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
Austin TX 78712, USA
bhat@mail.utexas.edu

King Abdulaziz University
Jeddah 21589, Saudi Arabia

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

C. R. Bhat. *The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Models.* Foundations and Trends® in Econometrics, vol. 7, no. 1, pp. 1–117, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-60198-829-4

Full text available at: http://dx.doi.org/10.1561/0800000022

Foundations and Trends® in Econometrics

Volume 7, Issue 1, 2014

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Econometrics publishes survey and tutorial articles in the following topics:

- Econometric models
- Simultaneous equation models
- Estimation frameworks
- Biased estimation
- Computational problems
- Microeconometrics
- Treatment modeling
- Discrete choice modeling
- Models for count data
- Duration models
- Limited dependent variables
- Panel data
- Time series analysis
- Latent variable models
- Qualitative response models
- Hypothesis testing
- Econometric theory
- Financial econometrics
- Measurement error in survey data
- Productivity measurement and analysis
- Semiparametric and nonparametric estimation
- Bootstrap methods
- Nonstationary time series
- Robust estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2014, Volume 7, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Foundations and Trends® in Econometrics
Vol. 7, No. 1 (2014) 1–117

DOI: 10.1561/0800000022

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
Austin TX 78712, USA
bhat@mail.utexas.edu

King Abdulaziz University
Jeddah 21589, Saudi Arabia

Contents

1 Introduction **2**

1.1 Background . . . 2

1.2 Types of CL methods . . . 3

1.3 The composite marginal likelihood (CML) inference approach . . . 4

1.4 Asymptotic properties of the CML estimator with many independent replicates . . . 6

1.5 Asymptotic properties of the CML estimator for the case of very few or no independent replicates . . . 12

1.6 Relative efficiency of the CML estimator . . . 12

1.7 Robustness of consistency of the CML estimator . . . 19

1.8 Model selection in the CML inference approach . . . 19

1.9 Positive-definiteness of the implied multivariate covariance matrix . . . 22

1.10 The maximum approximate composite marginal likelihood approach . . . 23

2 Application to Traditional Discrete Choice Models **28**

2.1 Ordered and unordered-response model systems . . . 29

2.2 Aspatial formulations . . . 30

2.3 Spatial formulations 58
2.4 Application to count models 86

3 Application to Joint Mixed Model Systems 91

3.1 Joint mixed dependent variable model formulation 93
3.2 The joint mixed model system and the likelihood formation 96
3.3 The joint mixed model system and the MACML estimation approach 99
3.4 Positive definiteness 101

4 Conclusions 103

Acknowlegements 106

References 108

Abstract

This monograph 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 monograph 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, it develops a blueprint (complete with matrix notation) to apply the CML estimation technique to a wide variety of discrete and mixed dependent variable models.

C. R. Bhat. *The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Models*. Foundations and Trends® in Econometrics, vol. 7, no. 1, pp. 1–117, 2014.

DOI: 10.1561/0800000022.

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. 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 monograph, 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, Larribe and Fearnhead, 2011, for reviews]. However, the CL method has found little coverage in the econometrics field, and it is the hope that this monograph 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 $(\tilde{H} \times 1)$ vector $\boldsymbol{Y}$ with density function $f(\boldsymbol{y}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown $\tilde{K}$-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 $(A_1(\boldsymbol{m}), A_2(\boldsymbol{m}), \ldots, A_{\tilde{E}}(\boldsymbol{m}))$. 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:

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

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 monograph, 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, 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 $(t = 1, 2, 3, \ldots, T)$. 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]:

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

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:

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

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 even when there is dependence among the observations of the individual. 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:

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

[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.

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, 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 a sample $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3, \ldots, \boldsymbol{Y}_Q$, each $\boldsymbol{Y}_q$ $(q = 1, 2, \ldots, Q)$ being a $\tilde{H} \times 1$ vector. That is, $\boldsymbol{Y}_q = (Y_{q1}, Y_{q2}, \ldots, Y_{q\tilde{H}})$. $\tilde{H}$ 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 $\boldsymbol{y_q} = (y_{q1}, y_{q2}, \ldots, y_{q\tilde{H}})$. Define $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_Q)$. 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:

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

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, p. 191), the maximization of the logarithm of the CML function in the equation above is achieved by solving the composite score equations given by:

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

where $\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'}) = \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}}$. Since the equations $\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}, \boldsymbol{y})$ 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:

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

where the expectation above is taken with respect to the full distribution of $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{\tilde{H}})$. 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:

$$\begin{aligned} E[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}, y_{qh}, y_{qh'})] &= \int_{y_q} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} f(\boldsymbol{y_q}) \boldsymbol{dy_q} \\ &= \int_{y_{qd}} \int_{y_{qd'}} \int_{\boldsymbol{y_{-qdd'}}} \frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} \\ &\quad \times f(y_{qh}, y_{qh'}, \boldsymbol{y_{-qhh'}}) dy_{qh} dy_{qh'} \boldsymbol{dy_{-qhh'}}, \end{aligned} \tag{1.8}$$

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

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

$$= \int_{y_{qh}} \int_{y_{qh'}} \frac{\partial L_{qhh'}}{\partial \boldsymbol{\theta}} dy_{qh} dy_{qh'}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{y_{qh}} \int_{y_{qh'}} L_{qhh'} dy_{qh} dy_{qh'} = \frac{\partial}{\partial \boldsymbol{\theta}}(1) = \mathbf{0}. \tag{1.9}$$

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 $\boldsymbol{J} = \text{Var}[\boldsymbol{s_q}(\boldsymbol{\theta_0}, \boldsymbol{y_q})]$, 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

$$\sqrt{Q}\boldsymbol{s}(\boldsymbol{\theta_0}, \boldsymbol{y}) \xrightarrow{d} \text{MVN}_{\tilde{K}}(\mathbf{0}, \boldsymbol{J}), \tag{1.10}$$

where $\text{MVN}_{\tilde{K}}(.,.)$ stands for the multivariate normal distribution of $\tilde{K}$ dimensions. Next, let $\hat{\boldsymbol{\theta}}_{\text{CML}}$ be the CML estimator, so that, by design of the CML estimator, $\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{\text{CML}}, \boldsymbol{y}) = \mathbf{0}$. Expanding $\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{\text{CML}}, \boldsymbol{y})$ around $\boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})$ in a first-order Taylor series, we obtain $\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{\text{CML}}, \boldsymbol{y}) = \mathbf{0} = \boldsymbol{s}(\boldsymbol{\theta_0}, \boldsymbol{y}) + \nabla \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})[\hat{\boldsymbol{\theta}}_{\text{CML}} - \boldsymbol{\theta}_0]$, or equivalently,

$$\sqrt{Q}[\hat{\boldsymbol{\theta}}_{\text{CML}} - \boldsymbol{\theta}_0] = \sqrt{Q}[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})]^{-1} \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y}). \tag{1.11}$$

From the law of large numbers (LLN), we also have that $\nabla \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})$, which is the sample mean of $\nabla \boldsymbol{s}_q(\boldsymbol{\theta}_0, \boldsymbol{y}_q)$, converges to the population mean for the quantity. That is,

$$[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})] \xrightarrow{d} \boldsymbol{H} = E[-\nabla \boldsymbol{s}(\boldsymbol{\theta}_0, \boldsymbol{y})] \tag{1.12}$$

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:

$$\sqrt{Q}[\hat{\boldsymbol{\theta}}_{\text{CML}} - \boldsymbol{\theta}_0] \xrightarrow{d} MVN_{\tilde{K}}(\boldsymbol{0}, \boldsymbol{G}^{-1}), \quad \text{where } \boldsymbol{G} = \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}, \tag{1.13}$$

where $\boldsymbol{G}$ is the Godambe [1960] information matrix. Thus, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{CML}}$ is centered on the true parameter vector $\boldsymbol{\theta}_0$. Further, the variance of $\hat{\boldsymbol{\theta}}_{\text{CML}}$ reduces as the number of sample points Q increases. The net result is that $\hat{\boldsymbol{\theta}}_{\text{CML}}$ converges in probability to $\boldsymbol{\theta}_0$ as $Q \to \infty$ (with $\tilde{H}$ fixed), leading to the consistency of the estimator. In addition, $\hat{\boldsymbol{\theta}}_{\text{CML}}$ 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 $\boldsymbol{\theta}_0$. But $\boldsymbol{J}$ and $\boldsymbol{H}$ may be estimated in a straightforward manner at the CML estimate $\hat{\boldsymbol{\theta}}_{\text{CML}}$ as follows:

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

and

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

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

$$\begin{aligned} \boldsymbol{J_q} &= \mathrm{Var}[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}] = -\boldsymbol{H_q} = -E[-\nabla \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}] \\ &= E[\nabla \boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}], \end{aligned} \tag{1.16}$$

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

$$\begin{aligned} \hat{\boldsymbol{H}} &= \frac{1}{Q} \sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=1}^{\tilde{H}} \mathrm{Var}[\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}]_{\hat{\boldsymbol{\theta}}_{\mathrm{CML}}} \\ &= \frac{1}{Q} \sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=1}^{\tilde{H}} ([\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}][\boldsymbol{s}_{qhh'}(\boldsymbol{\theta}_0, y_{qh}, y_{qh'}]')_{\hat{\boldsymbol{\theta}}_{\mathrm{CML}}} \\ &= \frac{1}{Q} \sum_{q=1}^{Q} \sum_{h=1}^{\tilde{H}-1} \sum_{h'=1}^{\tilde{H}} \left(\left[\frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial \log L_{qhh'}}{\partial \boldsymbol{\theta}'} \right] \right)_{\hat{\boldsymbol{\theta}}_{\mathrm{CML}}} \end{aligned} \tag{1.17}$$

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

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]:

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{\mathrm{CML}}) = \frac{Q-1}{Q} \sum_{q=1}^{Q} (\hat{\boldsymbol{\theta}}_{\mathrm{CML}}^{(-q)} - \hat{\boldsymbol{\theta}}_{\mathrm{CML}})(\hat{\boldsymbol{\theta}}_{\mathrm{CML}}^{(-q)} - \hat{\boldsymbol{\theta}}_{\mathrm{CML}})', \tag{1.18}$$

where $\hat{\boldsymbol{\theta}}_{\mathbf{CML}}^{(-q)}$ 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 $\hat{\boldsymbol{\theta}}_{\mathbf{CML}}^{(-q)}$ with a single step of the Newton–Raphson algorithm with $\hat{\boldsymbol{\theta}}_{\mathbf{CML}}$ 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).

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 $\log L_{\text{CML}}(\boldsymbol{\theta})$, 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

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

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 $E[\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})] = \mathbf{0}$, 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 $\boldsymbol{Y}$):

$$E[\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})] = \mathbf{0} = \int_y \frac{\partial \log L_{\text{CML}}}{\partial \boldsymbol{\theta}} f(\boldsymbol{y}) \boldsymbol{dy} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$
$$= \int_y \frac{\partial \log L_{\text{CML}}}{\partial \boldsymbol{\theta}} L_{\text{ML}} \boldsymbol{dy} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \quad (1.19)$$

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

$$\mathbf{0} = \int_y \frac{\partial^2 \log L_{\text{CML}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L_{\text{ML}} \boldsymbol{dy} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$
$$+ \int_y \frac{\partial \log L_{\text{CML}}}{\partial \boldsymbol{\theta}} \frac{\partial \log L_{\text{ML}}}{\partial \boldsymbol{\theta}} L_{\text{ML}} \boldsymbol{dy} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$
$$= E[\nabla \boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})] + E[\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y}) \boldsymbol{s}_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y})], \quad (1.20)$$

where $\boldsymbol{s}_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y})$ is the score function of the full likelihood. From above, we get the following:

$$\boldsymbol{H} = -E[\nabla \boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})] = \text{Cov}[\boldsymbol{s}_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y}), \boldsymbol{s}'_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})]|,$$

and

$$\boldsymbol{G} = \text{Cov}[\boldsymbol{s}_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y}), \boldsymbol{s}'_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y})]|[\text{Var}(\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y}))]^{-1}]$$
$$\times \text{Cov}[\boldsymbol{s}_{\text{CML}}(\boldsymbol{\theta}_0, \boldsymbol{y}), \boldsymbol{s}'_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y}),] \quad (1.21)$$

Then, using the multivariate version of the Cauchy–Schwartz inequality [Lindsay, 1988], we obtain the following:

$$\boldsymbol{IFISHER} = \text{Var}[\boldsymbol{s}_{\text{ML}}(\boldsymbol{\theta}_0, \boldsymbol{y})] \geq \boldsymbol{G}. \quad (1.22)$$

Thus, from a theoretical standpoint, the difference between the regular ML information matrix (i.e., ***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 $\boldsymbol{IFISHER}$ and the Godambe matrix $\boldsymbol{G}$ 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 $\boldsymbol{IFISHER}$ and $\boldsymbol{G}$ 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 $\boldsymbol{IFISHER}$ and $\boldsymbol{G}$ 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 $\boldsymbol{IFISHER}$ and $\boldsymbol{G}$ 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. [2014]. 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. Fieuws 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. [2014] 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 $H_0\colon \boldsymbol{\tau} = \boldsymbol{\tau_0}$ against $H_1\colon \boldsymbol{\tau} \neq \boldsymbol{\tau_0}$, where τ is a subvector of $\boldsymbol{\theta}$ of dimension $\tilde{d}$; i.e., $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\alpha}')'$. The statistic takes the familiar form shown below:

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

where $\hat{\boldsymbol{\theta}}_{\mathbf{R}}$ is the composite marginal likelihood estimator under the null hypothesis $(\boldsymbol{\tau}_{\mathbf{0}}', \hat{\boldsymbol{\alpha}}_{\text{CML}}'(\boldsymbol{\tau}_{\mathbf{0}}))$. 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 $\tilde{d} \times \tilde{d}$ 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:

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

where $\tilde{W}_i^2$ for $i = 1, 2, \ldots, \tilde{d}$ are independent χ_1^2 variates and $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$ 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 $\tilde{d}$ 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 $\tilde{d}$ degrees of freedom:

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

where $\boldsymbol{S}_\tau(\boldsymbol{\theta})$ is the $\tilde{d} \times 1$ submatrix of $S(\boldsymbol{\theta}) = \left(\frac{\partial \log L_{\text{CML}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)$ corresponding to the vector τ, 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 $\chi^2_{\tilde{d}}$ null distribution. Thus, $ADCLRT$ is also very close to being an asymptotic $\chi^2_{\tilde{d}}$ 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 $\tilde{\boldsymbol{y}}_{\mathbf{obs}}$, and the observed $CLRT$ value as $CLRT(\tilde{\boldsymbol{y}}_{\mathbf{obs}})$.
2. Generate C sample data sets $\tilde{\boldsymbol{y}}_{\mathbf{1}}, \tilde{\boldsymbol{y}}_{\mathbf{2}}, \tilde{\boldsymbol{y}}_{\mathbf{3}}, \ldots, \tilde{\boldsymbol{y}}_{\boldsymbol{C}}$ 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:

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

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]:

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

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

[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., $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta})$) and the CLIC in this case is exactly the AIC [$= \log L_{\mathrm{ML}}(\hat{\boldsymbol{\theta}})$ — (# of model parameters)].

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 $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ matrix. The first technique is to use Bhat and Srinivasan's (2005) strategy of reparameterizing the correlation matrix $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ matrix, and then pick off the appropriate elements of this matrix to 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 monograph). 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 integration of a multivariate normal cumulative distribution (MVNCD) function of dimension equal to $2 \times (I-1)$. The evaluation of such an integration 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, 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, 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 $(W_1, W_2, W_3, \ldots, W_I)$ 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:

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

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

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

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

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

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

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

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

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

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

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

where

$$\begin{aligned}\mathbf{\Omega}_{<i} &= \mathrm{Cov}(\boldsymbol{I}_{<i}, \boldsymbol{I}_{<i})\\ &= \begin{bmatrix} \mathrm{Cov}(\tilde{I}_1, \tilde{I}_1) & \mathrm{Cov}(\tilde{I}_1, \tilde{I}_2) & \mathrm{Cov}(\tilde{I}_1, \tilde{I}_3) & \cdots & \mathrm{Cov}(\tilde{I}_1, \tilde{I}_{i-1}) \\ \mathrm{Cov}(\tilde{I}_2, \tilde{I}_1) & \mathrm{Cov}(\tilde{I}_2, \tilde{I}_2) & \mathrm{Cov}(\tilde{I}_2, \tilde{I}_3) & \cdots & \mathrm{Cov}(\tilde{I}_2, \tilde{I}_{i-1}) \\ \mathrm{Cov}(\tilde{I}_3, \tilde{I}_1) & \mathrm{Cov}(I_3, I_2) & \mathrm{Cov}(\tilde{I}_3, \tilde{I}_3) & \cdots & \mathrm{Cov}(\tilde{I}_3, \tilde{I}_{i-1}) \\ \vdots & & & & \\ \mathrm{Cov}(\tilde{I}_{i-1}, \tilde{I}_1) & \mathrm{Cov}(\tilde{I}_{i-1}, \tilde{I}_2) & \mathrm{Cov}(\tilde{I}_{i-1}, \tilde{I}_3) & \cdots & \mathrm{Cov}(\tilde{I}_{i-1}, \tilde{I}_{i-1}) \end{bmatrix},\end{aligned}$$

[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 $\tilde{I}_i$ as the "dependent" random variable and $\breve{I}_{<i} = (\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_{i-1}, \tilde{I}_{12}, \tilde{I}_{13}, \dots I_{1,i-1}, \tilde{I}_{23}, \tilde{I}_{24}, \dots \tilde{I}_{2,i-1}, \dots I_{i-2,i-1})$ as the independent random variable vector, where $\tilde{I}_{i'j'} = \tilde{I}_{i'}\tilde{I}_{j'}$. 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 monograph, we will use the term approximation to refer to the first-order approximation evaluation of the MVNCD function.

and

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

Finally, putting the estimate of $\hat{\boldsymbol{\alpha}}$ back in Equation (1.32), and predicting the expected value of $\tilde{I}_i$ conditional on $\mathbf{2}\tilde{\boldsymbol{I}}_{\boldsymbol{<i}} = \mathbf{1}$ (i.e., $\tilde{I}_1 = 1, \tilde{I}_2 = 1, \tilde{I}_{i-1} = 1$), we get the following approximation for Equation (1.31):

$$\begin{aligned} &\Pr(W_i < w_i | W_1 < w_1, W_2 < w_2, \ldots, W_{i-1} < w_{i-1}) \\ &\quad \approx \Phi(w_i) + (\boldsymbol{\Omega}_{\boldsymbol{<i}}^{\mathbf{-1}} \cdot \boldsymbol{\Omega}_{\boldsymbol{i,<i}})'(1 - \Phi(w_1), 1 - \Phi(w_2) \ldots 1 - \Phi(w_{i-1}))' \end{aligned} \tag{1.34}$$

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 $\boldsymbol{W} = (W_1, W_2, W_3, \ldots, W_I)$) 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).

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C. R. Bhat and R. Sidharthan. 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, 2011.

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

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

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

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

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

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

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

C. R. Bhat, R. Paleti, and P. Singh. A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, 54(3):462–502, 2014b.

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

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

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

M. Castro, R. Paleti, and C. R. Bhat. 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, 2012.

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

D. R. Cox. The analysis of multivariate binary data. *Journal of the Royal Statistical Society*, 21C(2):113–120, 1972.

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

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

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

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

J. Eidsvik, B. A. Shaby, B. J. Reich, M. Wheeler, and J. Niemi. Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23(2):295–315, 2014.

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

N. Eluru, C. R. Bhat, and D. A. Hensher. 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, 2008.

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

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

N. Ferdous and C. R. Bhat. 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, 2013.

N. Ferdous, N. Eluru, C. R. Bhat, and I. Meloni. 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, 2010.

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

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

S. Fieuws and G. Verbeke. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431, 2006.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

I. Jeliazkov, J. Graves, and M. Kutzbach. Fitting and comparison of models for multivariate ordinal outcomes. In S. Chib, W. Griffiths, G. Koop, and D. Terrell, editors, *Advances in Econometrics*, vol. 23*, Bayesian Econometrics*, pages 115–156. Emerald Group Publishing Limited, Bingley, UK, 2008.

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

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

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

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

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

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

J. J. LaMondia and C. R. Bhat. 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, 2011.

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

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

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

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

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

K. V. Mardia, G. Hughes, and C. C. Taylor. Efficiency of the pseudo-likelihood 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, 2007.

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

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

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

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

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

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

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

J. Mitchell and M. Weale. The reliability of expectations reported by british households: Micro evidence from the bhps. National Institute of Economic and Social Research Discussion Paper, 2007.

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

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

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

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

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

M. D. Partridge, M. Boarnet, S. Brakman, and G. Ottaviano. Introduction: Whither spatial econometrics? *Journal of Regional Science*, 57(2):167–171, 2012.

R. Rebonato and P. Jaeckel. 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, 1999.

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

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

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

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

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

C. Scotti. 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.), 2006.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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