mathbf{Q}} \otimes \mathbf{\Lambda})\right]\mathbf{S}'$.[13]

To develop the likelihood function, construct a matrix $\mathbf{M}$ as follows. First, for each unordered variable g and individual q, construct a matrix $\mathbf{M}_{\mathbf{qg}}$ with $(I_g - 1)$ rows and I_g columns. This matrix corresponds to an $(I_g - 1)$ identity matrix with an extra column of '-1' values added as the m_{qg}^{th} column. Then, define the following:

[13] One can also obtain $\widetilde{\mathbf{\Omega}}$ as $\widetilde{\mathbf{\Omega}} = \tilde{\mathbf{x}}\left(\mathbf{IDEN}_{\mathbf{Q}} \otimes \begin{bmatrix} \widetilde{\mathbf{\Omega}}_{\mathbf{1}} & 0 & 0 & 0.\cdots 0 \\ 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{2}} & 0 & 0\cdots 0 \\ 0 & 0 & \widetilde{\mathbf{\Omega}}_{\mathbf{3}} & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \widetilde{\mathbf{\Omega}}_{\mathbf{G}} \end{bmatrix}\right)\tilde{\mathbf{x}}'$

$$\mathbf{M_q} = \begin{bmatrix} \mathbf{M_{q1}} & 0 & 0 & 0.\cdots & 0 \\ 0 & \mathbf{M_{q2}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{M_{q3}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \mathbf{M_{qG}} \end{bmatrix} (\breve{G} \times \widetilde{G} \text{ matrix}), \text{ where } \breve{G} = \sum_{g=1}^{G} (I_g - 1), \text{ and} \tag{2.83}$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & 0 & 0 & 0.\cdots & 0 \\ 0 & \mathbf{M}_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{M}_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \mathbf{M}_Q \end{bmatrix} (Q\breve{G} \times Q\widetilde{G} \text{ matrix}). \tag{2.84}$$

The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\widetilde{\boldsymbol{\Omega}}}_1, \overline{\widetilde{\boldsymbol{\Omega}}}_2, \boldsymbol{...}, \overline{\widetilde{\boldsymbol{\Omega}}}_G \overline{\boldsymbol{\Lambda}}', \overline{\widetilde{\boldsymbol{\delta}}})'$. Using the usual notations, the likelihood function is:

$$L(\boldsymbol{\theta}) = \Phi_{Q\breve{G}}(\boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}(-\boldsymbol{B}), \boldsymbol{\Xi}^*), \text{ where } \boldsymbol{B} = \mathbf{M}\mathbf{V},\ \boldsymbol{\Xi} = \mathbf{M}\widetilde{\boldsymbol{\Xi}}\mathbf{M}', \text{ and } \boldsymbol{\Xi}^* = \boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Xi}\boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1} \tag{2.85}$$

The likelihood function is of a very high dimensionality. Instead, consider the (pairwise) composite marginal likelihood function. Further, as in Section 2.1.2.2, we can reduce the pairings by testing different distance bands and determining the "optimal" distance for including pairings across individuals based on minimizing the trace of the asymptotic covariance matrix. Define a set $\widetilde{C}_q$ that includes the set of individuals *q'* (including *q*) that are within a specified threshold distance of individual *q*. Then, the CML function reduces to the following expression:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{\substack{q'=q \\ q' \in \widetilde{C}_q}}^{Q} \prod_{g=1}^{G} \prod_{g'=g}^{G'} L_{qq'gg'} \right) \text{ with } q' \neq q \text{ when } g = g', \text{ where} \tag{2.86}$$

$$L_{qq'gg'} = \Pr(d_{qg} = m_{qg}, d_{q'g'} = m_{q'g'}) = \Phi_{\breve{J}_{gg'}}(\boldsymbol{\omega}_{\breve{\Xi}_{qq'gg'}}^{-1}(\breve{\boldsymbol{B}}_{qq'gg'}), \breve{\boldsymbol{\Xi}}_{qq'gg'}^*),$$

and $\breve{J}_{gg'} = I_g + I_{g'} - 2$, $\breve{\boldsymbol{B}}_{qq'gg'} = \boldsymbol{\Delta}_{qq'gg'}\boldsymbol{B}$, $\breve{\boldsymbol{\Xi}}_{\mathbf{qq'gg'}} = \boldsymbol{\Delta}_{qq'gg'}\boldsymbol{\Xi}_{\boldsymbol{qq'gg'}}\boldsymbol{\Delta}'_{qq'gg'}$, $\breve{\boldsymbol{\Xi}}^*_{\boldsymbol{qq'gg'}} = \boldsymbol{\omega}^{-1}_{\breve{\Xi}_{qq'gg'}}\breve{\boldsymbol{\Xi}}_{\boldsymbol{qq'gg'}}\boldsymbol{\omega}^{-1}_{\breve{\Xi}_{qq'gg'}}$, and $\boldsymbol{\Delta}_{qq'gg'}$ is a $\breve{J} \times Q\breve{G}$-selection matrix with an identity matrix of size $(I_g - 1)$ occupying the first $(I_g - 1)$ rows and the $\left[(q-1)\times\breve{G} + \sum_{l=1}^{g-1} I_l + 1\right]^{th}$ through $\left[(q-1)\times\breve{G} + \sum_{l=1}^{g} I_l\right]^{th}$ columns, and another identity matrix of size $(I_{g'} - 1)$ occupying the last $(I_{g'} - 1)$ rows and the

$\left[(q'-1)\times\breve{G}+\sum_{l=1}^{g'-1}I_l+1\right]^{th}$ through $\left[(q'-1)\times\breve{G}+\sum_{l=1}^{g'}I_l\right]^{th}$ columns (with the convention that $\sum_{l=1}^{0}I_l=0$). The model can now be estimated using the MACML method. The computation of the covariance matrix is identical to the case in Section 2.2.2.2, with the use of $L_{qq'gg'}$ as in Equation (2.86) above. Once again, the analyst can consider further cutting down the number of pairings by ignoring the pairings among different individuals (observation units) across the *G* variables.

References for the CML Estimation of the Spatial MNP (SMNP) Model
No known applications thus far.

2.2.2.3 The Spatial Panel MNP Model
Consider the following model with '*t*' now being an index for choice occasion:

$$U_{qit}=\delta\sum_{q'}w_{qq'}\,U_{q'ti}+\boldsymbol{\beta}_{\boldsymbol{q}}'\boldsymbol{x}_{\boldsymbol{qti}}+\xi_{qti},\ \boldsymbol{\beta}_{\boldsymbol{q}}\sim MVN_L(\boldsymbol{b},\boldsymbol{\Omega}),\ q=1,2,...,Q,\ \ t=1,2,...,T,\ \ i=1,2,...,I. \tag{2.87}$$

We assume that ξ_{qit} is independent and identically normally distributed across *individuals and choice occasions*, but allow a general covariance structure across alternatives for each choice instance of each individual. Specifically, let $\boldsymbol{\xi}_{\boldsymbol{qt}}=(\xi_{qt1},\xi_{qt2},...\xi_{qtI})'$ ($I\times1$ vector). Then, we assume $\boldsymbol{\xi}_{\boldsymbol{qt}}\sim MVN_I(0,\boldsymbol{\Lambda})$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Lambda}$ for identifiability. Next, define the following vectors and matrices: $\mathbf{U}_{qt}=(U_{qt1},U_{qt2},...,U_{qtI})'$ ($I\times1$ vector), $\mathbf{U}_q=(\mathbf{U}_{q1},\mathbf{U}_{q2},...,\mathbf{U}_{qI})'$ ($TI\times1$ vector), $\boldsymbol{\xi}_{\boldsymbol{q}}=(\boldsymbol{\xi}_{\boldsymbol{q1}}',\boldsymbol{\xi}_{\boldsymbol{q2}}',...\boldsymbol{\xi}_{\boldsymbol{qT}}')'$ ($TI\times1$ vector), $\boldsymbol{x}_{qt}=(\boldsymbol{x}_{qt1},\boldsymbol{x}_{qt2},...,\boldsymbol{x}_{qtI})'$ ($I\times L$ matrix), $\boldsymbol{x}_q=(\boldsymbol{x}_{q1}',\boldsymbol{x}_{q2}'...,\boldsymbol{x}_{qT}')'$ ($TI\times L$ matrix),
$\mathbf{U}=(\mathbf{U}_1',\mathbf{U}_2',...\mathbf{U}_Q')'$, $\boldsymbol{\xi}=(\boldsymbol{\xi}_{\boldsymbol{1}}',\boldsymbol{\xi}_{\boldsymbol{2}}',...\boldsymbol{\xi}_{\boldsymbol{Q}}')'$ ($QTI\times1$ vectors), and $\mathbf{x}=(\mathbf{x}_1',\mathbf{x}_2',...,\mathbf{x}_Q')'$ ($QTI\times L$ matrix). Let $\boldsymbol{\beta}_{\boldsymbol{q}}=\boldsymbol{b}+\widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}}$, $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}}\sim MVN_L(\mathbf{0},\boldsymbol{\Omega})$, $\widetilde{\boldsymbol{\beta}}=\left(\widetilde{\boldsymbol{\beta}}_1',\widetilde{\boldsymbol{\beta}}_2',...,\widetilde{\boldsymbol{\beta}}_Q'\right)'$, and

$$\widetilde{\mathbf{x}}=\begin{bmatrix}\boldsymbol{x}_1 & 0 & 0 & 0.\cdots 0\\ 0 & \boldsymbol{x}_2 & 0 & 0\cdots 0\\ 0 & 0 & \boldsymbol{x}_3 & 0\cdots 0\\ \vdots & \vdots & \vdots & \vdots\ \ldots\ \vdots\\ 0 & 0 & 0 & 0\ldots\boldsymbol{x}_Q\end{bmatrix}\ (QTI\times QK\ \text{matrix}), \tag{2.88}$$

Then, we can write Equation (2.87) in matrix notation as:

$$\mathbf{U}=\mathbf{S}\left[\mathbf{x}\boldsymbol{b}+\widetilde{\mathbf{x}}\widetilde{\boldsymbol{\beta}}+\boldsymbol{\xi}\right], \tag{2.89}$$

with $\mathbf{S} = \left[\mathbf{IDEN}_{QTI} - \left\{(\delta\mathbf{W} \otimes \mathbf{IDEN}_T) \otimes \mathbf{IDEN}_I\right\}\right]^{-1}$ ($QTI \times QTI$ matrix).

Then, $\mathbf{U} \sim MVN_{QTI}(\mathbf{V}, \widetilde{\Xi})$, where $\mathbf{V} = \mathbf{Sx}\boldsymbol{b}$ and $\widetilde{\Xi} = \mathbf{S}\left[\mathbf{IDEN}_{\mathbf{Q}} \otimes (\mathbf{\Omega} + \mathbf{\Lambda})\right]\mathbf{S}'$. To develop the likelihood function, define **M** as an $[QT(I-1)] \times [QTI]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the t^{th} observation time period on individual q. This $(I-1) \times I$ matrix for parcel q and observation time period t corresponds to an $(I-1)$ identity matrix with an extra column of "-1" values added as the $m_{qt}{}^{\text{th}}$ column. For instance, consider the case of Q = 2, T = 2, and I = 4. Let individual 1 be observed to choose alternative 2 in time period 1 and alternative 1 in time period 2, and let individual choose alternative 3 in time period 1 and in alternative 4 in time period 2. Then **M** takes the form below.

$$\mathbf{M} = \left[\begin{array}{cccc|cccc|cccc|cccc}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{array}\right] \tag{2.90}$$

Let $\boldsymbol{B} = \mathbf{MV}$ and $\Xi = \mathbf{M}\widetilde{\Xi}\mathbf{M}'$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\mathbf{\Omega}}', \overline{\mathbf{\Lambda}}', \delta')'$, and the likelihood function is:

$$L(\boldsymbol{\theta}) = \Phi_{QT(I-1)}(\boldsymbol{\omega}_{\Xi}^{-1}(-\mathbf{B}\ ), \Xi^{*}), \tag{2.91}$$

where $\Xi^{*} = \boldsymbol{\omega}_{\Xi}^{-1}\Xi\boldsymbol{\omega}_{\Xi}^{-1}$.

Now, consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the T choice occasions) of the joint pairwise probability of the chosen alternatives m_{qt} for the t^{th} choice occasion and m_{qg} for the g^{th} choice occasion for individual q. To reduce the number of pairings, define a set $\widetilde{C}_q$ as in the previous section that includes the set of individuals q' (including q) that are within a specified threshold distance of individual q. Then, the CML function reduces to the following expression:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{q=1}^{Q} \prod_{\substack{q'=q \\ q' \in \tilde{C}_q}}^{Q} \prod_{t=1}^{T} \prod_{t'=i}^{T} L_{qq'tt'} \right) \text{ with } q' \neq q \text{ when } t = t', \text{ where} \tag{2.92}$$

$$L_{qq'tt'} = \Pr(d_{qt} = m_{qt}, d_{q't'} = m_{q't'}) = \Phi_{2(I-1)}(\boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qq'tt'}}(-\ddot{\boldsymbol{B}}_{qq'tt'}), \ddot{\boldsymbol{\Xi}}^{*}_{qq'tt'}),$$

where $\ddot{\boldsymbol{\Xi}}^{*}_{\mathbf{qq'tt'}} = \boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qq'tt'}} \ddot{\boldsymbol{\Xi}}_{\mathbf{qq'tt'}} \boldsymbol{\omega}^{-1}_{\tilde{\Xi}_{qq'tt'}}$, $\ddot{\boldsymbol{B}}_{qq'tt'} = \boldsymbol{\Delta}_{qq'tt'} \boldsymbol{B}$, $\ddot{\boldsymbol{\Xi}}_{\mathbf{qq'tt'}} = \boldsymbol{\Delta}_{qq'tt'} \boldsymbol{\Xi}_{\mathbf{qq'tt'}} \boldsymbol{\Delta}'_{qq'tt'}$, and $\boldsymbol{\Delta}_{qq'tt'}$ is a $2(I-1) \times QT(I-1)$-selection matrix with an identity matrix of size $(I-1)$ occupying the first $(I-1)$ rows and the $[(q-1)\times(I-1)\times T + (t-1)\times(I-1)+1]^{th}$ through $[(q-1)\times(I-1)\times T + t\times(I-1)]^{th}$ columns, and another identity matrix of size $(I-1)$ occupying the last $(I-1)$ rows and the $[(q'-1)\times(I-1)\times T + (t'-1)\times(I-1)+1]^{th}$ through $[(q'-1)\times(I-1)\times T + t'\times(I-1)]^{th}$ columns. The model can now be estimated using the MACML method. The computation of the covariance matrix is identical to the case in Section 2.2.2.2 with the use of $L_{qq'tt'}$ as in Equation (2.92) above. The analyst can consider further cutting down the number of pairings by ignoring the pairings among different individuals (observation units) across the *T* time periods.

References for the CML Estimation of the Spatial MNP (SMNP) Model

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

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

Sidharthan, R., Bhat, C.R., 2012. Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis* 44(4), 321-349.

2.4. Application to Count Models

Count data models are used in several disciplines to analyze discrete and non-negative outcomes without an explicit upper limit. Applications of such count data models abound in the scholarly literature, both in number (a count in and of itself!) as well as diversity of topics. Applications include the analysis of (a) the number of doctor visits, the number of homes affected by cholera, the number of cancer incidents, and the number of milk formula bottles supplied to infants by breastfeeding mothers in the *medicine field*, (b) the number of crimes and the number of drug possession convictions in the *criminology field*, (c) the number of mergers and acquisitions of foreign direct investments, the number of faults in a bolt, the frequency of contract change orders, and the number of jobs by space unit in the *economics field*, (d) the number of harbor seals hauled out on glacial ice and the count of birds at sanctuaries in the *ecology field*, and (e)

roadway crash frequency, counts of flights from airports, and the number of drinking under intoxication (DUI) infractions in the *transportation field*.

Count data models assume a discrete probability distribution for the count variables, followed by the parameterization of the mean of the discrete distribution as a function of explanatory variables. The two most commonly used discrete probability distributions are the Poisson and the negative binomial (NB) distributions, though other distributions such as the binomial and logarithmic distributions have also been occasionally considered. Several modifications and generalizations of the Poisson and negative binomial distributions have also been used. For example, in many count data contexts, there are a large number of zero count values. The most commonly used approach to accommodate this issue is the zero-inflated approach. The approach identifies two separate states for the count generating process – one that corresponds to a "zero" state in which the expected value of counts is so close to zero as being indistinguishable from zero, and another "normal" state in which a typical count model (with either a Poisson or NB distribution) operates. Effectively, the zero-inflated approach is a discrete-mixture model involving a discrete error distribution that modifies the probability of the zero outcome. Another similar approach to account for excess zeros is the hurdle-count approach (in which a binary outcome process of the count being below or above a hurdle (zero) is combined with a truncated discrete distribution for the count process being above the hurdle (zero) point. While the modifications and generalizations such as those just described have been effective for use with univariate count models, they are difficult to infeasible to implement in the case when there are inter-related multivariate counts at play (see Castro, Paleti and Bhat, 2012 (or CPB hereafter) and Herriges *et al*., 2008 for discussions). Also, including spatial dependence within the framework of traditional count formulations is very cumbersome. To address these situations, we can re-formulate the traditional count models as a special case of a generalized ordered-response probit (GORP) formulation (see CPB). Indeed, in this re-formulation, any count model can be formulated as a special case of a GORP formulation. Once this is achieved, all the GORP-related formulations in the earlier sections immediately carry over to count models. In this section, we will consider a single count variable based on a negative binomial distribution and show its aspatial GORP formulation, because extension to include multivariate and spatial contexts exactly mirror the previous GORP discussions.

Consider the recasting of the count model using a specific functional form for the random-coefficients generalized ordered-response probit (GORP) structure of Section 2.2.1.1 as follows:

$$y_q^* = \boldsymbol{\beta}_q' \boldsymbol{x}_q + \varepsilon_q, y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{2.93}$$

where $\boldsymbol{x}_q$ is an ($L\times1$) vector of exogenous variables (not including a constant), $\boldsymbol{\beta}_q$ is a corresponding ($L\times1$) vector of individual-specific coefficients to be estimated, ε_q is an idiosyncratic random error term that we will assume in the presentation below is independent of the elements of the vectors $\boldsymbol{\beta}_q$ and $\boldsymbol{x}_q$, and $\psi_{q,k}$ is the individual-specific upper bound threshold for discrete level k. The ε_q terms are assumed independent and identically standard normally

distributed across individuals. The typical assumption for ε_q is that it is either normally or logistically distributed, though non-parametric or mixtures-of-normal distributions may also be considered. Also, $\boldsymbol{\beta}_q = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_q$, where $\widetilde{\boldsymbol{\beta}}_q \sim MVN_L(0, \boldsymbol{\Omega})$. y_q^* is an underlying latent continuous variable that maps into the observed count variable y_q l_q through the $\boldsymbol{\psi}_q$ vector (which is a vertically stacked column vector of thresholds $(\psi_{q,-1}, \psi_{q0}, \psi_{q1}, \psi_{q2}, ..., \infty)'$. The $\psi_{q,k}$ thresholds are parameterized as a function of a vector of observable covariates $\boldsymbol{z_q}$ (including a constant) as follows (see Bhat *et al.*, 2014b):

$$\psi_{q,k} = \Phi^{-1}\left[\frac{(1-c_q)^{\theta}}{\Gamma(\theta)}\sum_{r=0}^{k}\left(\frac{\Gamma(\theta+r)}{r!}c_q^r\right)\right] + \varphi_k,\ c_q = \frac{\lambda_q}{\lambda_q+\theta}, \text{ and } \lambda_q = e^{\boldsymbol{\gamma}'\boldsymbol{z_q}}. \tag{2.94}$$

In the above equation, $\Phi^{-1}[.]$ is the inverse function of the univariate cumulative standard normal. θ is a parameter that provides flexibility to the count formulation, and, as we will see later, serves the same purpose as the dispersion parameter in a traditional negative binomial model ($\theta > 0$). $\Gamma(\theta)$ is the traditional gamma function; $\Gamma(\theta) = \int_{h=0}^{\infty} h^{\theta-1} e^{-h}\, dh$. The threshold terms in the $\boldsymbol{\psi}_q$ vector satisfy the ordering condition (*i.e.*, $\psi_{q,-1} < \psi_{q,0} < \psi_{q,1} < \psi_{q,2} .... < \infty\ \forall\ q$) as long as $\varphi_{-1} < \varphi_0 < \varphi_1 < \varphi_2 .... < \infty$. The presence of these φ terms provides substantial flexibility to accommodate high or low probability masses for specific count outcomes, beyond what can be offered by traditional treatments using zero-inflated or related mechanisms. For identification, we set $\varphi_{-1} = -\infty, \psi_{q,-1} = -\infty\ \forall q$, and $\varphi_0 = 0$. In addition, we identify a count value e^* ($e^* \in \{0,1,2,.....\}$) above which φ_e ($e \in \{0,1,2,.....\}$) is held fixed at φ_{e^*}; that is, $\varphi_e = \varphi_{e^*}$ if $e > e^*$, where the value of e^* can be based on empirical testing. For later use, let $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots \varphi_{e^*})'$ ($e^* \times 1$ vector).

The specification of the GORP model in the equation above provides a very flexible mechanism to model count data. It subsumes the traditional count models as very specific and restrictive cases. In particular, if the vector $\boldsymbol{\beta_q}$ is degenerate with all its elements taking the fixed value of zero, and all elements of the $\boldsymbol{\varphi}$ vector are zero, the model in Equation (2.93) collapses to a traditional negative binomial model with dispersion parameter θ. To see this, note that the probability expression in the GORP model of Equation (2.93) with the restrictions may be written as:

$$P[y_q = k] = P\left[\Phi^{-1}\left[\frac{(1-c_q)^{\theta}}{\Gamma(\theta)} \sum_{r=0}^{k-1} \left(\frac{\Gamma(\theta+r)}{r!} c_q^r \right) \right] < y_q^* < \Phi^{-1}\left[\frac{(1-c_q)^{\theta}}{\Gamma(\theta)} \sum_{r=0}^{k} \left(\frac{\Gamma(\theta+r)}{r!} c_q^r \right) \right] \right]$$
$$= \Phi\left(\Phi^{-1}\left[\frac{(1-c_q)^{\theta}}{\Gamma(\theta)} \sum_{r=0}^{k} \left(\frac{\Gamma(\theta+r)}{r!} c_q^r \right) \right] \right) - \Phi\left(\Phi^{-1}\left[\frac{(1-c_q)^{\theta}}{\Gamma(\theta)} \sum_{r=0}^{k} \left(\frac{\Gamma(\theta+r)}{r!} c_q^r \right) \right] \right) \qquad (2.95)$$
$$= \frac{(1-c_q)^{\theta}}{\Gamma(\theta)} \left(\frac{\Gamma(\theta+k)}{k!} c_q^k \right),$$

which is the probability expression of the negative binomial count model. If, in addition, $\theta \to \infty$, the result can be shown to be the Poisson count model.

In an empirical context of crash counts at intersections, CPB interpret the GORP recasting of the count model as follows. There is a latent "long-term" crash propensity y_q^* associated with intersection *q* that is a linear function of a set of intersection-related attributes $\boldsymbol{x_q}$ On the other hand, there may be some specific intersection characteristics (embedded in $\mathbf{z}_q$ within the threshold terms) that may dictate the likelihood of a crash occurring at any given *instant of time* for a given long-term crash propensity y_q^*. Thus, two intersections may have the same latent long-term crash propensity y_q^*, but may show quite different observed number of crashes over a certain time period because of different y_q^*- to -y_q mappings through the cut points (y_q is the observed count variable). CPB postulated that factors such as intersection traffic volumes, traffic control type and signal coordination, driveways between intersections, and roadway alignment are likely to affect "long-term" latent crash propensity at intersections and perhaps also the thresholds. On the other hand, they postulate that there may be some specific intersection characteristics such as approach roadway types and curb radii at the intersection that will load more on the thresholds that affect the translation of the crash propensity to crash outcomes. Of course, one can develop similar interpretations of the latent propensity and thresholds in other count contexts (see, for example, the interpretation provided by Bhat *et al*., 2014a, in a count context characterized by the birth of new firms in Texas counties).

To summarize, the GORP framework represents a generalization of the traditional count data model, has the ability to retain all the desirable traits of count models and relax constraints imposed by count models, leads to a much simpler modeling structure when flexible spatial and temporal dependencies are to be accommodated, and may also be justified from an intuitive/conceptual standpoint. Indeed, all the spatial, multivariate, and panel-based extensions discussed under ordered-response models immediately apply to count models based on the count reformulation as a GORP model.

References for the CML Estimation of Count Models

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B* 46(1), 253-272.

Bhat, C.R., Paleti, R., Singh, P., 2014a. A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, forthcoming.

Bhat, C.R., Born, K., Sidharthan, R., Bhat, P.C., 2014b. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. *Analytic Methods in Accident Research* 1, 53-71.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation Research Part B* 55, 245-264.

3. APPLICATION TO JOINT MIXED MODEL SYSTEMS

The joint modeling of data of mixed types of dependent variables (including ordered-response or ordinal variables, unordered-response or nominal variables, count variables, and continuous variables) is of interest in several fields, including biology, economics, epidemiology, social science, and transportation (see a good synthesis of applications in de Leon and Chough, 2013). For instance, in the transportation field, it is likely that households that are not auto-oriented choose to locate in transit and pedestrian friendly neighborhoods that are characterized by mixed and high land use density, and then the good transit service may also further structurally influence mode choice behaviors. If that is the case, then it is likely that the choices of residential location, vehicle ownership, and commute mode choice are being made jointly as a bundle. That is, residential location may structurally affect vehicle ownership and commute mode choice, but underlying propensities for vehicle ownership and commute mode may themselves affect residential location in the first place to create a bundled choice. This is distinct from a sequential decision process in which residential location choice is chosen first (with no effects whatsoever of underlying propensities for vehicle ownership and commute mode on residential choice), then residential location affects vehicle ownership (which is chosen second, and in which the underlying propensity for commute mode does not matter), and finally vehicle ownership affects commute mode choice (which is chosen third). The sequential model is likely to over-estimate the impacts of residential location (land use) attributes on activity-travel behavior because it ignores self-selection effects wherein people who locate themselves in mixed and high land use density neighborhoods were auto-disoriented to begin with. These lifestyle preferences and attitudes constitute unobserved factors that simultaneously impact long term location choices, medium term vehicle ownership choices, and short term activity-travel choices; the way to accurately reflect their impacts and capture the "bundling" of choices is to model the choice dimensions together in a joint equations modeling framework that accounts for correlated unobserved lifestyle (and other) effects as well as possible structural effects.

There are many approaches to model joint mixed systems (see Wu *et al.*, 2013 for a review), but the one we will focus on here is based on accommodating jointness through the specification of a distribution for the unobserved components of the latent continuous variables underlying the discrete (ordinal, nominal, or count) variables and the unobserved components of observed continuous variables. Very generally speaking, one can consider a specific marginal distribution for each of the unobserved components of the latent continuous variables (underlying the discrete variables) and the observed continuous variable, and then generate a joint system through a copula-based correlation on these continuous variables. However, here we will assume that the marginal distributions of the latent and observed continuous variables are all normally distributed, and assume a Gaussian Copula to stitch the error components together. This is equivalent to assuming a multivariate normal distribution on the error components. But the procedures can be extended to non-normal marginal distributions and non-Gaussian copulas in a relatively straightforward fashion.

From a methodological perspective, the simulation-based likelihood estimation of joint mixed models can become quite cumbersome and time-consuming. However, the use of the MACML estimation technique has once again opened up possibilities because of the dramatic breakthrough in the ease and computational feasibility of estimating joint mixed systems.

3.1. Joint Mixed Dependent Variable Model Formulation

In the following presentation, for ease in exposition, we assume fixed coefficients on variables, though extension to the case of random coefficients in conceptually very straightforward (as in earlier sections). We will also suppress the notation for individuals, and assume that all error terms are independent and identically distributed across individuals. Finally, we will develop the formulation in the context of ordinal, nominal, and continuous variables, though the formulation is immediately applicable to count variables too because count variables may be modeled as a specific case of the GORP-based formulation for ordinal variables.

Let there be N ordinal variables for an individual, and let n be the index for the ordinal variables $(n = 1, 2, ..., N)$. Also, let J_n be the number of outcome categories for the n^{th} ordinal variable $(J_n \geq 2)$ and let the corresponding index be j_n $(j_n = 1, 2, ..., J_n)$. Let y_n^* be the latent underlying variable whose horizontal partitioning leads to the observed choices for the n^{th} ordinal variable. Assume that the individual under consideration chooses the a_n^{th} ordinal category. Then, in the usual ordered response formulation:

$$y_n^* = \boldsymbol{\delta}_n' \boldsymbol{w} + \varepsilon_n, j_n = k \text{ if } \psi_{k-1}^n < y_{ql}^* < \psi_k^n, \tag{3.1}$$

where $\boldsymbol{w}$ is a fixed and constant vector of exogenous variables (not including a constant), $\boldsymbol{\delta}_n$ is a corresponding vector of coefficients to be estimated, the ψ terms represent thresholds, and ε_n is the standard normal random error for the n^{th} ordinal variable. We parameterize the thresholds as:

$$\psi_k^n = \psi_{k-1}^n + \exp(\alpha_{kn} + \boldsymbol{\gamma}_{kn}' \boldsymbol{z}) \tag{3.2}$$

In the above equation, α_{kn} is a scalar, and $\boldsymbol{\gamma}_{kn}$ is a vector of coefficients associated with ordinal level $k = 1, 2, ..., K-1$ for the n^{th} ordinal variable. The above parameterization immediately guarantees the ordering condition on the thresholds for each and every crash, while also enabling the identification of parameters on variables that are common to the $\boldsymbol{w}$ and $\boldsymbol{z}$ vectors. For identification reasons, we adopt the normalization that $\psi_1^n = \exp(\alpha_{1n})$ $\forall$ n. Stack the N latent variables y_n^* into an $(N \times 1)$ vector $\boldsymbol{y}^*$, and let $\mathbf{y}^* \sim N\left(\mathbf{f}, \boldsymbol{\Xi}_{\mathbf{y}^*}\right)$, where $\mathbf{f} = \left(= (\boldsymbol{\delta}_1' \boldsymbol{w}, \boldsymbol{\delta}_2' \boldsymbol{w}, ..., \boldsymbol{\delta}_N' \boldsymbol{w}\right)$ and $\boldsymbol{\Sigma}_{y^*}$ is the covariance matrix of $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)$. Also, stack the lower thresholds corresponding to the actual observed outcomes for the n ordinal variables into an $(N \times 1)$ vector $\boldsymbol{\psi}^{\mathbf{low}}$ and the upper thresholds into another vector $\boldsymbol{\psi}^{\mathbf{up}}$. For later use, define

$\boldsymbol{\gamma}_{\boldsymbol{n}} = (\boldsymbol{\gamma}'_{\boldsymbol{2n}}, \boldsymbol{\gamma}'_{\boldsymbol{3n}}, ..., \boldsymbol{\gamma}'_{\boldsymbol{J_n\text{-}1,n}})', \boldsymbol{\gamma} = (\boldsymbol{\gamma}'_{\boldsymbol{1}}, \boldsymbol{\gamma}'_{\boldsymbol{2}}, ..., \boldsymbol{\gamma}'_{\boldsymbol{N}})'$, $\boldsymbol{\alpha}_{\boldsymbol{n}} = (\alpha_{1n}, \alpha_{2n}, ..., \alpha_{J_n-1,n})'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_{\boldsymbol{1}}, \boldsymbol{\alpha}'_{\boldsymbol{2}}, ..., \boldsymbol{\alpha}'_{\boldsymbol{N}})'$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, ..., \boldsymbol{\delta}'_N)'$.

Let there be *G* nominal (unordered-response) variables for an individual, and let *g* be the index for the nominal variables (*g* = 1, 2, 3,…, *G*). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 3$) and let i_g be the corresponding index (i_g = 1, 2, 3,…, I_g). Consider the g^{th} nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual random utility structure for each alternative i_g.

$$U_{gi_g} = \boldsymbol{b}'_{\boldsymbol{g}} \boldsymbol{x}_{gi_g} + \xi_{gi_g}, \tag{3.3}$$

where $\boldsymbol{x}_{gi_g}$ is a $L \times 1-$column vector of exogenous attributes, $\boldsymbol{b}_g$ is a column vector of corresponding coefficients, and ξ_{gi_g} is a normal error term. Let $\boldsymbol{\xi}_{\boldsymbol{g}} = (\xi_{g1}, \xi_{g2}, ...\xi_{gI_g})'$ ($I_g \times 1$ vector), $\boldsymbol{\xi}_{\boldsymbol{g}} \sim MVN_{I_g}(0, \boldsymbol{\Lambda}_{\mathbf{g}})$. Let $\mathbf{U}_g = (U_{g1}, U_{g2}, ..., U_{gI_g})'$ ($I_g \times 1$ vector), $\boldsymbol{x}_g = (\boldsymbol{x}_{\boldsymbol{g1}}, \boldsymbol{x}_{\boldsymbol{g2}}, \boldsymbol{x}_{\boldsymbol{g3}}, ..., \boldsymbol{x}_{\boldsymbol{gI_q}})'$ ($I_g \times L$ matrix), $\mathbf{V}_g = \boldsymbol{x}_{\boldsymbol{g}} \boldsymbol{b}_{\mathbf{g}}$ ($I_g \times 1$ vector). Then $\mathbf{U}_g \sim MVN_{I_g}(\mathbf{V}_g, \boldsymbol{\Lambda}_{\mathbf{g}})$. Under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$, since the individual chose alternative m_g. Let $u^*_{gi_gm_g} = U_{gi_g} - U_{gm_g} (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{u}^*_{\boldsymbol{g}} = \left[\left(u^*_{g1m_g}, u^*_{g2m_g}, ..., u^*_{gI_gm_g} \right)'; i_g \neq m_g \right]$. As usual, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}_g$ of $\boldsymbol{\varsigma}_{\boldsymbol{g}} = (\varsigma_{g2}, \varsigma_{g3}, ..., \varsigma_{gI_g})'$, where $\varsigma_{gi} = \xi_{gi} - \xi_{g1}$ ($i \neq 1$), are estimable. However, the condition that $\boldsymbol{u}^*_{\boldsymbol{g}} < \boldsymbol{0}_{I_g - 1}$ takes the difference against the alternative m_g that is chosen for the nominal variable *g*. Thus, during estimation, the covariance matrix $\ddot{\boldsymbol{\Lambda}}_{\mathbf{g}}$ (of the error differences taken with respect to alternative m_g is desired). Since m_g will vary across households, $\ddot{\boldsymbol{\Lambda}}_{\mathbf{g}}$ will also vary across households. But all the $\ddot{\boldsymbol{\Lambda}}_{\mathbf{g}}$ matrices must originate in the same covariance matrix $\boldsymbol{\Lambda}_g$ for the original error term vector $\boldsymbol{\xi}_{\boldsymbol{g}}$. To achieve this consistency, $\boldsymbol{\Lambda}_g$ is constructed from $\breve{\boldsymbol{\Lambda}}_g$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. Also, an additional scale normalization needs to be imposed on $\breve{\boldsymbol{\Lambda}}_g$. For this, we normalize the first element of $\breve{\boldsymbol{\Lambda}}_g$ to the value of one. The discussion above focuses on a single nominal variable *g*. When there are *G* nominal variables, define $\ddot{G} = \sum_{g=1}^{G} I_g$ and $\tilde{G} = \sum_{g=1}^{G} (I_g - 1)$. Further, let $\breve{\boldsymbol{u}}^*_{\boldsymbol{g}} = \left(U_{g2} - U_{g1}, U_{g3} - U_{g1}, ..., U_{gI_g} - U_{g1} \right)'$, $\breve{\boldsymbol{u}}^* = \left(\left[\breve{\boldsymbol{u}}^*_{\boldsymbol{1}} \right]', \left[\breve{\boldsymbol{u}}^*_{\boldsymbol{2}} \right]', ..., \left[\breve{\boldsymbol{u}}^*_{\boldsymbol{G}} \right]' \right)'$, and $\boldsymbol{u}^* = \left(\left[\boldsymbol{u}^*_{\boldsymbol{1}} \right]', \left[\boldsymbol{u}^*_{\boldsymbol{2}} \right]', ..., \left[\boldsymbol{u}^*_{\boldsymbol{G}} \right]' \right)'$ (so $\breve{\boldsymbol{u}}^*$ is the vector of utility

differences taken with respect to the first alternative for each nominal variable, while $\boldsymbol{u}^*$ is the vector of utility differences taken with respect to the chosen alternative for each nominal variable). Now, construct a matrix of dimension $\widetilde{G}\times\widetilde{G}$ that represents the covariance matrix of $\breve{\boldsymbol{u}}^*$:

$$\boldsymbol{\Sigma}_{\breve{u}^*}=\begin{bmatrix} \breve{\boldsymbol{\Lambda}}_{\mathbf{1}} & \breve{\boldsymbol{\Lambda}}_{\mathbf{12}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{1G}} \\ \breve{\boldsymbol{\Lambda}}'_{\mathbf{12}} & \breve{\boldsymbol{\Lambda}}_{\mathbf{2}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{2G}} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \breve{\boldsymbol{\Lambda}}'_{\mathbf{1G}} & \breve{\boldsymbol{\Lambda}}'_{\mathbf{2G}} & . & . & . & \breve{\boldsymbol{\Lambda}}_{\mathbf{G}} \end{bmatrix} \tag{3.4}$$

In the general case, this allows the estimation of $\sum_{g=1}^{G}\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms across all the G nominal variables (originating from $\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms embedded in each $\breve{\boldsymbol{\Lambda}}_g$ matrix; g=1,2,…G) and the $\sum_{g=1}^{G-1}\sum_{l=g+1}^{G}(I_g-1)\times(I_l-1)$ covariance terms in the off-diagonal matrices of the $\boldsymbol{\Sigma}_{\breve{u}^*}$ matrix characterizing the dependence between the latent utility differentials (with respect to the first alternative) across the nominal variables (originating from $(I_g-1)\times(I_l-1)$ estimable covariance terms within each off-diagonal matrix in $\boldsymbol{\Sigma}_{\breve{u}^*}$). For later use, define the stacked $\breve{G}\times 1-$vectors $\boldsymbol{U}=\left(\boldsymbol{U}'_1,\boldsymbol{U}'_2,\ldots,\boldsymbol{U}'_G\right)'$, and $\boldsymbol{V}=\left(\boldsymbol{V}'_1,\boldsymbol{V}'_2,\ldots,\boldsymbol{V}'_G\right)'$.

Finally, let there be H continuous variables $(y_1,y_2,...,y_H)$ with an associated index h $(h=1,2,...,H)$. Let $y_h=\boldsymbol{\lambda}'_{\mathbf{h}}\boldsymbol{s}_h+\eta_h$ in the usual linear regression fashion, and $\boldsymbol{\lambda}=(\boldsymbol{\lambda}'_1,\boldsymbol{\lambda}'_2,...,\boldsymbol{\lambda}'_H)'$. Stacking the H continuous variables into a $(H\times 1)$-vector $\boldsymbol{y}$, one may write $\boldsymbol{y}=MVN_h(\boldsymbol{c},\boldsymbol{\Sigma}_y)$, where $\boldsymbol{c}=\left(\boldsymbol{\lambda}'_1\boldsymbol{s}_1,\boldsymbol{\lambda}'_2\boldsymbol{s}_2,...,\boldsymbol{\lambda}'_H\boldsymbol{s}_H\right)'$, and $\boldsymbol{\Sigma}_y$ is the covariance matrix of $\boldsymbol{\eta}=\left(\eta_1,\eta_2,....\eta_H\right)$.

3.2. The Joint Mixed Model System and the Likelihood Formation

The jointness across the different types of dependent variables may be specified by writing the covariance matrix of $\breve{\boldsymbol{y}}=\left(\breve{\boldsymbol{u}}^*,\boldsymbol{y}^*,\boldsymbol{y}\right)$ as:

$$\mathrm{Var}(\breve{\boldsymbol{y}})=\breve{\boldsymbol{\Omega}}=\begin{bmatrix} \boldsymbol{\Sigma}_{\breve{u}^*} & \boldsymbol{\Sigma}_{\breve{u}^*y^*} & \boldsymbol{\Sigma}_{\breve{u}^*y} \\ \boldsymbol{\Sigma}'_{\breve{u}^*y^*} & \boldsymbol{\Sigma}_{y^*} & \boldsymbol{\Sigma}_{y^*y} \\ \boldsymbol{\Sigma}'_{\breve{u}^*y} & \boldsymbol{\Sigma}'_{y^*y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}, \tag{3.5}$$

where $\boldsymbol{\Sigma}_{\breve{u}^*\mathbf{y}^*}$ is a $\widetilde{G}\times N$ matrix capturing covariance effects between the $\breve{\boldsymbol{u}}^*$ vector and the $\boldsymbol{y}^*$ vector, $\boldsymbol{\Sigma}_{\breve{u}^*\mathbf{y}}$ is a $\widetilde{G}\times H$ matrix capturing covariance effects between the $\breve{\boldsymbol{u}}^*$ vector and the $\boldsymbol{y}$ vector, and $\boldsymbol{\Sigma}_{y^*y}$ is an $N\times H$ matrix capturing covariance effects between the $\boldsymbol{y}^*$ vector and the $\boldsymbol{y}$ vector. All elements of the matrix above are identifiable. However, the matrix represents the covariance of latent utility differentials taken with respect to the first alternative for each of the nominal variables. For estimation, the corresponding matrix with respect to the latent utility differentials with respect to the chosen alternative for each nominal variable, say $\widetilde{\boldsymbol{\Omega}}$, is needed. For this purpose, first construct the general covariance matrix $\boldsymbol{\Omega}$ for the original $\left[\ddot{G}+N+H\right]\times 1$ vector $\mathbf{UY}=\left(\mathbf{U}',\boldsymbol{y}^{*'},\boldsymbol{y}'\right)'$, while also ensuring all parameters are identifiable (note that $\boldsymbol{\Omega}$ is equivalently the covariance matrix of $\boldsymbol{\tau}=(\boldsymbol{\varepsilon}',\boldsymbol{\xi}',\boldsymbol{\eta}')'$. To do so, define a matrix $\mathbf{D}$ of size $\left[\ddot{G}+N+H\right]\times\left[\widetilde{G}+N+H\right]$. The first I_1 rows and (I_1-1) columns correspond to the first nominal variable. Insert an identity matrix of size (I_1-1) after supplementing with a first row of zeros in the first through (I_1-1)th columns of the matrix. The rest of the elements in the first I_1 rows and the first (I_1-1) columns take a value of zero. Next, rows (I_1+1) through (I_1+I_2) and columns (I_1) through (I_1+I_2-2) correspond to the second nominal variable. Again position an identity matrix of size (I_2-1) after supplementing with a first row of zeros into this position. Continue this for all *G* nominal variables. Put zero values in all cells without any value up to this point. Finally, insert an identity matrix of size *N*+*H* into the last *N*+*H* rows and *N*+*H* columns of the matrix **D**. Thus, for the case with two nominal variables, one nominal variable with 3 alternatives and the second with four alternatives, one ordinal variable, and one continuous variable, the matrix **D** takes the form shown below:

$$\left[\begin{array}{cc:ccc:cc}
0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0\\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0\\
\hdashline
0 & 0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]_{9*7} \tag{3.6}$$

Then, the general covariance matrix of $\mathbf{UY}$ may be developed as $\boldsymbol{\Omega}=\mathbf{D}\breve{\boldsymbol{\Omega}}\mathbf{D}'$. All parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on utility differences and, at the same time, it provides a consistent means to obtain the covariance matrix

$\widetilde{\boldsymbol{\Omega}}$ that is needed for estimation (and is with respect to each individual's chosen alternative for each nominal variable). Specifically, to develop the distribution for the vector $\widetilde{\boldsymbol{y}} = \left(\boldsymbol{u}^*, \boldsymbol{y}^{*'}, \boldsymbol{y}'\right)'$, define a matrix **M** of size $\left[\widetilde{G} + N + H\right] \times \left[\vec{G} + N + H\right]$. The first $(I_1 - 1)$ rows and I_1 columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. The rest of the columns for the first $(I_1 - 1)$ rows and the rest of the rows for the first I_1 columns take a value of zero. Next, rows (I_1) through $(I_1 + I_2 - 2)$ and columns $(I_1 + 1)$ through $(I_1 + I_2)$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Continue this procedure for all *G* nominal variables. Finally, insert an identity matrix of size *N* +*H* into the last *N* +*H* rows and *N* +*H* columns of the matrix **M**. With the matrix **M** as defined, the covariance matrix $\widetilde{\boldsymbol{\Omega}}$ is given by $\widetilde{\boldsymbol{\Omega}} = \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'$.

Next, define $\widetilde{\boldsymbol{u}} = \left(\boldsymbol{u}^{*'}, \boldsymbol{y}^{*'}\right)'$ and $\widetilde{\boldsymbol{g}} = \left((\mathbf{M}\boldsymbol{V})', \boldsymbol{f}'\right)'$. Also, partition $\widetilde{\boldsymbol{\Omega}}$ so that

$$\widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_{u^*} & \widetilde{\boldsymbol{\Sigma}}_{u^*y^*} & \widetilde{\boldsymbol{\Sigma}}_{u^*y} \\ \widetilde{\boldsymbol{\Sigma}}'_{u^*y^*} & \boldsymbol{\Sigma}_{y^*} & \boldsymbol{\Sigma}_{y^*y} \\ \widetilde{\boldsymbol{\Sigma}}'_{u^*y} & \boldsymbol{\Sigma}'_{y^*y} & \boldsymbol{\Sigma}_{y} \end{bmatrix} \tag{3.7}$$

Let $\widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}} = \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_{u^*} & \widetilde{\boldsymbol{\Sigma}}_{u^*y^*} \\ \widetilde{\boldsymbol{\Sigma}}'_{u^*\mathbf{y}^*} & \boldsymbol{\Sigma}_{\mathbf{y}^*} \end{bmatrix}$ and $\mathrm{Var}(\widetilde{\boldsymbol{y}}) = \widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}} & \widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}y} \\ \widetilde{\boldsymbol{\Sigma}}'_{\widetilde{u}y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}$, where $\widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}y} = \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_{u^*y} \\ \boldsymbol{\Sigma}_{y^*y} \end{bmatrix}$ $(\widetilde{G} + N) \times H$ matrix. Also, supplement the threshold vectors defined earlier as follows: $\widetilde{\boldsymbol{\psi}}^{\mathbf{low}} = \left[\left(-\infty_{\widetilde{G}}\right)', \left(\boldsymbol{\psi}^{\mathbf{low}}\right)'\right]$, and $\widetilde{\boldsymbol{\psi}}^{\mathbf{up}} = \left[\left(\mathbf{0}_{\widetilde{G}}\right)', \left(\boldsymbol{\psi}^{\mathbf{up}}\right)'\right]$, where $-\infty_{\widetilde{G}}$ is a $(\widetilde{G} \times 1)$-column vector of negative infinities, and $\mathbf{0}_{\widetilde{G}}$ is another $(\widetilde{G} \times 1)$-column vector of zeros. The conditional distribution of $\widetilde{\boldsymbol{u}}$ given $\boldsymbol{y}$, is multivariate normal with mean $\widetilde{\widetilde{\boldsymbol{g}}} = \widetilde{\boldsymbol{g}} + \widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}y}\boldsymbol{\Sigma}_y^{-1}\left(\boldsymbol{y} - \boldsymbol{d}\right)$ and variance $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{\widetilde{u}} = \widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}} - \widetilde{\boldsymbol{\Sigma}}_{\widetilde{u}y}\boldsymbol{\Sigma}_y^{-1}\widetilde{\boldsymbol{\Sigma}}'_{\widetilde{u}y}$.

Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_{\boldsymbol{G}};\ \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\alpha};\ \mathrm{Vech}(\boldsymbol{\Sigma}_{\widetilde{u}});\ \boldsymbol{\lambda};\ \mathrm{Vech}(\boldsymbol{\Sigma}_y);\ \mathrm{Vech}(\boldsymbol{\Sigma}_{\widetilde{u}\mathbf{y}})]$. Then the likelihood function for the household may be written as:

$$L(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_y) \times \Pr\left[\widetilde{\boldsymbol{\psi}}^{\mathbf{low}} \le \widetilde{\boldsymbol{u}} \le \widetilde{\boldsymbol{\psi}}^{\mathbf{up}}\right], \tag{3.8}$$

$$= \phi_H(\boldsymbol{y} - \boldsymbol{d} \mid \boldsymbol{\Sigma}_y) \times \int_{D_{\widetilde{u}}} f_{\widetilde{G}+N}(\widetilde{\boldsymbol{u}} \mid \widetilde{\widetilde{\boldsymbol{g}}}, \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{\widetilde{u}}) d\widetilde{u},$$

where the integration domain $D_{\tilde{u}} = \{\tilde{\mathbf{u}} : \tilde{\boldsymbol{\psi}}^{\text{low}} \le \tilde{\mathbf{u}} \le \tilde{\boldsymbol{\psi}}^{\text{up}}\}$ is simply the multivariate region of the elements of the $\tilde{\boldsymbol{u}}$ vector determined by the range $(-\infty, 0)$ for the nominal variables and by the observed outcomes of the ordinal variables, and $f_{\tilde{G}+N}(.)$ is the multivariate normal density function of dimension $\tilde{G}+N$. The likelihood function for a sample of Q observations is obtained as the product of the observation-level likelihood functions.

The above likelihood function involves the evaluation of a $\tilde{G}+N$-dimensional rectangular integral for each household, which can be computationally expensive. So, the Maximum Approximate Composite Marginal Likelihood (MACML) approach of Bhat (2011) may be used.

3.3. The Joint Mixed Model System and the MACML Estimation Approach

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the N ordinal variables *and* G nominal variables) of the joint pairwise probability of the chosen alternatives for an individual, and computed using the analytic approximation of the multivariate normal cumulative distribution (MVNCD) function.

$$L_{MACML}(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_{\mathbf{y}}) \times \left(\prod_{n=1}^{N-1} \prod_{n'=n+1}^{N} \Pr(j_n = a_n, j_{n'} = a'_n) \right) \times \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr(d_{i_g} = m_g, d_{i_{g'}} = m_{g'}) \right) \times \left(\prod_{g=1}^{G} \prod_{n=1}^{N} \Pr(d_{i_g} = m_g, j_n = a_n) \right). \tag{3.9}$$

where d_{i_g} is an index for the individual's choice for the g^{th} nominal variable. The net result is that the pairwise likelihood function now only needs the evaluation of $\tilde{G}_{gg'}, \tilde{G}_{nn'}$, and $\tilde{G}_{gn}$ dimensional cumulative normal distribution functions (rather than the $\tilde{G}+N$-dimensional cumulative distribution function in the maximum likelihood function), where $\tilde{G}_{gg'} = I_g + I_{g'} - 2$, $\tilde{G}_{nn'} = 2$, and $\tilde{G}_{gn} = I_g$. This leads to substantial computational efficiency. However, in cases where there are several alternatives for one or more nominal variables, the dimension $\tilde{G}_{gg'}$ and $\tilde{G}_{gn}$ can still be quite high. This is where the use of an analytic approximation of the MVNCD function comes in handy. The resulting maximum approximated composite marginal likelihood (MACML) is solely based on bivariate and univariate cumulative normal computations. Also note that the probabilities in the MACML function in Equation (3.9) can be computed by selecting out the appropriate sub-matrices of the mean vector $\tilde{\tilde{\boldsymbol{g}}}$ and the covariance matrix $\tilde{\tilde{\boldsymbol{\Sigma}}}_{\tilde{u}}$ of the vector $\tilde{\boldsymbol{u}}$, and the appropriate sub-vectors of the threshold vectors $\tilde{\boldsymbol{\psi}}^{\text{low}}$ and $\tilde{\boldsymbol{\psi}}^{\text{up}}$. The covariance matrix of the parameters $\boldsymbol{\theta}$ may be estimated as:

$$\frac{\hat{\boldsymbol{G}}^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}, \tag{3.10}$$

$$\text{with } \hat{\boldsymbol{H}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{MACML}}$$

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\theta}_{MACML}} \tag{3.11}$$

An alternative estimator for $\hat{\boldsymbol{H}}$ is as below:

$$\hat{\boldsymbol{H}} = \frac{1}{Q}\sum_{q=1}^{Q}\left(\begin{array}{l}
\left[\frac{\partial \log\left[\phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_y)\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log\left[\phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_y)\right]}{\partial\boldsymbol{\theta}'}\right] + \\
\sum_{n=1}^{N-1}\sum_{n'=n+1}^{N}\left[\frac{\partial \log\left[\Pr(j_n = a_n, j_{n'} = a'_n)\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log\left[\Pr(j_n = a_n, j_{n'} = a'_n)\right]}{\partial\boldsymbol{\theta}'}\right] + \\
\sum_{g=1}^{G-1}\sum_{g'=g+1}^{G}\left[\frac{\partial \log\left[\Pr(d_{i_g} = m_g, d_{i_{g'}} = m_{g'})\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log\left[\Pr(d_{i_g} = m_g, d_{i_{g'}} = m_{g'})\right]}{\partial\boldsymbol{\theta}'}\right] + \\
\sum_{g=1}^{G}\sum_{n=1}^{N}\left[\frac{\partial \log\left[\Pr(d_{i_g} = m_g, j_n = a_n)\right]}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log\left[\Pr(d_{i_g} = m_g, j_n = a_n)\right]}{\partial\boldsymbol{\theta}'}\right]
\end{array}\right)$$

3.4. Positive Definiteness

The matrix $\tilde{\boldsymbol{\Omega}}$ for each household has to be positive definite. The simplest way to guarantee this is to ensure that the matrix $\breve{\boldsymbol{\Omega}}$ is positive definite. To do so, the Cholesky matrix of $\breve{\boldsymbol{\Omega}}$ may be used as the matrix of parameters to be estimated. However, note that the top diagonal element of each $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$ in $\breve{\boldsymbol{\Omega}}$ is normalized to one for identification, and this restriction should be recognized when using the Cholesky factor of $\breve{\boldsymbol{\Omega}}$. Further, the diagonal elements of $\boldsymbol{\Sigma}_{y^*}$ in $\breve{\boldsymbol{\Omega}}$ are also normalized to one. These restrictions can be maintained by appropriately parameterizing the diagonal elements of the Cholesky decomposition matrix. Thus, consider the lower triangular Cholesky matrix $\breve{\boldsymbol{L}}$ of the same size as $\breve{\boldsymbol{\Omega}}$. Whenever a diagonal element (say the kk^{th} element) of $\breve{\boldsymbol{\Omega}}$ is to be normalized to one, the corresponding diagonal element of $\breve{\boldsymbol{L}}$ is written as $\sqrt{1-\sum_{j=1}^{a-1} d_{kj}^2}$, where the d_{kj} elements are the Cholesky factors that are to be estimated. With this parameterization, $\breve{\boldsymbol{\Omega}}$ obtained as $\breve{\boldsymbol{L}}\breve{\boldsymbol{L}}'$ is positive definite and adheres to the scaling conditions.

References for the CML Estimation of the Mixed Variable Model

Bhat, C.R., Born, K., Sidharthan, R., Bhat, P.C., 2014b. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. *Analytic Methods in Accident Research* 1, 53-71.

Khan, M., Paleti, R., Bhat, C.R., Pendyala, R.M., 2012. Joint household-level analysis of individuals' work arrangement choices. *Transportation Research Record* 2323, 56-66.

Paleti, R., Bhat, C.R., Pendyala, R.M., 2013. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research* Record 2382, 162-172.

Paleti, R., Pendyala, R.M., Bhat, C.R., Konduri, K.C., 2011. A joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length. Technical paper, School of Sustainable Engineering and the Built Environment, Arizona State University.

Singh, P., Paleti, R., Jenkins, S., Bhat, C.R., 2013. On modeling telecommuting behavior: option, choice, and frequency. *Transportation* 40(2), 373-396.

4. CONCLUSIONS

This paper presents the basics of the composite marginal likelihood (CML) inference approach, discussing the asymptotic properties of the CML estimator and possible applications of the approach for a suite of different types of discrete and mixed dependent variable models. The approach can be applied using simple optimization software for likelihood estimation. In the case of models with complex and analytically intractable full likelihoods, the CML also represents a conceptually and pedagogically simpler simulation-free procedure relative to simulation techniques, and has the advantage of reproducibility of the results. For instance, in a panel application, Varin and Czado (2010) examine 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 computer-intensive simulation techniques. In another panel spatial application, Sidharthan and Bhat (2012) examine the case of spatial dependence in land-use of spatial grids, and the full information likelihood estimator would have entailed integration of the order of 4800 dimensions. Despite advances in simulation techniques and computational power, the evaluation of such high dimensional integrals is literally infeasible using traditional frequentist and Bayesian simulation techniques. For instance, in frequentist methods, where estimation is typically undertaken using pseudo-Monte Carlo or quasi-Monte Carlo simulation approaches (combined with a quasi-Newton optimization routine in a maximum simulated likelihood (MSL) inference), the computational cost to ensure good asymptotic estimator properties becomes prohibitive for the number of dimensions just discussed. Similar problems arise in Bayesian Markov Chain Monte Carlo (MCMC) simulation approaches, which remain cumbersome, require extensive simulation, are time consuming, and pose convergence assessment problems as the number of dimensions increases (see Ver Hoef and Jansen, 2007, and Franzese *et al*., 2010 for discussions).

Even when the full likelihood involves a lower and more practically feasible dimensionality of integration, the accuracy of simulation techniques is known to degrade rapidly as the dimensionality increases, and the simulation noise increases substantially. This leads to convergence problems during estimation, unless a very high number of simulation draws is used. Several studies have demonstrated so in a variety of econometric modeling contexts (see, for example, Bhat and Sidharthan, 2011 and Paleti and Bhat, 2013). Besides, an issue generally ignored in simulation-based approaches is the accuracy (or lack thereof) of the covariance matrix of the estimator, which is critical for good inference even if the asymptotic properties of the estimator are well established. Thus, the CML can present a very attractive alternative to the traditional MSL method in many situations.

Of course, there are some special cases where the MSL approach may be preferable to the CML approach. For example, consider a panel binary discrete choice case with J choice occasions per individual and K random coefficients on variables. Let the kernel error term be normally distributed and assume that the random coefficients are multivariate normally distributed, so that the overall error is also normally distributed. Here, when $K < J$, and $K \leq 3$, the

MSL estimation with the full likelihood function is likely to be preferable to the CML. This is because integrating up to three dimensions is quite fast and accurate using quasi-Monte Carlo simulation techniques. This is particularly so when J is also large, because the number of pairings in the CML is high. For the case when $K < J$ and $K > 3$, or $K \geq J > 3$, the CML is likely to become attractive, because of the MSL-related problems mentioned earlier for moderate dimensions of integration. For example, when $K = J = 5$, the CML is fast since it entails the evaluation of only 10 probability pairings for each individual (each pairing involving bivariate normal cumulative distribution function evaluations) rather than a five-dimensional integration for each individual in the MSL estimation. Note that one may be tempted to think that the CML loses this edge when J becomes large. For instance, when $J = 10$, there would be 45 probability pairings for each individual in a pairwise likelihood approach. But the surrogate likelihood function in the CML estimation can be formulated in many different ways rather than the full pairings approach presented here. Thus, one could consider only the pairing combinations of the first five (or five randomly selected) choice occasions for each individual, and assume independence between the remaining five choice occasions and between each of these remaining choice occasions and the choice occasions chosen for the pairings. Basically, the CML approach is flexible, and allows customization based on the problem at hand. The issue then becomes one of balancing between speed gain/convergence improvement and efficiency loss. Besides, the CML can also use triplets or quadruplets rather than the couplets considered here.

If the probabilities of the lower dimensional events in the CML approach themselves have a multivariate normal cumulative distribution (MVNCD) form, then one can use the MACML approach proposed by Bhat to evaluate the MVNCD function using an analytic approximation.

One potential limitation of the CML approach is the need to compute the Godambe information matrix to compute the asymptotic standard errors of parameters. However, even when an MSL method is used, the Godambe matrix is recommended to accommodate the simulation error that accrues because of the use of a finite number of draws. Another limitation of the CML approach is the need to compute the ADCLRT statistic, which is somewhat more complicated than the traditional likelihood ration test (LRT) statistic. It is hoped that such practical issues will be resolved once standard econometric software packages start accommodating the CML inference approach as an option for high dimensional model systems.

In summary, the CML inference approach (and the associated MACML approach) can be very effective for the estimation and analysis of high-dimensional heterogeneous data. This has been shown in many recent studies, and there are many more empirical contexts that can gainfully use the CML approach using the formulations discussed in this paper. In terms of future research on the CML approach itself, one wide open area pertains to how best to form a CML function in a given modeling and empirical context (especially because a precise theoretical analysis of the properties of the CML estimator is not possible except for the simplest of models).

ACKNOWLEDGEMENTS

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The author would also like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany. One anonymous reviewer provided valuable comments on an earlier version of the paper. Finally, the author is grateful to Lisa Macias for her help in formatting this document.

REFERENCES

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422), 669-679.

Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht, The Netherlands.

Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89(1), 3-25.

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

Arbia, G., Kelejian, H., 2010. Advances in spatial econometrics. *Regional Science and Urban Economics* 40(5), 253-366.

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

Bartels, R., Fiebig, D.G., van Soest, A., 2006. Consumers and experts: an econometric analysis of the demand for water heaters. *Empirical Economics* 31(2), 369-391.

Beck, N., Gleditsch, K.S., Beardsley, K., 2006. Space is more than geography: using spatial econometrics in the study of political economy. *International Studies Quarterly* 50(1), 27-44.

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

Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B* 36(2), 192-236.

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

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

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

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B* 38(2), 147-168.

Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioral mechanisms for car ownership decisions. *Transportation Research Part B* 32(1), 61-75.

Bhat, C.R., Sener, I.N., 2009. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems* 11(3), 243-272.

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

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

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

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

Bhat, C.R., Eluru, N., Copperman, R.B., 2008. Flexible model structures for discrete choice analysis. In *Handbook of Transport Modelling, 2nd edition*, Chapter 5, Hensher, D.A., Button, K.J. (eds.), Elsevier Science, 75-104.

Bhat, C.R., Paleti, R., Singh, P., 2014a. A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, forthcoming.

Bhat, C.R., Sener, I.N., Eluru, N., 2010a. 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.

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

Bhat, C.R., Born, K., Sidharthan, R., Bhat, P.C., 2014b. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. *Analytic Methods in Accident Research* 1, 53-71.

Bradlow, E.T., Bronnenberg, B., Russell, G.J., Arora, N., Bell, D.R., Duvvuri, S.D., Hofstede, F.T., Sismeiro, C., Thomadsen, R., Yang, S., 2005. Spatial models in marketing. *Marketing Letters* 16(3), 267-278.

Brady, M., Irwin, E., 2011. Accounting for spatial effects in economic models of land use: recent developments and challenges ahead. *Environmental and Resource Economics* 48(3), 487-509.

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

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B* 46(1), 253-272.

Chen, M.-H., Dey, D.K., 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.R., 1972. The analysis of multivariate binary data. *Journal of the Royal Statistical Society* 21C(2), 113-120.

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

Davis, R.A. and Yau, C.Y. (2011). Comments of pairwise likelihood. *Statistica Sinica* 21, 255-277.

De Leon, A., Chough, K.C., 2013. *Analysis of Mixed Data: Methods and Applications*. CRC Press, Boca Raton.

Dube, J-P., Chintagunta, P., Petrin, A., Bronnenberg, B., Goettler, R., Seetharam, P.B., Sudhir, K., Tomadsen, R., Zhao, Y., 2002. Structural applications of the discrete choice model. *Marketing Letters* 13(3), 207-220.

Eidsvik, J., Shaby, B.A., Reich, B.J., Wheeler, M., and Niemi, J., 2013. Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, forthcoming.

Elhorst, J.P., 2010. Applied spatial econometrics: raising the bar. *Spatial Economic Analysis* 5(1), 9-28.

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

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

Feddag, M.-L., 2013. Composite likelihood estimation for multivariate probit latent traits models. *Communications in Statistics - Theory and Methods* 42(14), 2551-2566.

Ferdous, N., Bhat, C.R., 2013. A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. *Journal of Geographical Systems* 15(1), 1-29.

Ferdous, N., Eluru, N., Bhat, C.R., Meloni, I., 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.

Ferguson, T.S., 1996. *A Course in Large Sample Theory*. Chapman & Hall, London.

Fiebig, D.C., Keane, M.P., Louviere, J., Wasi, N., 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3), 393-421.

Fieuws, S., Verbeke, G., 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 62(2), 424-31.

Fleming, M.M., 2004. Techniques for estimating spatially dependent discrete choice models. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), Springer-Verlag, Berlin, 145-168.

Fotheringham, A.S., Brunsdon, C., 1999. Local forms of spatial analysis. *Geographical Analysis* 31(4), 340-358.

Franzese, R.J., Hays, J.C., 2008. Empirical models of spatial interdependence. In *The Oxford Handbook of Political Methodology*, Box-Steffensmeier, J.M., Brady, H.E., Collier, D., (eds.), Oxford University Press, Oxford, 570-604.

Franzese, R.J., Hays, J.C., Schaffer, L., 2010. Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: Estimation, interpretation, and presentation. *APSA 2010 Annual Meeting Paper*, August.

Gassmann, H.I., Deák, I., Szántai, T., 2002. Computing multivariate normal probabilities: A new look. *Journal of Computational and Graphical Statistics* 11(4), 920-949.

Girard, P., Parent, E., 2001. Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. *Technometrics*, 43(2), 180-190.

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

Greene, W.H., 2009. Models for count data with endogenous participation. *Empirical Economics* 36(1), 133-173.

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

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

Hays, J.C., Kachi, A., Franzese, R.J., 2010. A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology* 7(3), 406-428.

Heagerty, P.J., Lumley, T., 2000. Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* 95(449), 197-211.

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

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

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

Higham, N.J., 2002. Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis* 22(3), 329-343.

Hjort, N.L., Omre, H., 1994. Topics in spatial statistics (with discussion). *Scandinavian Journal of Statistics* 21(4), 289-357.

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

Huguenin, J., Pelgrin F., Holly A., 2009. Estimation of multivariate probit models by exact maximum likelihood. Working Paper 0902, University of Lausanne, Institute of Health Economics and Management (IEMS), Lausanne, Switzerland.

Jeliazkov, I., Graves, J., Kutzbach, M., 2008. Fitting and comparison of models for multivariate ordinal outcomes. In *Advances in Econometrics*, Volume 23, Bayesian Econometrics, Chib, S., Griffiths, W., Koop, G., Terrell, D. (eds.), Emerald Group Publishing Limited, Bingley, U.K., 115-156.

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

Joe, H., 2008. Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis* 52(12), 5066-5074.

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

Keane, M., 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10(2), 193-200.

Kent, J.T., 1982. Robust properties of likelihood ratio tests. *Biometrika* 69(1), 19-27.

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

LaMondia, J.J., Bhat, C.R., 2011. A study of visitors' leisure travel behavior in the northwest territories of Canada, *Transportation Letters: The International Journal of Transportation Research* 3(1), 1-19.

Larribe, F. Fearnhead, P., 2011. On composite likelihoods in statistical genetics. *Statistica Sinica* 21, 43-69.

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

LeSage, J.P., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL.

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

Luce, R., Suppes, P., 1965. Preference, utility and subjective probability. In *Handbook of Mathematical Probability*, Vol. 3, Luce, R., Bush, R., Galanter, E., (eds.), Wiley, New York.

Mardia, K.V., Hughes, G., Taylor, C.C., 2007. Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions. University of Leeds, UK. Available at: http://www.amsta.leeds.ac.uk/Statistics/research/reports/2007/STAT07-02.pdf

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

McCulloch, R.E., Rossi P.E., 2000. Bayesian analysis of the multinomial probit model. In *Simulation-Based Inference in Econometrics*, Mariano, R., Schuermann, T., Weeks, M.J., (eds.), Cambridge University Press, New York, 158-178.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, 105-142, Zarembka, P., (ed.), Academic Press, New York.

McFadden, D., 1978. Modeling the choice of residential location. *Transportation Research Record* 672, 72-77.

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

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

McMillen, D.P. 2010. Issues in spatial analysis. *Journal of Regional Science* 50(1), 119-141.

Mitchell, J., Weale, M., 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., Verbeke, G., 2005. *Models for Discrete Longitudinal Data*. Springer Series in Statistics, Springer Science + Business Media, Inc., New York.

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

Munkin, M.K., Trivedi, P.K., 2008. Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, 143(2), 334-348.

Pace, L., Salvan A., Sartori, N., 2011. Adjusting composite likelihood ratio statistics. *Statistica Sinica* 21(1), 129-148.

Paleti, R., Bhat, C.R., 2013. The composite marginal likelihood (CML) estimation of panel ordered-response models. *Journal of Choice Modelling* 7, 24-43.

Partridge, M.D., Boarnet, M., Brakman, S., Ottaviano, G., 2012. Introduction: Whither Spatial Econometrics? *Journal Of Regional Science* 57(2), 167-171.

Rebonato, R., Jaeckel, P., 1999. The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *The Journal of Risk* 2(2), 17-28.

Ruud, P.A., 2007. Estimating mixtures of discrete choice model, Technical Paper, University of California, Berkeley.

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

Schoettle. K., Werner, R., 2004. Improving "the most general methodology to create a valid correlation matrix". In *Risk Analysis IV, Management Information Systems*, Brebbia, C.A., ed., WIT Press, Southampton, U.K., 701-712.

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

Scott, D.M., Kanaroglou, P.S., 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.).

Sidharthan, R., Bhat, C.R., 2012. Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis* 44(4), 321-349.

Small, K.A., Winston, C., Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica* 73(4), 1367-1382.

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

Switzer, P., 1977. Estimation of spatial distribution from point sources with application to air pollution measurement. *Bulletin of the International Statistical Institute* 47(2), 123-137.

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

Vandekerkhove, P., 2005. Consistent and asymptotically normal parameter estimates for hidden markov mixtures of markov models. *Bernoulli* 11(1), 103-129.

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

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

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. *Biometrika* 92(3), 519-528.

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

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

Varin, C., Reid, N., Firth, D, 2011. An overview of composite marginal likelihoods. *Statistica Sinica* 21(1), 5-42.

Vasdekis, V.G.S., Cagnone, S., Moustaki, I., 2012. A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika* 77(3), 425-441.

Ver Hoef, J.M., Jansen, J.K., 2007. Space-time zero-inflated count models of harbor seals. *Environmetrics* 18(7), 697-712.

Wang, H., E. M. Iglesias and J. M. Wooldridge (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, 172, 77-89.

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

Wu, B., de Leon, A.R., Withanage, N., 2013. Joint analysis of mixed discrete and continuous outcomes via copula models. In de Leon, A. and K.C. Chough (eds.) *Analysis of Mixed Data: Methods and Applications*, CRC Press, Boca Raton, 139-156.

Xu, X., Reid, N., 2011. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference*, 141(9), 3047-3054.

Yi, G.Y., Zeng, L., Cook, R.J., 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *The Canadian Journal of Statistics* 39(1), 34-51.

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

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