Efficient specification and Estimation of Choice Models in Activity-Travel Model Systems

Chandra R. Bhat
University of Texas at Austin
Presentation outline

- Introduction
- Discrete-Continuous
- Spatial Dependency
- Emerging Estimation Technique for discrete choice models
- Conclusion
Acknowledgments

- TxDOT, NCTCOG, SCAG
- Kostas Goulias and Ram Pendyala
- Jessica Guo, Siva Srinivasan, Abdul Pinjari, Rachel Copperman
- Naveen Eluru, Rajesh Paleti, Nazneen Ferdous
Introduction - Discrete-Continuous frameworks (D/C)

- Characterized by a continuous variable related to a discrete variable
  - Selective sample observation effect: Continuous outcome observed only if a discrete condition is met. Examples: Household income observation, GPS-based data
  - Endogenous treatment effect: The continuous equation depends on a discrete explanatory variable that is determined endogenously with the continuous variable. Examples: job training program – wages, seat belt use – injury severity
Introduction – Spatially and Socially dependent choice processes (SD)

- Characterized by a choice process influenced by unobserved error dependency based on spatial location
  - Spatial dependence across alternatives
  - Spatial dependence across observational units

- Tendency of data points to be similar when closer in space
  - Diffusion effects
  - Social interaction effects
  - Unobserved location-related effects

- Examples: Residential location choice, Physical activity participation
State of the field – D/C

- Discrete choice models have seen substantial advancement in recent years
  - Mixed logit and advances in simulation
- Not the same level of maturity in discrete-continuous frameworks
  - “The field is still expanding more than it is coalescing”
    - Train
- Approaches
  - Heckman or Lee’s approach
  - Semi-parametric and Non-parametric approaches
  - More recently Copula approach
State of the field – SD

- Spatial correlation across alternatives: choices correspond to spatial units.
  - Transportation and geography literature.
  - Common model structures include mixed logit, multinomial probit, GEV-based spatially correlated models.

- Spatial correlation across observational units: choices among the aspatial alternatives may be moderated by space.
  - Regional science and political science literature
  - Common model structure include Binary spatial probit model estimated using McMillen’s EM, LeSage’s MCMC etc.
D/C frameworks

Direct and indirect utility approaches to modeling discrete/continuous frameworks

- Typically D/C approaches begin with constrained direct utility functions.
- This constrained direct utility function can be equivalently represented by an indirect utility function.
- Once an indirect utility function is chosen, deriving demand functions is relatively easy.
- However, recently studies have started employing direct utility functions to model D/C frameworks particularly for multiple-discrete choices.
- An explicit framework employing direct utility functions applicable to multiple discrete problems is discussed in detail.
Why multiple-discreteness

Several consumer demand choices are characterized by multiple discreteness

- Vehicle type holdings and usage
- Household consumption patterns on consumer services/goods
- Activity type choice and duration of participation
- Airline fleet mix and usage
- Carrier choice and transaction level
- Brand choice and purchase quantity
- Stock choice and investment amount
Modeling methodologies of multiple discrete situations

- Traditional random utility-based (RUM) single discrete choice models
  - Number of composite alternatives explodes with the number of elemental alternatives

- Multivariate probit (logit) methods
  - Not based on a rigorous underlying utility-maximizing framework of multiple discreteness

- Other issues with these methods
  - Cannot accommodate the diminishing marginal returns (i.e., satiation) in the consumption of an alternative
  - Cumbersome to include a continuous dimension of choice
Modeling methodologies of multiple discrete situations

- Two alternative methods proposed by Wales and Woodland (1983)
  - Amemiya-Tobin approach
  - Kuhn-Tucker approach

- Both approaches assume a direct utility function $U(x)$ that is assumed to be quasi-concave, increasing, and continuously differentiable with respect to the consumption quantity vector $x$

- Approaches differ in how stochasticity, non-negativity of consumption, and corner solutions (i.e., zero consumption of some goods) are accommodated
Methods proposed by Wales and Woodland

- Amemiya-Tobin approach
  - Extension of the classic microeconomic approach of adding normally distributed stochastic terms to the budget-constrained utility-maximizing share equations
  - Direct utility function $U(x)$ assumed to be deterministic by the analyst, and stochasticity is introduced post-utility maximization

- Kuhn-Tucker (KT) approach
  - Based on the Kuhn Tucker or KT (1951) first-order conditions for constrained random utility maximization
  - Employs a direct stochastic specification by assuming the utility function $U(x)$ to be random (from the analyst’s perspective) over the population
  - Derives the consumption vector for the random utility specification subject to the linear budget constraint by using the KT conditions for constrained optimization
  - Stochastic nature of the consumption vector in the KT approach is based fundamentally on the stochastic nature of the utility function
Advantages of KT approach

- Constitutes a more theoretically unified and consistent framework for dealing with multiple discreteness consumption patterns
- Satisfies all the restrictions of utility theory
- Stochastic KT first-order conditions provide the basis for deriving the probabilities for each possible combination of corner solutions (zero consumption) for some goods and interior solutions (strictly positive consumption) for other goods
- Accommodates for the singularity imposed by the “adding-up” constraint

Problems with KT approach used by Wade and Woodland

- Random utility distribution assumptions lead to a complicated likelihood function that entails multi-dimensional integration
Studies that used the KT approach for multiple discreteness

- Kim et al. (2002)
  - Used the GHK simulator to evaluate the multivariate normal integral appearing in the likelihood function in the KT approach
  - Used a generalized variant of the well-known translated constant elasticity of substitution (CES) direct utility function
  - Not realistic for practical applications and is unnecessarily complicated

- Bhat (2005)
  - Introduced a simple and parsimonious econometric approach to handle multiple discreteness
  - Based on the generalized variant of the translated CES utility function but with a multiplicative log-extreme value error term
  - Labeled as the multiple discrete-continuous extreme value (MDCEV) model
  - MDCEV model represents the multinomial logit (MNL) form-equivalent for multiple discrete-continuous choice analysis and collapses exactly to the MNL in the case that each (and every) decision-maker chooses only one alternative

Several studies in the environmental economics field

- Phaneuf et al., 2000; von Haefen et al., 2004; von Haefen, 2003a; von Haefen, 2004; von Haefen and Phaneuf, 2005; Phaneuf and Smith, 2005
  - Used variants of the linear expenditure system (LES) and the translated CES for the utility functions, and used multiplicative log-extreme value errors
MDCEV

Functional form of utility function

\[ U(x) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left\{ \left( \frac{x_k}{y_k} + 1 \right)^{\alpha_k} - 1 \right\} \]

- \( U(x) \) is a quasi-concave, increasing, and continuously differentiable function with respect to the consumption quantity vector \( x \)
- \( \psi_k, \gamma_k \) and \( \alpha_k \) are parameters associated with good \( k \)
Assumptions

- Additive separability
  - All the goods are strictly Hicksian substitutes
  - Marginal utility with respect to any good is independent of the level of consumption of other goods

- Weak complementarity
Role of $\psi_k$

\[
\frac{\partial U(x)}{\partial x_k} = \psi_k \left( \frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k^{-1}}
\]

- $\psi_k$ represents the baseline marginal utility, or the marginal utility at the point of zero consumption.
- Higher baseline $\psi_k$ implies less likelihood of a corner solution for good $k$. 
Role of $\gamma_k$

Indifference Curves Corresponding to Different Values of $\gamma_1$

- $\psi_1 = \psi_2 = 1$
- $\alpha_1 = \alpha_2 = 0.5$
- $\gamma_2 = 1$

$\gamma_1 = 0.25$

$\gamma_1 = 1$

$\gamma_1 = 2$

$\gamma_1 = 5$
Role of $\gamma_k$

Effect of $\gamma_k$ Value on Good $k$'s Subutility Function Profile
Role of $\alpha_k$

\[ \psi_k = 1 \quad \gamma_k = 1 \]

Effect of $\alpha_k$ Value on Good $k$'s Subutility Function Profile
Empirical identification issues associated with utility form

Alternative Profiles for Moderate Satiation Effects with Low $\alpha_k$ Value and High $\gamma_k$ Value

$\psi_k = 1$ for all profiles
Empirical identification issues associated with utility form—cont’d

Alternative Profiles for Moderate Satiation Effects with High $\alpha_k$ Value and Low $\gamma_k$ Value
Empirical identification issues associated with utility form-cont’d

Alternative Profiles for Low Satiation Effects with High $\alpha_k$ Value and High Value

- Combination profile ($\alpha_k = 0.7$, $\gamma_k = 30$)
- $\gamma_k$ - profile ($\alpha_k^* \rightarrow 0$, $\gamma_k^* = 218$)
- $\alpha_k$ - profile ($\alpha_k^{**} = 0.898$, $\gamma_k^{**} = 1$)

$\psi_k = 1$ for all profiles

Utility Accrued Due to Consumption of Good $k$

Consumption Quantity of Good $k$
Empirical identification issues associated with utility form—cont’d

Alternative Profiles for High Satiation Effects with Low $\alpha_k$ Value and Low $\gamma_k$ Value

$\psi_k = 1$ for all profiles

- Combination profile ($\alpha_k = 0.2, \gamma_k = 2$)
- $\gamma_k$-profile ($\alpha_k = 0, \gamma_k = 3.85$)
- $\alpha_k$-profile ($\alpha_k = 0.34, \gamma_k = 1$)
Stochastic form of utility function

- **Overall random utility function**

  \[
  U(x) = \sum_k \frac{\gamma_k}{\alpha_k} \left\{ \exp(\beta'z_k + \varepsilon_k) \cdot \left( \frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\}
  \]

- **Random utility function for optimal expenditure allocations**

  \[
  U(x) = \sum_k \frac{\gamma_k}{\alpha_k} \exp(\beta'z_k + \varepsilon_k) \cdot \left\{ \frac{e_k}{\gamma_k p_k} + 1 \right\}^{\alpha_k} - 1
  \]
KT conditions

\[ V_k + \varepsilon_k = V_1 + \varepsilon_1 \quad \text{if} \quad e_k^* > 0 \quad (k = 2, 3, \ldots, K) \]

\[ V_k + \varepsilon_k < V_1 + \varepsilon_1 \quad \text{if} \quad e_k^* = 0 \quad (k = 2, 3, \ldots, K), \quad \text{where} \]

\[ V_k = \beta'z_k + (\alpha_k - 1) \ln \left( \frac{e_k^*}{\gamma_k p_k} + 1 \right) - \ln p_k \quad (k = 1, 2, 3, \ldots, K) \]
General econometric model structure and identification

\[ P(e_1^*, e_2^*, e_3^*, \ldots, e_M^*, 0, 0, \ldots, 0) = |J| \int_{\varepsilon_1=\infty}^{+\infty} \int_{\varepsilon_{M+1}=\infty}^{+\infty} \int_{\varepsilon_{M+2}=\infty}^{+\infty} \cdots \int_{\varepsilon_{K-1}=\infty}^{+\infty} \int_{\varepsilon_{K}=\infty}^{+\infty} f(\varepsilon_1, V_1 - V_2 + \varepsilon_1, V_1 - V_3 + \varepsilon_1, \ldots, V_1 - V_M + \varepsilon_1, \varepsilon_{M+1}, \varepsilon_{M+2}, \ldots, \varepsilon_{K-1}, \varepsilon_K) \, d\varepsilon_K \, d\varepsilon_{K-1} \cdots d\varepsilon_{M+2} \, d\varepsilon_{M+1} \, d\varepsilon_1, \]

where \( J \) is the Jacobian whose elements are given by (see Bhat, 2005a):

\[ J_{ih} = \frac{\partial [V_{i+1} - V_i + \varepsilon_1]}{\partial e_{i+1}^*}; \ i, h = 1, 2, \ldots, M - 1 \]

\[ P(e_1^*, e_2^*, e_3^*, \ldots, e_M^*, 0, 0, \ldots, 0) = |J| \int_{\tilde\varepsilon_{M+1,1}=\infty}^{+\infty} \int_{\tilde\varepsilon_{M+2,1}=\infty}^{+\infty} \cdots \int_{\tilde\varepsilon_{K-1,1}=\infty}^{+\infty} \int_{\tilde\varepsilon_{K,1}=\infty}^{+\infty} g(V_1 - V_2, V_1 - V_3, \ldots, V_1 - V_M, \tilde\varepsilon_{M+1,1}, \tilde\varepsilon_{M+2,1}, \ldots, \tilde\varepsilon_{K,1}) \, d\tilde\varepsilon_{K,1} \, d\tilde\varepsilon_{K-1,1} \cdots d\tilde\varepsilon_{M+1,1} \]
Specific model structures

- The MDCEV model structure

\[ P \ e_1^*, e_2^*, e_3^*, ..., e_M^*, 0, 0, ..., 0 \]

\[
= |J| \int_{\varepsilon_1=\pm \infty} \left\{ \left( \prod_{i=2}^{M} \frac{1}{\sigma} \lambda \left[ \frac{V_1 - V_i + \varepsilon_1}{\sigma} \right] \right) \right\} \times \left\{ \prod_{s=M+1}^{K} \Lambda \left[ \frac{V_1 - V_s + \varepsilon_1}{\sigma} \right] \right\} \frac{1}{\sigma} \lambda \left( \frac{\varepsilon_1}{\sigma} \right) d\varepsilon_1
\]

\[
|J| = \left( \prod_{i=1}^{M} c_i \right) \left( \sum_{i=1}^{M} \frac{1}{c_i} \right), \text{ where } c_i = \left( \frac{1 - \alpha_i}{e_i^* + \gamma_i p_i} \right)
\]

\[
P \ e_1^*, e_2^*, e_3^*, ..., e_M^*, 0, 0, ..., 0 = \frac{1}{\sigma^{M-1}} \left[ \prod_{i=1}^{M} c_i \right] \left[ \sum_{i=1}^{M} \frac{1}{c_i} \right] \left[ \prod_{i=1}^{K} \frac{e^{V_i/\sigma}}{\left( \sum_{k=1}^{M} e^{V_k/\sigma} \right)^M} \right] (M-1)!
\]
MDCEV model structure cont’d

- Probability of the consumption pattern of the goods (rather than the expenditure pattern) is

\[
P\left(x_1^*, x_2^*, x_3^*, \ldots, x_M^*, 0, 0, \ldots, 0\right)
\]

\[
= \frac{1}{p_1} \cdot \frac{1}{\sigma^{M-1}} \left[ \prod_{i=1}^{M} f_i \right] \left[ \sum_{i=1}^{M} \frac{p_i}{f_i} \right] \left[ \prod_{i=1}^{M} e^{V_i / \sigma} \right] \left[ \sum_{k=1}^{K} e^{V_k / \sigma} \right]^M (M - 1)!,
\]

where

\[
f_i = \left( \frac{1 - \alpha_i}{x_i^* + \gamma_i} \right)
\]
MDCEV in an activity-based context

- Growing interest in accommodating joint activity participation across household members
- In conventional discrete choice frameworks, the need to generate mutually exclusive alternatives results in an explosion in choice sets
- MDCEV allows us to tackle the problem by considering activity participation as a household decision.
- MDCEV offers substantial computational and behavioral advantages
  - Employ one model to generate activity participation for all household members as opposed to one model per activity type and per person while simultaneously accommodating for joint activity participation
  - Accommodate substitution/complementarity in activity participation and household member dimensions
Activity Generation Framework

For household with \( P \) members

For each activity purpose:

- **Alone**: \( 1, 2, \ldots, P \)
- **Two**: \( 1,2, 1,3, 1,P, 2,3, \ldots, P \)
- **All Adults**: \( 1,2,\ldots,P \)

Alternatives for Activity Type = \( 2^P - 1 \)
MDCEV Framework

Overall choice process (for A activity purposes)

Total Choice Alternatives = \((2^p-1)(A) + 1\)
Traditional Framework

- Within single discrete choice models, there is explosion of alternatives for accommodating joint activities

- We need to determine the entire set of activity purposes pursued and with whom dimension for each of these
For two activity purposes A1 and A2, the possible activity participation:
- None
- A1 only
- A2 only
- A1 and A2

So on the activity dimension \(2^2 = 2^{^\text{ActNo}}\)

For 2 persons P1, P2, the possible combinations:
- P1 alone
- P2 alone
- P1 and P2

So on the person dimension \(2^2 - 1 = (2^{^\text{PerNo}} - 1)\)

For each additional person combination we will have \(2^{^\text{ActNo}}\) repeating \((2^p - 1)\) times

Now for 2 person and 2 activities we have:
- \(2^2 \times 2^2 \times 2^2\)
For A activity purposes, P household members the number of alternatives is given by

\[ 2^A \times 2^A \times 2^A \ldots \times (2^P - 1) \text{ times} \]

\[ = 2^{A(2^P - 1)} \]
Example 2 persons and 2 activity purposes – Single Discrete Case

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P1 P2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>A1</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>A2</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>A1 A2</td>
<td>A1</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Each box represents an alternative
Example 2 persons and 2 activity purposes – Multiple Discrete Case

- Total 7 alternatives versus 64 in traditional case
## Total choice set size comparison for 3 activity purposes

<table>
<thead>
<tr>
<th>Household Size</th>
<th>Single Discrete Model (MNL)</th>
<th>MDCEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>512</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>2097152</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>$3.52 \times 10^{13}$</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>$9.9 \times 10^{27}$</td>
<td>93</td>
</tr>
<tr>
<td>Total</td>
<td>$9.9 \times 10^{27}$</td>
<td>171</td>
</tr>
</tbody>
</table>

Once the number of activities increases the difference will be even stark!
MDCEV in Activity-Based Model

- Currently, most activity based models accommodate activity type choice as a series of activity type specific binary logit models for each individual in the household.

- These approaches do not explicitly recognize that activity participation is a collective decision of household members.

- MDCEV approach, because of its simplicity and relatively inexpensive computational requirement, facilitates modeling activity participation at a household level with joint activity participation incorporated in a simple fashion.

- CEMDAP (within SimAGENT) now features MDCEV for activity participation.
## Comparison of some models

<table>
<thead>
<tr>
<th>Model Aspect</th>
<th>SF-CHAMP</th>
<th>SACSIM</th>
<th>MORPC</th>
<th>CEMDAP</th>
<th>SimAGENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPO</td>
<td>San Francisco County Transportation Authority</td>
<td>Sacramento Area Council of Governments</td>
<td>Mid-Ohio Regional Planning Commission</td>
<td>North Central Texas Council of Governments</td>
<td>Southern California Association of Governments (SCAG)</td>
</tr>
<tr>
<td>Region</td>
<td>San Francisco County, CA</td>
<td>Sacramento, CA</td>
<td>Columbus, Ohio</td>
<td>Dallas Fort-Worth, TX</td>
<td>Los Angeles, CA</td>
</tr>
<tr>
<td>Population in base year</td>
<td>0.3 Million Households 0.8 Million Individuals</td>
<td>0.7 Million Households 1.8 Million Individuals</td>
<td>0.6 Million Households 1.4 Million Individuals</td>
<td>1.8 Million Households 4.8 Million Individuals</td>
<td>5.6 Million Households 17.6 Million Individuals</td>
</tr>
</tbody>
</table>
## Data: Summary of Reviewed Models

<table>
<thead>
<tr>
<th>Model Aspect</th>
<th>SF-CHAMP</th>
<th>SACSIM</th>
<th>MORPC</th>
<th>CEMDAP</th>
<th>SimAGENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model estimation data</td>
<td>1990 SF Bay Area Household Travel Survey Data of 1100 HHs on SF County, stated preference survey of 609HHs for transit related Preferences</td>
<td>Household activity diary survey</td>
<td>1999 Household travel survey data of 5500 HHs in the Columbus region, on-board transit survey data</td>
<td>1996 Household travel survey data of 3500 households in DFW, on-board transit survey data</td>
<td>California Department of Finance (DOF) E-5 Population and Housing Estimates; California Employment Development Department (EDD) 2005 Benchmark</td>
</tr>
<tr>
<td>Network zones (TAZs)</td>
<td>1,900</td>
<td>1,300 (as well as parcels)</td>
<td>2,000 (w/ 3 transit access zones in each zone)</td>
<td>4784</td>
<td>4192 (as well as parcels)</td>
</tr>
<tr>
<td>Network time periods</td>
<td>5 per day</td>
<td>4 per day</td>
<td>5 per day</td>
<td>5 per day</td>
<td>4 per day</td>
</tr>
<tr>
<td>Predicted time periods</td>
<td>5 per day</td>
<td>30 min</td>
<td>1 hour</td>
<td>continuous time (1 min)</td>
<td>continuous time (1 min)</td>
</tr>
</tbody>
</table>
For each child not undertaking joint discretionary activity
Decision to undertake independent discretionary activity
(model GA16)

Decision of household to undertake grocery shopping
(model GA17)

For every household model activity participation using MDCEV Model

How MDCEV alters CEMDAP

For each adult
Activity allocated to the single adult

For each adult
Decision to undertake shopping given that household undertakes grocery shopping
(model GA18)

For each adult
Decision to undertake personal/household business activities
(model GA19)

For each adult
Decision to undertake social/recreational activities
(model GA20)

For each adult
Decision to undertake eat-out activities
(model GA21)

For each adult
Decision to undertake other serve-passenger activities
(model GA22)
Incorporating joint activity alters travel scheduling process substantially.

MDCEV provides us the household members for joint activity.

Need to ensure spatial and temporal consistency among the joint activity participants.

Determining when and where the joint activity is pursued forms an additional pin around which individual travel is scheduled.

Currently we follow the following precedence:
- Children travel needs
- Commuter travel needs
- Joint travel (excluding children travel needs)
- Individual travel
Drop-off child at School

Travel from home to school zone
Activity duration at stop = 5 minutes

Does non-worker participate in joint activity?

Yes

Is travel to joint activity joint or separate?

Joint

If joint activity start time – (current time + travel time to home from school) > 15

Yes

Go Home

No

Go to Joint Activity Location

Separate

If (joint activity start time) – (current time + travel time to joint activity location) > 15

Yes

No

Go Home

Does non-worker undertake independent activities?

No

Go Home

Yes

Number of stops in the work to home commute (model WSCH2)

No Stops

Go Home

Stops Module

Illustration of Non-work Drop-off tour

If “no joint activity” or “joint travel” or “(joint activity start time) – (current time + travel time to joint activity location) > 15”

Yes

Go to Joint Activity Location

No

Go Home
Summary

- Several multiple discrete continuous choice contexts can be modeled using MDCEV.

- MDCEV is an effective tool to address the computational and behavioral challenges to model activity participation (while incorporating joint activity participation seamlessly).

- The computational advantages are evident based on the numbers provided.

- CEMDAP (within SimAGENT), in its latest version, will feature MDCEV.
A NEW ESTIMATION APPROACH FOR DISCRETE CHOICE MODELING SYSTEMS
Simulation techniques:

- Maximum Simulated Likelihood (MSL) Approach
  
  *Approach gets imprecise, develops convergence problems, and becomes computationally expensive with the increase in the number of ordered-response outcomes*

- Bayesian Inference Approach
  
  *Unfortunately, the method remains cumbersome, requires extensive simulation, and is time-consuming*

- Overall, simulation-based approaches become impractical or even infeasible as the number of ordered-response outcomes increases
Another solution to such problems is the use of the Composite Marginal Likelihood (CML) approach. The CML approach...

- Belongs to the more general class of composite likelihood function approaches
- Is based on forming a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods
- Represents a conceptually and pedagogically simpler simulation-free procedure relative to simulation techniques
- Can be applied using simple optimization software for likelihood estimation
- Is typically more robust, and has the advantage of reproducibility of the results
We demonstrate the use of the CML approach in a pairwise marginal likelihood setting

- The pairwise marginal likelihood is formed by the product of likelihood contributions of all subset of couplets (i.e., pairs of variables or pairs of observations)

- Under the usual regularity assumptions, the CML (and hence, the pairwise marginal likelihood) estimator is consistent and asymptotically normal distributed

- This is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood
Motivation

- Compare the performance of the MSL approach with the CML approach when the MSL approach is feasible.

- Undertake a comparison in the context of an ordered-response setting with different correlation structures and with:
  - cross-sectional data, and
  - panel data.

- Use simulated data sets to evaluate the two estimation approaches.

- Examine the performance of the MSL and CML approaches in terms of:
  - The ability of the two approaches to recover model parameters,
  - Relative efficiency, and
  - Non-convergence and computational cost.
Ordered response model systems are used when analyzing ordinal discrete outcome data that may be considered as manifestations of an underlying scale that is endowed with a natural ordering.

Examples include:

- Ratings data (of consumer products, bonds, credit evaluation, movies, etc.)
- Likert-scale type attitudinal/opinion data (of air pollution levels, traffic congestion levels, school academic curriculum satisfaction levels, teacher evaluations, etc.)
- Grouped data (such as bracketed income data in surveys or discretized rainfall data)
- Count data (such as the number of trips made by a household, the number of episodes of physical activity pursued by an individual, and the number of cars owned by a household)
Focus on Ordered Response Systems

- There is an abundance of applications of the ordered-response model in the literature
  - Examples include applications in the sociological, biological, marketing, and transportation sciences

- Mostly one outcome variable, though there have been some applications with 2 to 3 outcome variables

- However, the examination of more than three correlated outcomes is rare because of difficulty associated with medium-to-high dimensional integration
Focus on Ordered Response Systems

- Cross-sectional examples of multiple outcome variables
  - Number of episodes of each of several activities
  - Satisfaction levels associated with a related set of products/services
  - Multiple ratings measures regarding the state of health of an individual/organization

- Time-series or panel examples of multiple outcome variables
  - Rainfall levels (measured in grouped categories) over time in each of several spatial regions
  - Individual stop-making behavior over multiple days in a week
  - Individual headache severity levels at different points in time
Econometric Framework

Multivariate Ordered-Response Model System - Cross-Sectional Formulation (CMOP Model)

\[ y_{qi}^* = \beta_i' x_{qi} + \epsilon_{qi}, y_{qi} = m_{qi} \text{ if } \theta_{i}^{m_{qi}-1} < y_{qi}^* < \theta_{i}^{m_{qi}} \]

\[ y_{qi}^* = \text{The latent propensity of individual } q \text{ for ordered-response variable } i \]

\[ x_{qi} = \text{A } (L \times 1) \text{ vector of exogenous variables (not including a constant)} \]

\[ \beta_i = \text{A corresponding } (L \times 1) \text{ vector of coefficients (to be estimated)} \]

\[ \epsilon_{qi} = \text{A standard normal error term,} \]

\[ y_{qi} = \text{“Observed” count value for individual } q \text{ for variable } i \]

\[ \theta_{i}^{m_{qi}-1} = \text{The lower bound threshold for discrete level } m_{qi} \text{ of variable } i, \text{ and} \]

\[ \theta_{i}^{m_{qi}} = \text{The upper bound threshold for discrete level } m_{qi} \text{ of variable } i \]
Econometric Framework

\[ \varepsilon_q \sim N \left( \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \rho_{11} \\ \rho_{12} \\ \rho_{13} \\ \cdots \\ \rho_{1I} \\ \rho_{21} \\ \rho_{22} \\ \rho_{23} \\ \cdots \\ \rho_{2I} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \rho_{I1} \\ \rho_{I2} \\ \rho_{I3} \\ \cdots \\ 1 \end{pmatrix} \right) \]
Econometric Framework

$$\delta = (\beta_1', \beta_2', \ldots, \beta_I'; \theta_1', \theta_2', \ldots, \theta_I'; \Omega')',$$

$$\theta_i = (\theta_i^1, \theta_i^2, \ldots, \theta_i^K_i)'$$

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

$$L_q (\delta) = \int_{v_1=\theta_i^{m,q_1}-\beta_i' x_{q1}}^{\theta_i^{m,q_1+1}-\beta_i' x_{q1}} \int_{v_2=\theta_i^{m,q_2}-\beta_i' x_{q2}}^{\theta_i^{m,q_2+1}-\beta_i' x_{q2}} \cdots \int_{v_I=\theta_i^{m,q_I}-\beta_i' x_{qI}}^{\theta_i^{m,q_I+1}-\beta_i' x_{qI}} \phi_I(v_1, v_2, \ldots, v_I | \Omega) dv_1 dv_2 \ldots dv_I$$
Econometric Framework

Multivariate Ordered-Response Model System - Panel Formulation (PMOP Model)

\[ y_{qj}^* = \beta' x_{qj} + u_q + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \theta^{m_{qj}-1} < y_{qj}^* < \theta^{m_{qj}} \]

\( y_{qj}^* \) = The latent propensity of individual \( q \) for \( j \)th observation

\( \beta \) = A corresponding \((L \times 1)\) vector of coefficients (to be estimated)

\( x_{qj} \) = A \((L \times 1)\) vector of exogenous variables (not including a constant)

\( u_q \) = An individual-specific random term, \( u_q \sim N(0, \sigma^2) \)

\( \varepsilon_{qj} \) = A standard normal error term uncorrelated across individuals \( q \), but serially correlated across observations \( j \) for individual \( q \)

\( y_{qj} \) = “Observed” count value for individual \( q \) for variable \( j \)

\( \theta^{m_{qj}-1} \) = The lower bound threshold for discrete level \( m_{qj} \)

\( \theta^{m_{qj}} \) = The upper bound threshold for discrete level \( m_{qj} \)
Econometric Framework

The joint distribution of the latent variables \((y_{q1}^*, y_{q2}^*, \ldots, y_{qJ}^*)\) for the \(q\)th subject is multivariate normal with standardized mean vector \((β'x_{q1} / μ, β'x_{q2} / μ, \ldots, β'x_{qJ} / μ)\) and a correlation matrix with constant non-diagonal entries \(σ^2 / μ^2\), where \(μ = \sqrt{1 + σ^2}\)

\[
\delta = (β', θ^1, θ^2, \ldots, θ^{K-1}; σ, ρ)'
\]

\[
L_q(δ) = Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \ldots, y_{qJ} = m_{qJ})
\]

\[
L_q(δ) = \int_{v_1 = α_{mq1}^{-1}}^{α_{mq1}} \int_{v_2 = α_{mq2}^{-1}}^{α_{mq2}} \cdots \int_{v_J = α_{mqJ}^{-1}}^{α_{mqJ}} φ_J(v_1, v_2, \ldots, v_J \mid R_q)dv_1dv_2\ldots dv_J
\]

where \(α_{mqj} = (θ_{mqj} - β'x_{qj}) / μ\).
Estimation Approaches

- Simulation Approaches
  - The Frequentist Approach – Maximum Simulated Likelihood (MSL) Method
  - The Bayesian Approach
  - Simulators Used in the Current Paper
    - *The GHK Probability Simulator for the CMOP Model*
    - *The GB Simulator for the PMOP Model*

- The CML Technique – The Pairwise Marginal Likelihood Inference Approach
  - Pairwise Likelihood Approach
    - *The CMOP Model*
    - *The PMOP Model*
    - Positive-Definiteness of the Correlation Matrix
Estimation Approaches

The GHK Probability Simulator for the CMOP Model

- Named after John F. Geweke, Vassilis A. Hajivassiliou, and Michael P. Keane

- The GHK is perhaps the most widely used probability simulator for integration of the multivariate normal density function

- The simulator is based on directly approximating the probability of a multivariate rectangular region of the multivariate normal density distribution
The GHK Probability Simulator for the CMOP Model (cont.)

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

\[ L_q(\delta) = \Pr(y_{q1} = m_{q1}) \Pr(y_{q2} = m_{q2} \mid y_{q1} = m_{q1}) \Pr(y_{q3} = m_{q3} \mid y_{q1} = m_{q1}, y_{q2} = m_{q2}) ... \]
\[ ... \Pr(y_{ql} = m_{ql} \mid y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{ql-1} = m_{ql-1}) \]

Also,

\[ \begin{bmatrix} \varepsilon_{q1} \\ \varepsilon_{q2} \\ \vdots \\ \varepsilon_{ql} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{l1} & l_{l2} & l_{l3} & \cdots & l_{ll} \end{bmatrix} \begin{bmatrix} v_{q1} \\ v_{q2} \\ \vdots \\ v_{ql} \end{bmatrix} \]

\[ \varepsilon_q = Lv_q \]
Estimation Approaches

The GHK Probability Simulator for the CMOP Model (cont.)

- $L$ is the lower triangular Cholesky decomposition of the correlation matrix $\Sigma$, and $\nu_q$ terms are independent and identically distributed standard normal deviates.

- $\nu_q$ are drawn $d$ times ($d = 1, 2, \ldots, 100$) from the univariate standard normal distribution with pre-specified lower and upper bounds.

- We use a randomized Halton draw procedure to generate the $d$ realizations.

- The positive definiteness of $\Sigma$ was ensured by parameterizing the likelihood function with the elements of $L$. 
The GB Simulator for the PMOP Model

- Named after Alan Genz and Frank Bretz
- Provides an alternative simulation-based approximation of multivariate normal probabilities
- The approach involves
  - Transforming the original hyper-rectangle integral region to an integral over a unit hypercube
  - Filling the transformed integral region by randomized lattice points
  - Deriving robust integration error bounds by means of additional shifts of the integration nodes in random directions
- The positive-definite correlation matrix is ensured by defining the parameter spaces, so that $\sigma > 0$ and $0 < \rho < 1$
Estimation Approaches

Pairwise Likelihood Approach for the CMOP Model

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

\[
= \prod_{i=1}^{I-1} \prod_{g=i+1}^{I} \left[ \Phi_2 \left( m_{qi} - \beta_i'x_{qi}, \theta_g^{m_{qg}} - \beta_g'x_{qg}, \rho_{ig} \right) - \Phi_2 \left( m_{qi} - \beta_i'x_{qi}, \theta_g^{m_{qg}-1} - \beta_g'x_{qg}, \rho_{ig} \right) \right] - \Phi_2 \left( m_{qi} - \beta_i'x_{qi}, \theta_g^{m_{qg}-1} - \beta_g'x_{qg}, \rho_{ig} \right) + \Phi_2 \left( m_{qi} - \beta_i'x_{qi}, \theta_g^{m_{qg}-1} - \beta_g'x_{qg}, \rho_{ig} \right)
\]

\[
L_{CML}^{CMOP}(\delta) = \prod_{q} L_{CML,q}^{CMOP}(\delta)
\]
Estimation Approaches

Pairwise Likelihood Approach for the PMOP Model

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

\[ = \prod_{j=1}^{J-1} \prod_{g=j+1}^{J} \left[ \Phi_2 \left( \chi_{qj}^{m_{qj}}, \alpha_{qg}^{m_{qg}}, \rho_{jg} \right) - \Phi_2 \left( \chi_{qj}^{m_{qj}^{-1}}, \alpha_{qg}^{m_{qg}^{-1}}, \rho_{jg} \right) - \Phi_2 \left( \chi_{qj}^{m_{qj}^{-1}}, \alpha_{qg}^{m_{qg}^{-1}}, \rho_{jg} \right) + \Phi_2 \left( \chi_{qj}^{m_{qj}}, \alpha_{qg}^{m_{qg}}, \rho_{jg} \right) \right] \]

\[ L_{CML}^{PMOP}(\delta) = \prod_{q} L_{CML, q}^{PMOP}(\delta) \]

Where,
\[ \alpha_{qj}^{m_{qj}} = (\theta_{qj}^{m_{qj}} - \beta'x_{qj}) / \mu, \quad \mu = \sqrt{1 + \sigma^2}, \quad \text{and} \quad \rho_{jg} = (\sigma^2 + \rho_{\left| t_{qj} - t_{qg} \right|}) / \mu^2 \]
Experimental Design

The CMOP Model

- Multivariate ordered response system with five ordinal variables
  - Low error correlation structure ($\Sigma_{low}$)
  - High error correlation structure ($\Sigma_{high}$)
- $\delta$ vector with pre-specified values
- 20 independent data sets with 1000 data points
  - The GHK simulator is applied to each data set
    - Using 100 draws per individual of the randomized Halton sequence
    - 10 times with different (independent) randomized Halton draw sequences
The PMOP Model

- Multivariate ordered response system with six ordinal variables
  - Low autoregressive correlation parameter ($\rho=0.3$)
  - High autoregressive correlation parameter ($\rho=0.7$)
- $\delta$ vector with pre-specified values
- 100 independent data sets with 200 subjects and 6 “observation” per subject
  - The GB simulator is applied to each data set
    - 10 times with different (independent) random draw sequences
    - With an absolute error tolerance of 0.001
Performance Measures

- **Parameter Estimates**
  - Mean estimate
  - Absolute percentage bias

- **Standard Error Estimates**
  - Finite sample standard error
  - Asymptotic standard error
  - In addition, for MSL approach we estimated:
    - Simulation standard error
    - Simulation adjusted standard error
Performance Measures

- **Relative Efficiencies**
  - Ratio between the MSL and CML asymptotic standard errors
  - Ratio between the simulation adjusted standard error and the CML asymptotic standard error

- **Non-convergence Rates**

- **Relative Computational Time Factor (RCTF)**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
<td>Asymptotic SE (A)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.5000</td>
<td>0.5167</td>
<td>3.34%</td>
<td>0.0481</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>1.0000</td>
<td>1.0077</td>
<td>0.77%</td>
<td>0.0474</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.2500</td>
<td>0.2501</td>
<td>0.06%</td>
<td>0.0445</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.7500</td>
<td>0.7461</td>
<td>0.52%</td>
<td>0.0641</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>1.0000</td>
<td>0.9984</td>
<td>0.16%</td>
<td>0.0477</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>0.5000</td>
<td>0.4884</td>
<td>2.31%</td>
<td>0.0413</td>
</tr>
<tr>
<td>$\beta_{42}$</td>
<td>0.2500</td>
<td>0.2605</td>
<td>4.19%</td>
<td>0.0372</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>0.2500</td>
<td>0.2445</td>
<td>2.21%</td>
<td>0.0401</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>0.5000</td>
<td>0.4967</td>
<td>0.66%</td>
<td>0.0420</td>
</tr>
<tr>
<td>$\beta_{33}$</td>
<td>0.7500</td>
<td>0.7526</td>
<td>0.34%</td>
<td>0.0348</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>0.7500</td>
<td>0.7593</td>
<td>1.24%</td>
<td>0.0530</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>0.2500</td>
<td>0.2536</td>
<td>1.46%</td>
<td>0.0420</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>1.0000</td>
<td>0.9976</td>
<td>0.24%</td>
<td>0.0832</td>
</tr>
<tr>
<td>$\beta_{44}$</td>
<td>0.3000</td>
<td>0.2898</td>
<td>3.39%</td>
<td>0.0481</td>
</tr>
<tr>
<td>$\beta_{15}$</td>
<td>0.4000</td>
<td>0.3946</td>
<td>1.34%</td>
<td>0.0333</td>
</tr>
<tr>
<td>$\beta_{25}$</td>
<td>1.0000</td>
<td>0.9911</td>
<td>0.89%</td>
<td>0.0434</td>
</tr>
<tr>
<td>$\beta_{35}$</td>
<td>0.6000</td>
<td>0.5987</td>
<td>0.22%</td>
<td>0.0322</td>
</tr>
</tbody>
</table>
Simulation Results – The CMOP Model ($\Sigma_{low}$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>0.3000</td>
<td>0.2857</td>
<td>4.76%</td>
<td>0.0496</td>
</tr>
<tr>
<td>$\rho_{13}$</td>
<td>0.2000</td>
<td>0.2013</td>
<td>0.66%</td>
<td>0.0477</td>
</tr>
<tr>
<td>$\rho_{14}$</td>
<td>0.2200</td>
<td>0.1919</td>
<td>12.76%</td>
<td>0.0535</td>
</tr>
<tr>
<td>$\rho_{15}$</td>
<td>0.1500</td>
<td>0.1739</td>
<td>15.95%</td>
<td>0.0388</td>
</tr>
<tr>
<td>$\rho_{23}$</td>
<td>0.2500</td>
<td>0.2414</td>
<td>3.46%</td>
<td>0.0546</td>
</tr>
<tr>
<td>$\rho_{24}$</td>
<td>0.3000</td>
<td>0.2960</td>
<td>1.34%</td>
<td>0.0619</td>
</tr>
<tr>
<td>$\rho_{25}$</td>
<td>0.1200</td>
<td>0.1117</td>
<td>6.94%</td>
<td>0.0676</td>
</tr>
<tr>
<td>$\rho_{34}$</td>
<td>0.2700</td>
<td>0.2737</td>
<td>1.37%</td>
<td>0.0488</td>
</tr>
<tr>
<td>$\rho_{35}$</td>
<td>0.2000</td>
<td>0.2052</td>
<td>2.62%</td>
<td>0.0434</td>
</tr>
<tr>
<td>$\rho_{45}$</td>
<td>0.2500</td>
<td>0.2419</td>
<td>3.25%</td>
<td>0.0465</td>
</tr>
</tbody>
</table>

**Correlation Coefficients**
Simulation Results – The CMOP Model \( \Sigma_{\text{low}} \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
</tr>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>-1.0000</td>
<td>-1.0172</td>
<td>1.72%</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>1.0000</td>
<td>0.9985</td>
<td>0.15%</td>
</tr>
<tr>
<td>( \theta_3 )</td>
<td>3.0000</td>
<td>2.9992</td>
<td>0.03%</td>
</tr>
<tr>
<td>( \theta_4 )</td>
<td>0.0000</td>
<td>-0.0172</td>
<td>-</td>
</tr>
<tr>
<td>( \theta_5 )</td>
<td>2.0000</td>
<td>1.9935</td>
<td>0.32%</td>
</tr>
<tr>
<td>( \theta_6 )</td>
<td>-2.0000</td>
<td>-2.0193</td>
<td>0.97%</td>
</tr>
<tr>
<td>( \theta_7 )</td>
<td>-0.5000</td>
<td>-0.5173</td>
<td>3.47%</td>
</tr>
<tr>
<td>( \theta_8 )</td>
<td>1.0000</td>
<td>0.9956</td>
<td>0.44%</td>
</tr>
<tr>
<td>( \theta_9 )</td>
<td>2.5000</td>
<td>2.4871</td>
<td>0.52%</td>
</tr>
<tr>
<td>( \theta_{10} )</td>
<td>1.0000</td>
<td>0.9908</td>
<td>0.92%</td>
</tr>
<tr>
<td>( \theta_{11} )</td>
<td>3.0000</td>
<td>3.0135</td>
<td>0.45%</td>
</tr>
<tr>
<td>( \theta_{12} )</td>
<td>-1.5000</td>
<td>-1.5084</td>
<td>0.56%</td>
</tr>
<tr>
<td>( \theta_{13} )</td>
<td>0.5000</td>
<td>0.4925</td>
<td>1.50%</td>
</tr>
<tr>
<td>( \theta_{14} )</td>
<td>2.0000</td>
<td>2.0201</td>
<td>1.01%</td>
</tr>
<tr>
<td>Overall mean</td>
<td>-</td>
<td>2.21%</td>
<td></td>
</tr>
</tbody>
</table>

Threshold Parameters
### Simulation Results – The CMOP Model ($\sum_{\text{high}}$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
<td>Asymptotic SE (A)</td>
<td>Simulation SE</td>
<td>Simulation Adj. SE (B)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.5000</td>
<td>0.5063</td>
<td>1.27%</td>
<td>0.0300</td>
<td>0.0294</td>
<td>0.0020</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>1.0000</td>
<td>1.0089</td>
<td>0.89%</td>
<td>0.0410</td>
<td>0.0391</td>
<td>0.0026</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.2500</td>
<td>0.2571</td>
<td>2.85%</td>
<td>0.0215</td>
<td>0.0288</td>
<td>0.0017</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.7500</td>
<td>0.7596</td>
<td>1.27%</td>
<td>0.0495</td>
<td>0.0373</td>
<td>0.0028</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>1.0000</td>
<td>1.0184</td>
<td>1.84%</td>
<td>0.0439</td>
<td>0.0436</td>
<td>0.0036</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>0.5000</td>
<td>0.5099</td>
<td>0.17%</td>
<td>0.0343</td>
<td>0.0314</td>
<td>0.0023</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>0.2500</td>
<td>0.2524</td>
<td>0.96%</td>
<td>0.0284</td>
<td>0.0294</td>
<td>0.0021</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>0.7500</td>
<td>0.7498</td>
<td>0.02%</td>
<td>0.0302</td>
<td>0.0291</td>
<td>0.0019</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>0.7500</td>
<td>0.7580</td>
<td>0.11%</td>
<td>0.0416</td>
<td>0.0419</td>
<td>0.0039</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>0.2500</td>
<td>0.2407</td>
<td>3.70%</td>
<td>0.0311</td>
<td>0.0326</td>
<td>0.0033</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>1.0000</td>
<td>1.0160</td>
<td>1.60%</td>
<td>0.0483</td>
<td>0.0489</td>
<td>0.0041</td>
</tr>
<tr>
<td>$\beta_{44}$</td>
<td>0.3000</td>
<td>0.3172</td>
<td>5.72%</td>
<td>0.0481</td>
<td>0.0336</td>
<td>0.0028</td>
</tr>
<tr>
<td>$\beta_{15}$</td>
<td>0.4000</td>
<td>0.3899</td>
<td>2.54%</td>
<td>0.0279</td>
<td>0.0286</td>
<td>0.0026</td>
</tr>
<tr>
<td>$\beta_{25}$</td>
<td>0.1000</td>
<td>0.9875</td>
<td>1.25%</td>
<td>0.0365</td>
<td>0.0391</td>
<td>0.0036</td>
</tr>
<tr>
<td>$\beta_{35}$</td>
<td>0.6000</td>
<td>0.5923</td>
<td>1.28%</td>
<td>0.0309</td>
<td>0.0316</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

**Coefficients**

The CMOP Model ($\sum_{\text{high}}$)
Simulation Results – The CMOP Model ($\Sigma_{\text{high}}$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Abs. % Bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finite Sample SE</td>
<td>Asymptotic SE (A)</td>
<td>Simulation SE</td>
<td>Simulation Adj. SE (B)</td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finite Sample SE</td>
<td>Asymptotic SE (C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>0.9000</td>
<td>0.8969</td>
<td>0.34%</td>
<td>0.0224</td>
<td>0.0177</td>
<td>0.0034</td>
<td>0.0180</td>
<td>0.9019</td>
<td>0.21%</td>
<td>0.0233</td>
<td>0.0183</td>
<td>0.9669</td>
<td>0.9845</td>
<td></td>
</tr>
<tr>
<td>$\rho_{13}$</td>
<td>0.8000</td>
<td>0.8041</td>
<td>0.51%</td>
<td>0.0174</td>
<td>0.0201</td>
<td>0.0035</td>
<td>0.0204</td>
<td>0.8009</td>
<td>0.11%</td>
<td>0.0195</td>
<td>0.0203</td>
<td>0.9874</td>
<td>1.0023</td>
<td></td>
</tr>
<tr>
<td>$\rho_{14}$</td>
<td>0.8200</td>
<td>0.8249</td>
<td>0.60%</td>
<td>0.0284</td>
<td>0.0265</td>
<td>0.0061</td>
<td>0.0272</td>
<td>0.8151</td>
<td>0.60%</td>
<td>0.0296</td>
<td>0.0297</td>
<td>0.8933</td>
<td>0.9165</td>
<td></td>
</tr>
<tr>
<td>$\rho_{15}$</td>
<td>0.7500</td>
<td>0.7536</td>
<td>0.49%</td>
<td>0.0248</td>
<td>0.0243</td>
<td>0.0046</td>
<td>0.0247</td>
<td>0.7501</td>
<td>0.01%</td>
<td>0.0242</td>
<td>0.0251</td>
<td>0.9678</td>
<td>0.9849</td>
<td></td>
</tr>
<tr>
<td>$\rho_{23}$</td>
<td>0.8500</td>
<td>0.8426</td>
<td>0.87%</td>
<td>0.0181</td>
<td>0.0190</td>
<td>0.0081</td>
<td>0.0207</td>
<td>0.8468</td>
<td>0.38%</td>
<td>0.0190</td>
<td>0.0198</td>
<td>0.9606</td>
<td>1.0438</td>
<td></td>
</tr>
<tr>
<td>$\rho_{24}$</td>
<td>0.9000</td>
<td>0.8842</td>
<td>1.75%</td>
<td>0.0187</td>
<td>0.0231</td>
<td>0.0097</td>
<td>0.0251</td>
<td>0.9023</td>
<td>0.26%</td>
<td>0.0289</td>
<td>0.0244</td>
<td>0.9484</td>
<td>1.0284</td>
<td></td>
</tr>
<tr>
<td>$\rho_{25}$</td>
<td>0.7200</td>
<td>0.7184</td>
<td>0.22%</td>
<td>0.0241</td>
<td>0.0280</td>
<td>0.0072</td>
<td>0.0289</td>
<td>0.7207</td>
<td>0.09%</td>
<td>0.0295</td>
<td>0.0301</td>
<td>0.9298</td>
<td>0.9600</td>
<td></td>
</tr>
<tr>
<td>$\rho_{34}$</td>
<td>0.8700</td>
<td>0.8724</td>
<td>0.27%</td>
<td>0.0176</td>
<td>0.0197</td>
<td>0.0036</td>
<td>0.0200</td>
<td>0.8644</td>
<td>0.65%</td>
<td>0.0208</td>
<td>0.0220</td>
<td>0.8972</td>
<td>0.9124</td>
<td></td>
</tr>
<tr>
<td>$\rho_{35}$</td>
<td>0.8000</td>
<td>0.7997</td>
<td>0.04%</td>
<td>0.0265</td>
<td>0.0191</td>
<td>0.0039</td>
<td>0.0195</td>
<td>0.7988</td>
<td>0.15%</td>
<td>0.0193</td>
<td>0.0198</td>
<td>0.9645</td>
<td>0.9848</td>
<td></td>
</tr>
<tr>
<td>$\rho_{45}$</td>
<td>0.8500</td>
<td>0.8421</td>
<td>0.93%</td>
<td>0.0242</td>
<td>0.0231</td>
<td>0.0128</td>
<td>0.0264</td>
<td>0.8576</td>
<td>0.89%</td>
<td>0.0192</td>
<td>0.0252</td>
<td>0.9156</td>
<td>1.0480</td>
<td></td>
</tr>
</tbody>
</table>

**Correlation Coefficients**

- $\rho_{12}$
- $\rho_{13}$
- $\rho_{14}$
- $\rho_{15}$
- $\rho_{23}$
- $\rho_{24}$
- $\rho_{25}$
- $\rho_{34}$
- $\rho_{35}$
- $\rho_{45}$
Simulation Results – The CMOP Model ($\Sigma_{\text{high}}$)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Parameter Estimates</th>
<th>Standard Error (SE) Estimates</th>
<th>Parameter Estimates</th>
<th>Standard Error (SE) Estimates</th>
<th>( \frac{A}{C} )</th>
<th>( \frac{B}{C} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
<td>Asymptotic SE (A)</td>
<td>Simulation SE</td>
<td>Simulation Adj. SE (B)</td>
</tr>
<tr>
<td>( \theta_1^1 )</td>
<td>-1.0000</td>
<td>-1.0110</td>
<td>1.10%</td>
<td>0.0600</td>
<td>0.0520</td>
<td>0.0023</td>
<td>0.0520</td>
</tr>
<tr>
<td>( \theta_1^2 )</td>
<td>1.0000</td>
<td>0.9907</td>
<td>0.93%</td>
<td>0.0551</td>
<td>0.0515</td>
<td>0.0022</td>
<td>0.0515</td>
</tr>
<tr>
<td>( \theta_1^3 )</td>
<td>3.0000</td>
<td>3.0213</td>
<td>0.71%</td>
<td>0.0819</td>
<td>0.1177</td>
<td>0.0065</td>
<td>0.1179</td>
</tr>
<tr>
<td>( \theta_2^1 )</td>
<td>0.0000</td>
<td>-0.0234</td>
<td>-</td>
<td>0.0376</td>
<td>0.0435</td>
<td>0.0028</td>
<td>0.0436</td>
</tr>
<tr>
<td>( \theta_2^2 )</td>
<td>2.0000</td>
<td>2.0089</td>
<td>0.44%</td>
<td>0.0859</td>
<td>0.0781</td>
<td>0.0066</td>
<td>0.0784</td>
</tr>
<tr>
<td>( \theta_3^1 )</td>
<td>-2.0000</td>
<td>-2.0266</td>
<td>1.33%</td>
<td>0.0838</td>
<td>0.0754</td>
<td>0.0060</td>
<td>0.0757</td>
</tr>
<tr>
<td>( \theta_3^2 )</td>
<td>-0.5000</td>
<td>-0.5086</td>
<td>1.73%</td>
<td>0.0305</td>
<td>0.0440</td>
<td>0.0030</td>
<td>0.0441</td>
</tr>
<tr>
<td>( \theta_3^3 )</td>
<td>1.0000</td>
<td>0.9917</td>
<td>0.83%</td>
<td>0.0516</td>
<td>0.0498</td>
<td>0.0035</td>
<td>0.0499</td>
</tr>
<tr>
<td>( \theta_4^1 )</td>
<td>2.5000</td>
<td>2.4890</td>
<td>0.44%</td>
<td>0.0750</td>
<td>0.0928</td>
<td>0.0066</td>
<td>0.0930</td>
</tr>
<tr>
<td>( \theta_4^2 )</td>
<td>1.0000</td>
<td>0.9976</td>
<td>0.24%</td>
<td>0.0574</td>
<td>0.0540</td>
<td>0.0050</td>
<td>0.0542</td>
</tr>
<tr>
<td>( \theta_5^1 )</td>
<td>3.0000</td>
<td>3.0101</td>
<td>0.34%</td>
<td>0.1107</td>
<td>0.1193</td>
<td>0.0125</td>
<td>0.1200</td>
</tr>
<tr>
<td>( \theta_5^2 )</td>
<td>-1.5000</td>
<td>-1.4875</td>
<td>0.84%</td>
<td>0.0694</td>
<td>0.0629</td>
<td>0.0056</td>
<td>0.0632</td>
</tr>
<tr>
<td>( \theta_5^3 )</td>
<td>0.5000</td>
<td>0.4822</td>
<td>3.55%</td>
<td>0.0581</td>
<td>0.0465</td>
<td>0.0041</td>
<td>0.0467</td>
</tr>
<tr>
<td>( \theta_5^4 )</td>
<td>2.0000</td>
<td>1.9593</td>
<td>2.03%</td>
<td>0.0850</td>
<td>0.0741</td>
<td>0.0064</td>
<td>0.0744</td>
</tr>
<tr>
<td>Overall mean value across parameters</td>
<td>-</td>
<td>1.22%</td>
<td>0.0429</td>
<td>0.0428</td>
<td>0.0044</td>
<td>0.0432</td>
<td>-</td>
</tr>
</tbody>
</table>
### Simulation Results – The PMOP Model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Abs. % Bias</td>
<td>Finite Sample SE</td>
<td>Asymptotic SE (A)</td>
</tr>
<tr>
<td>$\rho = 0.30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.0000</td>
<td>0.9899</td>
<td>1.01%</td>
<td>0.1824</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.0000</td>
<td>1.0093</td>
<td>0.93%</td>
<td>0.1729</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.3000</td>
<td>0.2871</td>
<td>4.29%</td>
<td>0.0635</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.0000</td>
<td>1.0166</td>
<td>1.66%</td>
<td>0.2040</td>
</tr>
<tr>
<td>$\theta^1$</td>
<td>1.5000</td>
<td>1.5060</td>
<td>0.40%</td>
<td>0.2408</td>
</tr>
<tr>
<td>$\theta^2$</td>
<td>2.5000</td>
<td>2.5129</td>
<td>0.52%</td>
<td>0.2617</td>
</tr>
<tr>
<td>$\theta^3$</td>
<td>3.0000</td>
<td>3.0077</td>
<td>0.26%</td>
<td>0.2670</td>
</tr>
<tr>
<td>Overall mean value across parameters</td>
<td>-</td>
<td>1.29%</td>
<td>0.1989</td>
<td>0.2109</td>
</tr>
</tbody>
</table>

| $\rho = 0.70$ | | | | | | | | | | | | | | | | |
| $\beta_1$ | 1.0000 | 1.0045 | 0.45% | 0.2338 | 0.2267 | 0.0001 | 0.2267 | 1.0041 | 0.41% | 0.2450 | 0.2368 | 0.9572 | 0.9572 |
| $\beta_2$ | 1.0000 | 1.0183 | 1.83% | 0.1726 | 0.1812 | 0.0001 | 0.1812 | 1.0304 | 3.04% | 0.1969 | 0.2199 | 0.8239 | 0.8239 |
| $\rho$ | 0.7000 | 0.6854 | 2.08% | 0.0729 | 0.0673 | 0.0001 | 0.0673 | 0.6848 | 2.18% | 0.0744 | 0.0735 | 0.9159 | 0.9159 |
| $\sigma^2$ | 1.0000 | 1.0614 | 6.14% | 0.4634 | 0.4221 | 0.0004 | 0.4221 | 1.0571 | 5.71% | 0.4864 | 0.4578 | 0.9220 | 0.9220 |
| $\theta^1$ | 1.5000 | 1.5192 | 1.28% | 0.2815 | 0.2749 | 0.0002 | 0.2749 | 1.5304 | 2.03% | 0.3101 | 0.3065 | 0.8968 | 0.8968 |
| $\theta^2$ | 2.5000 | 2.5325 | 1.30% | 0.3618 | 0.3432 | 0.0003 | 0.3432 | 2.5433 | 1.73% | 0.3904 | 0.3781 | 0.9076 | 0.9076 |
| $\theta^3$ | 3.0000 | 3.0392 | 1.31% | 0.4033 | 0.3838 | 0.0003 | 0.3838 | 3.0514 | 1.71% | 0.4324 | 0.4207 | 0.9123 | 0.9123 |
| Overall mean value across parameters | - | 2.06% | 0.2842 | 0.2713 | 0.0002 | 0.2713 | - | 2.40% | 0.3051 | 0.2990 | 0.9051 | 0.9051 |
Simulation Results

- Non-convergence rates
  - The CMOP Model
    - Low correlation case: 28.5%
    - High correlation case: 35.5%
  - The PMOP Model
    - Low correlation case: 4.2%
    - High correlation case: 2.4%
Simulation Results

Relative Computational Time Factor (RCTF)

- The CMOP Model
  - Low correlation case: 18
  - High correlation case: 40

- The PMOP Model
  - Low correlation case: 332
  - High correlation case: 231
Summary and Conclusions

- Compared the performance of the MSL approach with the CML approach in multivariate ordered-response situations
  - Cross-sectional setting, and
  - Panel setting

- Simulation data sets with known parameter vectors were used

- The results indicate that the CML approach recovers parameters as well as the MSL estimation approach
  - In addition, the ability of the CML approach to recover the parameters seems to be independent of the correlation structure
The CML approach recovers parameters at a substantially reduced computational cost and improved computational stability.

Any reduction in the efficiency of the CML approach relative to the MSL approach is in the range of non-existent to small.
“Workhorse” multinomial logit is saddled with the problem of IIA

Several ways to relax the IID assumption
- Multinomial Probit
- GEV class of models
- Mixed MNL

Mixed MNL models are conceptually appealing

These methods employ simulation based approaches to tackle integration within the likelihood function.
- Accuracy of simulation techniques degrades rapidly at medium-to-high dimensions, and simulation noise increases convergence problems

Impractical in terms of computation time, or even infeasible, as the number of alternatives grows in the multinomial choice situation
Problem at Hand

- Consider a random utility formulation in which the utility that an individual $q$ associates with alternative $i$ ($i = 1, 2, ..., I$) is written as:

$$U_{qi} = \beta'x_{qi} + \varepsilon_{qi}$$

- The probability of choosing alternative $m$

$$P_{qm} = \operatorname{Prob}[U_{qm} > U_{qi} \forall i \neq m] = \operatorname{Prob}[\beta'x_{qm} + \varepsilon_{qm} > \beta'x_{qi} + \varepsilon_{qi} \forall i \neq m]$$

- Alternatively, $P_{qm} = \operatorname{Prob}[y_{qim}^* < 0 \forall i \neq m]$, where

$$y_{qim}^* = U_{qi} - U_{qm} = \beta'z_{qim} + \eta_{qim}, \quad z_{qim} = (x_{qi} - x_{qm}) \quad \text{and} \quad \eta_{qim} = (\varepsilon_{qi} - \varepsilon_{qm})$$
## Simulation Exercise – Cross-sectional MNP

Number of Runs : 50

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MNP MSL (150 Halton draws)</th>
<th>MNP MOPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>Mean Estimate</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.5000</td>
<td>1.3492</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-1.0000</td>
<td>-0.8924</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>2.0000</td>
<td>1.7869</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>1.0000</td>
<td>0.8977</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>2.0000</td>
<td>-1.7977</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>1.0000</td>
<td>0.8929</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>1.0000</td>
<td>0.8899</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>1.0000</td>
<td>0.8756</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>1.0000</td>
<td>0.8816</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>1.0000</td>
<td>0.8952</td>
</tr>
</tbody>
</table>
Simulation Exercise – Cross-sectional MNP

Number of Runs : 50

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MNP MSL (150 Scrambled Halton draws)</th>
<th>MNP MOPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>Mean Estimate</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.5000</td>
<td>1.3395</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-1.0000</td>
<td>-0.8866</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>2.0000</td>
<td>1.7731</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>1.0000</td>
<td>0.8900</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>2.0000</td>
<td>-1.7830</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>1.0000</td>
<td>0.8837</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>1.0000</td>
<td>0.8814</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>1.0000</td>
<td>0.8729</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>1.0000</td>
<td>0.8680</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>1.0000</td>
<td>0.8927</td>
</tr>
</tbody>
</table>
All these factors, combined with the conceptual and implementation simplicity of our approach, makes the approach promising.
Thank You

Web Site:
http://www.ce.utexas.edu/prof/bhat