A latent route choice model in Switzerland

M. Bierlaire, E. Frejinger and J. Stojanovic

transp-or.epfl.ch

Transport and Mobility Laboratory, EPFL
Outline

• Swiss mobility pricing project
• Aggregate observations and latent choices
• Modeling approach
• Empirical results
• Conclusion
Mobility Pricing

How’s the congestion charge trial going?

Not well.

Isn’t it keeping people off the road?

Yes...

...but not in the way we’d like...
Swiss Mobility Pricing Project

- A part of a major study on various mobility pricing scenarios in Switzerland
- A collaboration with ETH Zurich and USI Lugano
- Revealed Preferences (RP) and Stated Preferences (SP) data has been collected
- RP data concern long distance route choice by car
  - Route descriptions are approximative
  - Route choices are latent
Objective

- Estimate route choice models based on latent chosen routes
- Literature on latent choice models
  - Ben-Akiva et al. (1984), label path approach
  - Ben-Akiva and Lerman (1985), destination choice
  - Toledo et al. (2003), lane choice
Observations

- Exact descriptions of chosen routes are difficult and expensive to obtain
- The concept of path and network as we need for modeling is abstract for respondents
- Here, a chosen route is described by a sequence of cities and locations
- *Aggregate observations* (several paths in the network can correspond to the same observation)
Observations

- Better quality of the observations
- Travelers do not need to refer to the network used by the analyst
- Exact origin-destination pairs are not necessarily known
- Exact route is not known
Observations - Example
Modeling Approach

- Several possible modeling approaches
  - Construction of paths from the aggregate observations
    - Involves subjective judgments and generate noise
  - Alternatives in the model are aggregates instead of physical paths
    - Estimated model is of little use in practice
- Our approach: compute the likelihood of an aggregate observation for a classical route choice model
Modeling Approach

- Probability of an aggregate observation $i$:

\[
P(i) = \sum_{s \in S} P(s|i) \sum_{r \in C_s} P(r|i) P(r|C_s)
\]

- $s$: origin-destination pair
- $S$: set of all origin-destination pairs
- $r$: route
- $C_s$: set of all routes for origin-destination pair $s$
Modeling Approach

- Probability of an aggregate observation $i$:

$$P(i) = \sum_{s \in S} P(s|i) \sum_{r \in C_s} P(r|i)P(r|C_s)$$

- $P(s|i)$ and $P(r|i)$ can be modeled in several ways

- $P(r|C_s)$: route choice model that is identifiable if
  1. at least one of the routes in $C_s$ crosses the observed zones, and
  2. at least one route in $C_s$ does not cross the observed zones.

- This type of models can be estimated with BIOGEME
Empirical Results

- Simplified Swiss network (39411 links and 14841 nodes)
- RP data collection through telephone interviews
- Long distance car travel
- The chosen routes are described with the origin and destination cities as well as 1 to 3 cities or locations that the route pass by
- 940 observations available after data cleaning and verification
Empirical Results
Empirical Results

- This application is one of few presented in the literature that are based on RP data
- The network is to our knowledge the largest one used for evaluation of route choice modeling approaches
Empirical Results

- No information available on the exact origin destination pairs

\[ P(s|i) = \frac{1}{|S_i|} \quad \forall s \in S_i \]

- \( P(r|i) \) is modeled with a binary variable

\[ \delta_{ri} = \begin{cases} 
1 & \text{if } r \text{ corresponds to } i \\
0 & \text{otherwise} 
\end{cases} \]
Empirical Results

- Two origin-destination pairs are randomly chosen for each observation
- 46 routes per choice set are generated with a choice set generation algorithm
- After choice set generation 780 observations are available
  - 160 observations were removed because either all or none of the generated routes crossed the observed zones
Empirical Results

- Probability of an aggregate observation $i$

\[ P(i) = \sum_{s \in S_i} \frac{1}{|S_i|} \sum_{r \in C_s} \delta_{ri} P(r|C_s) \]

- We estimate Path Size Logit (Ben-Akiva and Bierlaire, 1999) and Subnetwork (Frejinger and Bierlaire, 2006) models

- BIOGEME (biogeme.epfl.ch) used for all model estimations
Empirical Results - Subnetwork

- Subnetwork: main motorways in Switzerland
- Correlation among routes is explicitly modeled on the subnetwork
- Combined with a Path Size attribute
- Linear-in-parameters utility specifications
Empirical Results - Subnetwork
<table>
<thead>
<tr>
<th>Parameter</th>
<th>PSL</th>
<th>Subnetwork</th>
</tr>
</thead>
<tbody>
<tr>
<td>In(path size) based on free-flow time</td>
<td>1.04 (0.134) 7.81</td>
<td>1.10 (0.141) 7.78</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>Freeway free-flow time 0-30 min</td>
<td>-7.12 (0.877) -8.12</td>
<td>-7.45 (0.984) -7.57</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-7.12</td>
<td>-7.04</td>
</tr>
<tr>
<td>Freeway free-flow time 30min - 1 hour</td>
<td>-1.69 (0.875) -1.93</td>
<td>-2.26 (1.03) -2.19</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-1.69</td>
<td>-2.14</td>
</tr>
<tr>
<td>Freeway free-flow time 1 hour +</td>
<td>-4.98 (0.772) -6.45</td>
<td>-5.64 (1.00) -5.61</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-4.98</td>
<td>-5.33</td>
</tr>
<tr>
<td>CN free-flow time 0-30 min</td>
<td>-6.03 (0.882) -6.84</td>
<td>-6.25 (0.975) -6.41</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-6.03</td>
<td>-5.91</td>
</tr>
<tr>
<td>CN free-flow time 30 min +</td>
<td>-1.87 (0.331) -5.64</td>
<td>-2.16 (0.384) -5.63</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-1.87</td>
<td>-2.04</td>
</tr>
<tr>
<td>Main free-flow travel time 10 min +</td>
<td>-2.03 (0.502) -4.05</td>
<td>-2.46 (0.624) -3.95</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-2.03</td>
<td>-2.33</td>
</tr>
<tr>
<td>Small free-flow travel time</td>
<td>-2.16 (0.685) -3.16</td>
<td>-2.75 (0.804) -3.42</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-2.16</td>
<td>-2.60</td>
</tr>
<tr>
<td>Proportion of time on freeways</td>
<td>-2.2 (0.812) -2.71</td>
<td>-2.31 (0.865) -2.67</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-2.2</td>
<td>-2.18</td>
</tr>
<tr>
<td>Proportion of time on CN</td>
<td>0 fixed</td>
<td>0 fixed</td>
</tr>
<tr>
<td>Proportion of time on main</td>
<td>-4.43 (0.752) -5.88</td>
<td>-4.40 (0.800) -5.51</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-4.43</td>
<td>-4.16</td>
</tr>
<tr>
<td>Proportion of time on small</td>
<td>-6.23 (0.992) -6.28</td>
<td>-6.02 (1.03) -5.83</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>-6.23</td>
<td>-5.69</td>
</tr>
<tr>
<td>Covariance parameter</td>
<td>0.217 (0.0543) 4.00</td>
<td>0.205</td>
</tr>
<tr>
<td>Scaled Estimate</td>
<td>0.205</td>
<td></td>
</tr>
</tbody>
</table>
## Empirical Results

<table>
<thead>
<tr>
<th></th>
<th>PSL</th>
<th>Subnetwork</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance parameter</td>
<td></td>
<td>0.217</td>
</tr>
<tr>
<td>(Rob. Std. Error) Rob. T-test</td>
<td>(0.0543)</td>
<td>(4.00)</td>
</tr>
<tr>
<td>Number of simulation draws</td>
<td>-</td>
<td>1000</td>
</tr>
<tr>
<td>Number of parameters</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-1164.850</td>
<td>-1161.472</td>
</tr>
<tr>
<td>Adjusted rho square</td>
<td>0.145</td>
<td>0.147</td>
</tr>
</tbody>
</table>

Sample size: 780, Null log-likelihood: -1375.851
Empirical Results

- All parameters have their expected signs and are significantly different from zero
- The values and significance level are stable across the two models
- The subnetwork model is significantly better than the Path Size Logit (PSL) model
Conclusion

- Aggregate observations are convenient to report paths
- They can be used for estimating route choice models
- Care must be taken about the level of aggregation
- Parameters of the RP model are significant and meaningful
- Available in Biogeme / Bioroute