Estimating Mixed Logit Models with Large Choice Sets

Roger H. von Haefen, NC State & NBER
Adam Domanski, NOAA
July 2013
Motivation

• Bayer et al. (JPE, 2007)
  – Sorting modeling / housing choice
  – 250,000 individuals / alternatives
  – Estimate conditional logit model
    • Why? Sampling of alternatives
    • But restrictive substitution patterns
Research Objectives

• Develop estimation strategy for mixed logit models applied to large choice set problems
  – Estimate latent class models with variation of the Expectation-Maximization (EM) algorithm
  – Quantify the efficiency/bias/run time tradeoffs in an outdoor recreation application
Outline

• Background
• Latent class models
• EM algorithm
• Simulations
• Application
• Future directions
Discrete Choice Analysis

- Choice from a large set of alternatives
Discrete Choice Analysis

- Conditional indirect utility:

\[ U_{ij} = X_{ij}\beta + \varepsilon_{ij} \]
Discrete Choice Analysis

- Decision rule:
  - Alternative $j$ chosen iff:
    
    $$U_{ij} = \max \{U_{i1}, \ldots, U_{iJ}\}$$
Discrete Choice Analysis

- Conditional Logit Model (McFadden 1974)
  - Assuming $\varepsilon_{ij}$ is iid type I extreme value, then:

$$P_{ij} = \frac{\exp(X_{ij}\beta)}{\sum_{j'}\exp(X_{ij'}\beta)}$$
Discrete Choice Analysis

- Independence of Irrelevant Alternatives (IIA)
  \[
  \frac{P_{ij}}{P_{ik}} = \frac{\exp(X_{ij}\beta)}{\exp(X_{ik}\beta)}
  \]

- Restrictive substitution patterns
Computational Challenges w/ Large Choice Sets

Three approaches:

• Aggregation

• Separability

• Sampling
Sampling of Alternatives

Ex: Five individuals, 15 alternatives

Chosen alternative in red

Full sample

<table>
<thead>
<tr>
<th>Individual</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 <strong>14</strong> 15</td>
</tr>
<tr>
<td>B</td>
<td>1 2 3 4 5 <strong>6</strong> 7 8 9 10 11 12 13 <strong>14</strong> 15</td>
</tr>
<tr>
<td>C</td>
<td>1 2 3 4 5 6 7 8 <strong>9</strong> 10 11 12 13 14 15</td>
</tr>
<tr>
<td>D</td>
<td>1 2 3 4 5 6 <strong>7</strong> 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>E</td>
<td>1 2 3 <strong>4</strong> 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
</tbody>
</table>
### Sampling of Alternatives

Ex: Five individuals, 15 alternatives

Chosen alternative in red

50% sample

<table>
<thead>
<tr>
<th>Individual</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>B</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>C</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>D</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>E</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
</tbody>
</table>
# Sampling of Alternatives

Ex: Five individuals, 15 alternatives

Chosen alternative in **red**

50% sample

<table>
<thead>
<tr>
<th>Individual</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 3 7 8 11 12 14 15</td>
</tr>
<tr>
<td>B</td>
<td>2 5 6 7 9 10 13 14</td>
</tr>
<tr>
<td>C</td>
<td>1 5 6 8 9 11 12 13</td>
</tr>
<tr>
<td>D</td>
<td>3 4 5 7 8 9 10 13</td>
</tr>
<tr>
<td>E</td>
<td>2 3 4 6 7 10 12 15</td>
</tr>
</tbody>
</table>
Sampling of Alternatives

- McFadden (1978) proved consistency of this approach
- But proof relies on independence of irrelevant alternatives (IIA) assumption
- Does not generalize to non-IIA models
  - So there is no theoretical justification for using sampling with mixed logit models
How should sampling work?

• Monte Carlo simulation #1
  – Fixed coefficient logit model
  – 500, 1000, or 2000 individuals making single discrete choice
  – 100 choice alternatives
  – 4 fixed coefficients
  – Sampling w/ 5, 10, 25 and 50 alternatives
  – Maximum likelihood estimation
  – 250 replications
How should sampling work?

Fixed Coefficient Means

Sample of Alternatives Size

Mean Parameter Bias

Relative Standard Error
Mixed Logit

Preference parameters vary randomly across population

• Continuous mixing distribution

\[ P_i = \int L_i(\beta_i)f(\beta_i | \theta)d\beta_i \]

• Finite mixing distribution

\[ P_i = \sum_{c} s_{ic}(\delta)L_{ic}(\beta_c) \]
What can go wrong?

- Monte Carlo simulation #2
  - Continuous mixing distribution (normal)
  - 500, 1000, or 2000 individuals making single discrete choice
  - 100 choice alternatives
  - 2 fixed coefficients, 2 random coefficients
  - Sampling w/ 5, 10, 25 and 50 alternatives
  - Maximum simulated likelihood estimation
  - 250 replications
What can go wrong?

Fixed Coefficient Means
What can go wrong?

Random Coefficient Means

Mean Parameter Bias

Sample of Alternatives Size

Relative Standard Error
What can go wrong?

Random Coefficient Standard Deviations

Sample of Alternatives Size

Mean Parameter Bias

Relative Standard Error

0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8

100 alt 50 alt 25 alt 10 alt 5 alt
What can go wrong?

• Monte Carlo simulation #3
  – Discrete mixing distribution (2 latent classes)
  – 500, 1000, or 2000 individuals making single discrete choice
  – 100 choice alternatives
  – 2 fixed coefficients, 2 random coefficients
  – Sampling w/ 5, 10, 25 and 50 alternatives
  – Maximum likelihood estimation
  – 250 replications
What can go wrong?

Fixed Coefficient Means

Mean Parameter Bias

Sample of Alternatives Size

Relative Standard Error
What can go wrong?

Latent Class Probability Coefficient Means

Sample of Alternatives Size

Mean Parameter Bias

Relative Standard Error
What can go wrong?

Random Coefficient Means

Mean Parameter Bias vs. Sample of Alternatives Size

Relative Standard Error
Practical Dilemma

• Mixed logit
  + Overcomes behavioral limitations of IIA
  + More flexibly accounts for unobserved pref. heterogeneity
    – Does not generate consistent estimates w/ sampling (McConnell and Tseng 2000; Nerella and Bhat 2004; our results)

• Fixed parameter logit
  – Limited by IIA
  – Limited ability to account for unobserved heterogeneity (nested logit?)
  + Does generate consistent estimates w/ sampling
Our Contribution

- Develop an expectation-maximization (EM) approach to estimate latent class mixed logit models for large choice set problems
  - Embeds sampling of alternatives at the M step
  - Computationally tractable for large (but not innumerable) choice sets
  - Can account for unobserved attributes / endogenity using Berry (1994) contraction mapping
Our Contribution

- Monte Carlo simulations suggest consistency
- Need relatively large sample size for precise estimates
- Quantify the small sample bias / precision / run time tradeoff with a recreation data set
Related Literature

• Fox (RAND, 2007), Spiller (Ph.D. diss., 2011)
  – Maximum score estimator using pairwise comparisons
    • Nonparametric approach that allows for heteroskedasticity in the errors across individuals but homoskedasticity and limited correlations across alternatives for a given individual
    • Works with choice sets that are effectively innumerable
      – Counterfactual analysis?
    • Assumes IIA
    • Only works with fixed parameter specifications
    • Can incorporate group specific (not alternative specific) constants
Latent Class Model

• Intuition:
  – Population can be segmented into finite number of types or classes
  – Analyst does not observe class membership (probabilistic)
  – Within each class, preferences are homogenous
  – But across classes, preferences are heterogeneous
Latent Class Model

• Setup:

\[ LL_i = \ln \left( \sum_c s_{ic} (\delta) L_{ic} (\beta_c) \right) \]

- where:

\[ s_{ic} (\delta) = \frac{\exp(z_i \delta_c)}{\sum_{c'=1}^{C} \exp(z_i \delta_{c'})} \]

\[ L_{ic} (\beta_c) = \prod_j \frac{\exp(x_{ij} \beta_c)}{\sum_{j'=1}^{J} \exp(x_{ij'} \beta_c)} \]
Expectation-Maximization (EM) Algorithm

• Attractive when estimating mixture models or models with latent data (i.e., class membership)

• Transforming the maximization of a log of a sum (mixed logit) into a recursive maximization of a sum of logs (logit)

• Because the M step involves logit estimation which embeds IIA assumption, can employ sampling
Latent Class Model via EM Algorithm

- **Expectation Step**
  - Construct expectation of likelihood conditional on data and current parameter estimates
  - Using Bayes rule, construct the probability of being in class \( c \) using **full** choice set

\[
\Pr_i(c | \delta^t, \beta^t, y) = \frac{s_{ic} \left( \delta^t \right) L_{ic} \left( \beta_c^t \right)}{\sum_{c'} s_{ic'} \left( \delta^t \right) L_{ic'} \left( \beta_{c'}^t \right)}
\]
Latent Class Model via EM Algorithm

- Maximization Step
  - Update parameter estimates by maximizing the conditional expected log-likelihood

\[
\text{Max}_{\delta, \beta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( s_{ic} (\delta) L_{ic} (\beta_c) \right) \right)
\]

\[
= \text{Max}_{\delta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( s_{ic} (\delta_c) \right) \right) + \text{Max}_{\beta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( L_{ic} (\beta_c) \right) \right)
\]

separate estimation
separatelogit estimation
Latent Class Model via EM Algorithm

- **Maximization Step**
  - Update parameter estimates by maximizing the conditional expected log-likelihood

$$\text{Max}_{\delta, \beta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( s_{ic}(\delta) L_{ic}(\beta_c) \right) \right)$$

$$= \text{Max}_{\delta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( s_{ic}(\delta) \right) \right) + \text{Max}_{\beta} \left( \sum_{i} \sum_{c} \text{Pr}(c | \delta^t, \beta^t, y) \ln \left( L_{ic}(\beta_c) \right) \right)$$

- Separate logit estimation can be used for sampling!
Latent Class Model via EM Algorithm

• Clarification:
  – **E step**: use full choice set
    • Generally straightforward, but problematic with innumerable choice sets
  – **M step**: use sample of alternatives
    • Logit estimation
Latent Class Model via EM Algorithm

• Iterate until convergence (small change in parameters)
Latent Class Model via EM Algorithm

Issues:

• Likelihood function is not globally concave
  – Try different starting values

• Inference
  – Three approaches
    • Bootstrapping
    • Plug in final estimates into full likelihood → Hessian
    • Gradients from final step of EM algorithm + OPG formula (Ruud 1991)
Latent Class Model via EM Algorithm

Issues:

• Likelihood function is not globally concave
  – Try different starting values

• Inference
  – Three approaches
    • Bootstrapping
    • Plug in final estimates into full likelihood → Hessian
    • Gradients from final step of EM algorithm + OPG formula (Ruud 1991) ★
Latent Class Model via EM Algorithm

Issues (cont.):

• Model selection
  – Information criteria

• Unobserved characteristics / endogenity
  – Because logit is a mean-fitting distribution, we can use the Berry (1994) contraction mapping to efficiently estimate alternative specific constants
Monte Carlo evidence

LC w/ EM

Fixed Coefficient Means

Sample of Alternatives Size

Mean Parameter Bias

Relative Standard Error
Monte Carlo evidence

LC w/ EM

Fixed Coefficient Means

Mean Parameter Bias vs. Sample of Alternatives Size
Monte Carlo evidence

LC w/ EM

Latent Class Probability Coefficient Means
Monte Carlo evidence
LC w/ EM

Random Coefficient Means

Mean Parameter Bias

Sample of Alternatives Size

Relative Standard Error
Monte Carlo evidence

LC w/ EM

Random Coefficient Means

Mean Parameter Bias

Sample of Alternatives Size

Relative Standard Error
Empirical Application

1997 Wisconsin angler data

- 512 anglers making site choices from 569 recreation sites (primarily lakes)
- Site choice influenced by travel costs, 15 site attributes (e.g., catch rates, bathrooms), and demographics (kids, income)
Conditional Logit Results
(average across 200 runs)

Scenario 4: Agricultural Runoff Mgmt
5% catch rate increase of all fish at all non-urban/forest/refuge sites
<table>
<thead>
<tr>
<th>Sample Size (%)</th>
<th>50%</th>
<th>25%</th>
<th>12.5%</th>
<th>5%</th>
<th>2%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (#)</td>
<td>285</td>
<td>142</td>
<td>71</td>
<td>28</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Efficiency Loss</td>
<td>6%</td>
<td>16%</td>
<td>33%</td>
<td>76%</td>
<td>165%</td>
<td>272%</td>
</tr>
<tr>
<td>Bias</td>
<td>1%</td>
<td>3%</td>
<td>6%</td>
<td>13%</td>
<td>28%</td>
<td>42%</td>
</tr>
<tr>
<td>Time Savings</td>
<td>56%</td>
<td>80%</td>
<td>90%</td>
<td>98%</td>
<td>99%</td>
<td>99%</td>
</tr>
</tbody>
</table>
Latent Class Results
(average across 25 runs)

Scenario 4: Agricultural Runoff Mgmt
5% catch rate increase of all fish at all non-urban/forest/refuge sites
## Latent Class Results
(average across 25 runs)

<table>
<thead>
<tr>
<th>Sample Size (%)</th>
<th>50%</th>
<th>25%</th>
<th>12.5%</th>
<th>5%</th>
<th>2%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (#)</td>
<td>285</td>
<td>142</td>
<td>71</td>
<td>28</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Efficiency Loss</td>
<td>10%</td>
<td>28%</td>
<td>51%</td>
<td>76%</td>
<td>84632%</td>
<td>18360%</td>
</tr>
<tr>
<td>Bias</td>
<td>19%</td>
<td>15%</td>
<td>6%</td>
<td>6%</td>
<td>34%</td>
<td>60%</td>
</tr>
<tr>
<td>Time Savings</td>
<td>33%</td>
<td>56%</td>
<td>75%</td>
<td>84%</td>
<td>81%</td>
<td>86%</td>
</tr>
</tbody>
</table>
Summary

- Exploiting modified EM algorithm, one can estimate random coefficient, discrete choice models
- Tradeoffs in terms of efficiency, bias & run time
- Our results suggest that moderately sized samples can generate good estimates in reasonable amounts of time
Extensions

• Mixed count data models
• Non-linear pricing models
Thank You!

Contact me with any comments:

roger_von_haefen@ncsu.edu