Specification of the cross nested logit model with sampling of alternatives for route choice models

Xinjun Lai    Michel Bierlaire

Transport and Mobility Laboratory  
School of Architecture, Civil and Environmental Engineering  
Ecole Polytechnique Fédérale de Lausanne

June 19, 2014
Outline

1. Introduction
2. Sampling of alternatives
3. MEV models
4. Validation on synthetic data
5. Case study with real data
Motivation

Route choice model
- Given an origin and a destination
- what is the preferred itinerary of a given traveler?

Difficulties
- Data
- Very large choice set
- Structural correlation among alternatives
Data

Revealed preferences
- Usually GPS data
- Unavailability of socio-economic variables

Stated preferences
- Hypothetical paths
- Simplified paths

In this paper...
GPS data
Very large choice set

Issue
Number of paths grows exponentially with the number of nodes

Literature
- link elimination  Azevedo et al. (1993)
- link penalty  de la Barra et al. (1993)
- labeled paths  Ben-Akiva et al. (1984)
- Sampling  Frejinger et al. (2009)
Structural correlation

Issue

Significant physical overlap

Literature

- **C-logit** Cascetta et al. (1996)
- **Path-size** Ben-Akiva and Bierlaire (1999)
- **Link-based cross-nested logit** Prashker and Bekhor (1999)
- **Error components** Ramming (2002); Frejinger and Bierlaire (2007)
In this paper...

Methodology
- Cross Nested logit
- Sampling of alternatives

Builds on...
- McFadden (1978)
- Vovsha and Bekhor (1998)
- Bierlaire et al. (2008)
- Frejinger et al. (2009)
- Guevara and Ben-Akiva (2013)
- Flötteröd and Bierlaire (2013)
Outline

1. Introduction

2. Sampling of alternatives

3. MEV models

4. Validation on synthetic data

5. Case study with real data
Logit model

\[ P(i|C) = \frac{e^{V_i}}{\sum_{j \in C} e^{V_j}} \]

McFadden (1978)

Sampling protocol

- Sample subset \( D \subseteq C \)
- Sampling probability \( q(D|j) \)
- Positive conditioning property

\[ q(D|i) > 0 \implies q(D|j) > 0 \ \forall j \in D. \]
Logit model

\[ P(i|C) \approx P(i|D) = \frac{e^{V_i + \ln q(D|i)}}{\sum_{j \in D} e^{V_j + \ln q(D|j)}} \]

Simple random sampling
- \( q(D|i) = q(D|j) \ \forall i, j \in C \)
- Correction terms cancel out
- Irrelevant, circuitous paths
- How to draw?

Importance sampling
- In \( q(D|i) \) are confounded with ASC
- How to draw?
How to draw?

Shortest path-based procedures
- link elimination: deterministic
- link penalty: deterministic
- labeled paths: deterministic
- SP on random costs:
  - some paths have 0 probability to be drawn
  - how to compute the sampling probability?
Sampling of alternatives

Metropolis-Hastings algorithm

Flötteröd and Bierlaire (2013)

Features

- Designed to draw from complex distributions
- Does not require the exact pmf/pdf
- Only a quantity proportional to it.
- For instance, to draw a path $p$ with probability

$$\frac{b_p}{\sum_{q \in C} b_q}$$

only $b_p$ are needed.
Sampling of alternatives

Metropolis-Hastings algorithm

Methodology

- Design a Markov chain $Q$ visiting the states/paths
- Accept/reject method
- Accept probability depends on
  - target (unnormalized) probabilities
  - transition probabilities of the Markov chain:

$$P(\text{accept}) = \min\left(\frac{b_q Q_{qp}}{b_p Q_{pq}}, 1\right)$$
Sampling of alternatives

Example

\[ b = (20, 8, 3, 1) \quad \pi = \left( \frac{5}{8}, \frac{1}{4}, \frac{3}{32}, \frac{1}{32} \right) \]

\[ Q = \begin{pmatrix}
  1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 \\
\end{pmatrix} \]

Run MH for 10000 iterations. Collect statistics after 1000

- Accept: \([2488, 1532, 801, 283]\]
- Reject: \([0, 952, 1705, 2239]\]
- Simulated: \([0.627, 0.250, 0.095, 0.028]\]
- Target: \([0.625, 0.250, 0.09375, 0.03125]\]
Difficulties

Design $Q$ such that

- Every path can be generated with nonzero probability
- Both $Q_{pq}$ and $Q_{qp}$ are known

Flötteröd and Bierlaire (2013)

- Proof of concept on synthetic data
- Application to Tel Aviv (17K links, 8K nodes)
Outline

1. Introduction
2. Sampling of alternatives
3. MEV models
4. Validation on synthetic data
5. Case study with real data
MEV models

Generic model

\[ P(i | C) = \frac{\exp(V_i + \ln G_i(C))}{\sum_{j \in C} \exp(V_j + \ln G_j(C))} \]

where \( G_i(C) = G_i(e^{V_1}, \ldots, e^{V_J}) \) is the derivative of the CPGF wrt \( e^{V_i} \).

Cross nested logit

\[ G_i(C) = \sum_{m=1}^{M} \left[ \mu_{\alpha m} e^{V_i(\mu_m - 1)} \left( \sum_{j \in C} \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu - \mu_m}{\mu_m}} \right], \]
MEV models

Generic model

\[ P(i|C) = \frac{\exp(V_i + \ln G_i(C))}{\sum_{j \in C} \exp(V_j + \ln G_j(C))} \]

where \( G_i(C) = G_i(e^{V_1}, \ldots, e^{V_j}) \) is the derivative of the CPGF wrt \( e^{V_i} \).

Cross nested logit

\[ G_i(C) = \sum_{m=1}^{M} \left[ \mu_{im} e^{V_i(\mu_m - 1)} \left( \sum_{j \in C} \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu - \mu_m}{\mu_m}} \right], \]
Sampling and MEV

\[ P(i|C) = \frac{\exp(V_i + \ln G_i(C))}{\sum_{j \in C} \exp(V_j + \ln G_j(C))} \]

Sampling correction

- If \( \ln G_j(C) \) is known, same idea as for logit

\[ \Pr(i|D) = \frac{\exp(V_i + \ln G_i(C) + \ln \Pr(D|i))}{\sum_{j \in D} \exp(V_j + \ln G_j(C) + \ln \Pr(D|j))}. \]

- Not confused with the constants anymore.
Correction term

\[ \Pr(\mathcal{D}|p) \propto \frac{k_p}{q(p)} \]

where

- \( k_p \) is the number of times path \( p \) has been generated
- \( q(p) \) is the sampling probability of path \( p \)
- \( q(p) \propto b_p \)
Model I

\[
\Pr(i|\mathcal{D}) = \frac{\exp(V_i + \ln G_i(C) + \ln \frac{k_i}{b_i})}{\sum_{j \in \mathcal{D}} \exp(V_j + \ln G_j(C) + \ln \frac{k_j}{b_j})},
\]
Approximation of $\ln G_i(C)$

Guevara and Ben-Akiva (2013)

$$G_i(C) \approx \hat{G}_i(D, w) = \sum_{m=1}^{M} \mu \alpha_{im} e^{V_i(\mu_m-1)} \left( \sum_{j \in D} w_j \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu - \mu_m}{\mu_m}}$$

where $w_j$ expansion factor to be defined.
Expansion factors: Guevara and Ben-Akiva (2013)

Realized / expected

\[ w_j^G = \frac{k_j}{E[k_j]} = \frac{k_j}{q(j)R} = \frac{k_jB}{b(j)R} \]

where

- \( R \) is the number of draws used to generate \( D \)
- \( B = \sum_{j \in C} b(j) \) [Requires enumeration of \( C \)]

Approximate \( B \)

\[ B = \sum_{j \in C} b(j) = |C| \frac{\sum_{i \in C} b(i)}{|C|} = |C| \bar{b}, \]

and

\[ \bar{b} = \frac{\sum_{i \in C} b(i)}{|C|} \approx \frac{\sum_{i \in D} b(i)}{|D|}. \]
Expansion factors: Guevara and Ben-Akiva (2013)

Approximation

\[ w_j^G = \frac{k_j}{b(j)R |D|} \sum_{i \in D} b(i) \]

which require \(|C|\)

Approximate \(|C|\)

Roberts and Kroese (2007)

\(N\) random walks in the network

\[ |C| \approx \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\ell(i)}. \]

\(\ell(i)\): likelihood of the path generated by the algorithm during run \(i\)
Expansion factors: Frejinger et al. (2009)

Account for the upper bound

\[ w_j^F = \begin{cases} 
1 & \text{if } b(j)R > B, \\
\frac{B}{b(j)R} & \text{otherwise.} 
\end{cases} \]

Same approximation of \( B \)

\[ B \approx \frac{|C|}{|D|} \sum_{i \in D} b(i) \]

Again, requires \(|C|\)
Avoiding $|C|$:

- Let $s$ be the path which has been sampled the most in $\mathcal{D}$.
- $k_s \geq k_p$, for each $p \in \mathcal{D}$.
- If sample is large enough, $k_s \approx q(s)R$

\[
    w_j^G = \frac{k_j}{q(j)R} \approx w_j^l = \frac{k_j}{q(j)R} \frac{q(s)R}{k_s} = \frac{k_j}{b(j)} \frac{b(s)}{k_s}
\]

which does not require $B$ or $|C|$. 
Expansion factors

- Guevara and Ben-Akiva (2013)
  \[ w_j^G = \frac{k_j}{b(j)R} B \quad \text{with} \quad B \approx \frac{|C|}{|D|} \sum_{i \in D} b(i) \]

- Frejinger et al. (2009)
  \[ w_j^F = \begin{cases} 
  1 & \text{if } b(j)R > B, \\
  \frac{B}{b(j)R} & \text{otherwise.} 
\end{cases} \quad \text{with} \quad B \approx \frac{|C|}{|D|} \sum_{i \in D} b(i). \]

- Lai and Bierlaire (2014)
  \[ w_j^L = \frac{k_j}{b(j)} \frac{b(s)}{k_s} \]
Models to be compared

- **Model I**: true \( G_i \) (impossible in practice)
  \[
  Pr(i|D) = \frac{\exp(V_i + \ln G_i(C) + \ln \frac{k_i}{b(i)})}{\sum_{j \in D} \exp(V_j + \ln G_j(C) + \ln \frac{k_j}{b(j)})}
  \]

- **Model II**: the proposed model
  \[
  Pr(i|D, D', w) = \frac{\exp(V_i + \ln G_i(D', w)) + \ln \frac{k_i}{b(i)})}{\sum_{j \in D} \exp(V_j + \ln G_j(D', w) + \ln \frac{k_j}{b(j)})}
  \]

- **Model III**: no expansion factor, no sampling correction (benchmark)
  \[
  Pr(i|D, D') = \frac{\exp(V_i + \ln G_i(D', 1))}{\sum_{j \in D} \exp(V_j + \ln G_j(D', 1))}
  \]
Outline

1. Introduction
2. Sampling of alternatives
3. MEV models
4. Validation on synthetic data
5. Case study with real data
The network: 170 paths (Frejinger (2008))
The true model: cross-nested logit

Utility

\[ V_i = \beta_L L_i + \beta_{SB} SB_i, \]

“True” parameters

- \( \beta_L = -0.5 \) and \( \beta_{SB} = -0.1 \)
- \( \mu_m = 1.5 \) for each link \( m \)
- \( \alpha_{im} = \ell_m / L_i \)

Data

3000 synthetic choices
Re-estimate the parameters of the true model

Full choice set

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Est.</th>
<th>Std err.</th>
<th>t-test (0)</th>
<th>t-test (true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_L$</td>
<td>-0.501</td>
<td>0.0118</td>
<td>43.1</td>
<td>0.678</td>
</tr>
<tr>
<td>$\beta_{SB}$</td>
<td>-0.0910</td>
<td>0.0240</td>
<td>3.19</td>
<td>0.375</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>1.49</td>
<td>0.0269</td>
<td>55.2</td>
<td>0.0535</td>
</tr>
</tbody>
</table>
Sampling paths

Metropolis-Hastings

\[ b(i) = \exp(-\theta L_i), \quad \theta \geq 0 \]
Validation on synthetic data

Number of generated paths

Number of generated paths vs. Number of draws for different values of \( \theta \):
- \( \theta = 0.01 \)
- \( \theta = 0.1 \)
- \( \theta = 0.3 \)
- \( \theta = 0.5 \)
- \( \theta = 1 \)

Lai & Bierlaire (EPFL)

CNL and sampling of alternatives

June 19, 2014
## Model I: true $G_i$ — MH $\theta = 0.5$

<table>
<thead>
<tr>
<th>Draws</th>
<th>Est.</th>
<th>Std err.</th>
<th>t-test(0)</th>
<th>t-test(true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$\beta_L (-0.5)$</td>
<td>-0.443</td>
<td>0.0163</td>
<td>27.3</td>
</tr>
<tr>
<td></td>
<td>$\beta_{SB} (-0.1)$</td>
<td>-0.0647</td>
<td>0.0427</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>$\mu_m (1.5)$</td>
<td>1.56</td>
<td>0.0340</td>
<td>45.8</td>
</tr>
<tr>
<td></td>
<td><strong>Estimation time:</strong></td>
<td>1362 seconds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>$\beta_L (-0.5)$</td>
<td>-0.479</td>
<td>0.0156</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>$\beta_{SB} (-0.1)$</td>
<td>-0.0720</td>
<td>0.0393</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>$\mu_m (1.5)$</td>
<td>1.51</td>
<td>0.0322</td>
<td>47.0</td>
</tr>
<tr>
<td></td>
<td><strong>Estimation time:</strong></td>
<td>4648 seconds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Validation on synthetic data

Model I: true $G_i$ — MH $\theta = 0.01$

<table>
<thead>
<tr>
<th>Draws</th>
<th>Est.</th>
<th>Std err.</th>
<th>t-test(0)</th>
<th>t-test(true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_L$ (-0.5)</td>
<td>-0.535</td>
<td>0.0174</td>
<td>30.8</td>
<td>2.01</td>
</tr>
<tr>
<td>$\beta_{SB}$ (-0.1)</td>
<td>-0.132</td>
<td>0.0545</td>
<td>2.42</td>
<td>0.580</td>
</tr>
<tr>
<td>$\mu_m$ (1.5)</td>
<td>1.41</td>
<td>0.0355</td>
<td>39.8</td>
<td>2.47</td>
</tr>
<tr>
<td>Estimation time: 1612 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Draws</th>
<th>Est.</th>
<th>Std err.</th>
<th>t-test(0)</th>
<th>t-test(true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_L$ (-0.5)</td>
<td>-0.544</td>
<td>0.0160</td>
<td>33.9</td>
<td>2.76</td>
</tr>
<tr>
<td>$\beta_{SB}$ (-0.1)</td>
<td>-0.130</td>
<td>0.0410</td>
<td>3.16</td>
<td>0.726</td>
</tr>
<tr>
<td>$\mu_m$ (1.5)</td>
<td>1.41</td>
<td>0.0322</td>
<td>43.8</td>
<td>2.85</td>
</tr>
<tr>
<td>Estimation time: 4914 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model I: comments

- Trade-off between dispersion (low $\theta$) and number of draws
- Lower value of $\theta$ requires more draws
- $\theta = 0.5$, 40 draws: parameters are correctly estimated
- First sampling scheme is validated
- No specific guideline for $\theta$ and $R$
Approximating $\bar{b}$ and $|C|$ 

**Protocol**

- For $\bar{b}$: generate $D$ using MH with 100 draws and $\theta = 0.01$
- For $|C|$: generate 10000 paths using random walk
- Repeat 100 times
- Compute the empirical mean and standard error

**Results**

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Mean</th>
<th>Std err</th>
<th>t-test(true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{b}$</td>
<td>0.688</td>
<td>0.684</td>
<td>0.0023</td>
<td>1.62</td>
</tr>
<tr>
<td>$</td>
<td>C</td>
<td>$</td>
<td>170</td>
<td>169.8</td>
</tr>
</tbody>
</table>
Validation on synthetic data

Model II

Protocol

- Denominator: $\mathcal{D}$ generated with MH (40 draws, $\theta = 0.5$)
- Expansion factor: $\mathcal{D}'$ MH with various values
Validation on synthetic data

Model II: 100 draws ($t$-test vs true value)

<table>
<thead>
<tr>
<th></th>
<th>$w^G$</th>
<th>$w^F$</th>
<th>$w^L$</th>
<th>$w = 1$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mod. II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>2.48</td>
<td>4.34</td>
<td>1.25</td>
<td>3.59</td>
<td>19.4</td>
</tr>
<tr>
<td>$\beta_{SB}$</td>
<td>0.910</td>
<td>0.867</td>
<td>0.722</td>
<td>0.179</td>
<td>0.221</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>2.02</td>
<td>3.09</td>
<td>0.437</td>
<td>2.98</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Sampling protocol for $D'$: $\theta = 0.5$

<table>
<thead>
<tr>
<th></th>
<th>$w^G$</th>
<th>$w^F$</th>
<th>$w^L$</th>
<th>$w = 1$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mod. II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>4.61</td>
<td>4.23</td>
<td>4.48</td>
<td>4.30</td>
<td>18.9</td>
</tr>
<tr>
<td>$\beta_{SB}$</td>
<td>0.303</td>
<td>0.297</td>
<td>0.254</td>
<td>0.467</td>
<td>0.634</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>4.70</td>
<td>4.71</td>
<td>5.38</td>
<td>4.55</td>
<td>3.63</td>
</tr>
</tbody>
</table>

Sampling protocol for $D'$: $\theta = 0.01$
Model II: 200 draws ($t$-test vs true value)

### Sampling protocol for $D'$: $\theta = 0.5$

|          | $w^G$ | $w^F$ | $w^L$ | $w = 1$ | $D'$
|----------|-------|-------|-------|---------|-------
| $\beta_L$ | 0.578 | 10.5  | 0.0374 | 3.38    | 18.9  |
| $\beta_{SB}$ | 0.513 | 0.194 | 0.440 | 0.259   | 0.269 |
| $\mu_m$    | 1.36  | 5.02  | 1.34  | 3.07    | 0.965 |

### Sampling protocol for $D'$: $\theta = 0.01$

|          | $w^G$ | $w^F$ | $w^L$ | $w = 1$ | $D'$
|----------|-------|-------|-------|---------|-------
| $\beta_L$ | 3.51  | 3.84  | 2.86  | 4.37    | 18.5  |
| $\beta_{SB}$ | 0.173 | 0.119 | 0.298 | 0.409   | 0.571 |
| $\mu_m$    | 9.11  | 8.65  | 7.19  | 5.41    | 3.72  |
Validation on synthetic data

**Model II: 300 draws (t-test vs true value)**

<table>
<thead>
<tr>
<th></th>
<th>Mod. II</th>
<th></th>
<th></th>
<th></th>
<th>Mod. III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_G$</td>
<td>$w_F$</td>
<td>$w_L$</td>
<td>$w = 1$</td>
<td></td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>0.981</td>
<td>3.62</td>
<td>0.703</td>
<td>0.981</td>
<td>19.3</td>
</tr>
<tr>
<td>$\beta_{SB}$</td>
<td>0.428</td>
<td>1.34</td>
<td>0.537</td>
<td>0.428</td>
<td>0.0052</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>2.28</td>
<td>3.12</td>
<td>1.70</td>
<td>2.28</td>
<td>1.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mod. II</th>
<th></th>
<th></th>
<th></th>
<th>Mod. III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_G$</td>
<td>$w_F$</td>
<td>$w_L$</td>
<td>$w = 1$</td>
<td></td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>0.809</td>
<td>0.0271</td>
<td>1.02</td>
<td>5.05</td>
<td>18.5</td>
</tr>
<tr>
<td>$\beta_{SB}$</td>
<td>0.565</td>
<td>0.780</td>
<td>0.480</td>
<td>0.564</td>
<td>0.654</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>1.66</td>
<td>0.650</td>
<td>1.84</td>
<td>5.19</td>
<td>3.01</td>
</tr>
</tbody>
</table>
Validation on synthetic data

Comments

- $\theta = 0.5$ seems again the most appropriate
- Model II outperforms Model III (no correction, no expansion factor)
- New expansion factor is the most appropriate (already good results with 100 draws)
- $\mu_m$ seems to be the most sensitive parameters
Validation on synthetic data

$t$-tests with $w^L$ and $\theta = 0.5$

The diagram shows the $t$-test results with respect to the true value for different draws and paths in the choice set. The x-axis represents the number of draws/number of paths in the choice set, while the y-axis represents the $t$-test value. The plot includes:

- t-test (beta_L)
- t-test (beta_SB)
- t-test (mu_m)

The graph illustrates the trend of how the $t$-test values change as the number of draws and paths in the choice set increases.
Outline

1. Introduction
2. Sampling of alternatives
3. MEV models
4. Validation on synthetic data
5. Case study with real data
Tianhe region (CBD) of Guangzhou (China)
Case study with real data

Data

Network
- 208 nodes
- 662 links
- 24 major roads
- 34 arterial streets
- 32 minor streets
- 57 signalized intersections

GPS traces from taxis
- 7 ODs
- 740 trips
Model

Utility

\[ V_i = \beta_L \text{Length}_i + \beta_{ARR} \text{ArteryRoadRatio}_i + \beta_S \text{Signal}_i. \]

Cross-nested logit

- Two nests: \( \mu \): non-artery roads, \( \mu_{mA} \): artery roads
- \( \alpha_{im} = \frac{\ell_m}{L_i} \)

MH sampling

| \( \theta \) | \( |D| \) | \( \theta \) | \( |D| \) |
|---|---|---|---|
| 0.005 | 29 | 0.0025 | 3813 |
| 0.004 | 54 | 0.0023 | 5624 |
| 0.003 | 201 | 0.002 | 7766 |
| 0.0028 | 2036 | 0.001 | 9836 |
Case study with real data

Estimation results (with Matlab, Intel i5 with 4GB RAM, one processor)

\[ \theta = 0.003 \]

<table>
<thead>
<tr>
<th>Model II</th>
<th>Est.</th>
<th>Std. err.</th>
<th>t-test (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_L )</td>
<td>-1.58</td>
<td>0.0566</td>
<td>27.9</td>
</tr>
<tr>
<td>( \beta_{ARR} )</td>
<td>8.09</td>
<td>0.636</td>
<td>12.7</td>
</tr>
<tr>
<td>( \beta_S )</td>
<td>-0.513</td>
<td>0.267</td>
<td>1.91</td>
</tr>
<tr>
<td>( \mu_m )</td>
<td>3.90</td>
<td>0.117</td>
<td>33.3</td>
</tr>
<tr>
<td>( \mu_{mA} )</td>
<td>2.22</td>
<td>0.257</td>
<td>8.62</td>
</tr>
</tbody>
</table>

Number of observations: 740 trips from 7 OD
Null log likelihood: -3.4078e+03
Final log likelihood: -1.9206e+03
Estimation time: 22.32 hours
Conclusion

Contributions

- Application of sampling of alternative for MEV and route choice
- New expansion factor
- Validity check: synthetic data
- Feasibility check: real data
- Heavy, but tractable

Future work

- Investigate other nesting structures
- Different ways to approximate $G_i$
- Estimation of $\alpha_{im}$ (?)
Bibliography I


Bibliography II


Bibliography IV


