Parameter estimation for activity-based models

Janody Pougala · Tim Hillel · Michel Bierlaire

STRC 2022
Introduction

Activity demand

Travel demand

Socio-economic characteristics
Social interactions
Cultural norms
Basic needs
...
(Chapin, 1974)

Time and space constraints
(Hägerstrand, 1970)
### Introduction

<table>
<thead>
<tr>
<th>Utility-based models</th>
<th>Rule-based models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision is made by maximizing utility derived from activities</td>
<td>Decision is made by considering context-dependent rules</td>
</tr>
<tr>
<td>e.g. Bowman &amp; Ben-Akiva, 2001</td>
<td>e.g. Gollegde et al., 1994</td>
</tr>
<tr>
<td>Bhat et al, 2004</td>
<td>Arentze &amp; Timmermans 2000</td>
</tr>
<tr>
<td>Pougala et al, 2021</td>
<td></td>
</tr>
</tbody>
</table>

|
Introduction

Utility-based models

*Decision is made by maximizing utility derived from activities*

- e.g.
  - Bowman & Ben-Akiva, 2001
  - Bhat et al, 2004
  - Pougala et al, 2021

Rule-based models

*Decision is made by considering context-dependent rules*

- e.g.
  - Gollegde et al., 1994
  - Arentze & Timmermans 2000
**Background**

- Optimisation-based simulation framework for activity-based models
  - Pougala et al, 2022

- Joint estimation
  - Activity participation
  - Activity scheduling
  - Mode choice
  - Location choice
Parameter estimation

- Data
- Literature
- Parameters $\beta_n$
- Synthetic individual $n$
- Disturbances $\epsilon$
- Optimisation $\Omega$
- Schedule $S^n_{\epsilon}$
- Indicators
Parameter estimation

- Maximum likelihood estimation (MLE) of parameters in DCM:

\[
\hat{\theta} = \arg \max L_n(\theta) \\
L_n = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}}
\]
Parameter estimation

- Maximum likelihood estimation (MLE) of parameters in DCM:

\[ \hat{\theta} = \arg \max L_n(\theta) \]

\[ L_n = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}} \]

- Common assumptions on choice set:
  - Universal across population
  - Fully observed or observable

Enumeration over choice set \( C_n \)
Parameter estimation

- Maximum likelihood estimation (MLE) of parameters in DCM:

\[
\hat{\theta} = \arg \max L_n(\theta)
\]

\[
L_n = \prod_{n=1}^{N} \prod_{i \in C_n^*} P_n(i|C_n^*)^{y_{in}}
\]

- Sample of alternatives \( C_n^* \subset C_n \):
  - Correction of the choice probabilities (Ben-Akiva & Lerman, 1985)
Choice set

Feasible schedules

Considered schedules

Unobserved and possibly infinite

Estimation choice set: sample of feasible schedules generated for estimation purposes

Actual choice set: Unobserved

Realized schedule

Based on Shocker (1991)
Choice set

- Choice set generation
  - Metropolis-Hastings sampling of feasible schedules
  - STRC 2021
Utility of a schedule: \( U_n = \sum_a U_{an} \)

- For an individual \( n \) considering an activity \( a \) with a flexibility \( k \):

\[
U_{an} = U_{\text{const}} + U_{\text{early}} + U_{\text{late}} + U_{\text{long}} + U_{\text{short}} + U_{\text{travel}} + \varepsilon_{an}
\]

Start time deviations:

\[
U_{\text{early}} = \theta_{ek} \max(0, x^*_a - x_a) \\
U_{\text{late}} = \theta_{lk} \max(0, x_a - x^*_a)
\]

Duration deviations:

\[
U_{\text{short}} = \theta_{dsk} \max(0, \tau^*_a - \tau_a) \\
U_{\text{long}} = \theta_{dlk} \max(0, \tau_a - \tau^*_a)
\]
Case study

- Lausanne population, MTMC 2015 (BFS & ARE, 2017)

  - 3 samples:
    - S1: Students (236 individuals)
    - S2: Workers (618 individuals)
    - S3: All occupations (1118 individuals)

  - Choice set sizes:
    - $N = 10$ alternatives for S1, S2
    - $N = 100$ for S3

Model 1 (14 parameters):
- Activity-specific constants
- Aggregated penalties (flexible vs. Non flexible)

Model 2 (30 parameters):
- Activity-specific constants
- Activity specific penalties
## Results

- **Model 1, Workers: constants**
  - Reference: ASC Home = 0
  - $\bar{R}^2 = 0.77$
  - Runtime: 1.36 s

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Param. estimate</th>
<th>Rob. Std error</th>
<th>Statistical significance $(p &lt; 0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC Education</td>
<td>10.8</td>
<td>2.50</td>
<td>Significant</td>
</tr>
<tr>
<td>ASC Errands</td>
<td>7.63</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>ASC Escort</td>
<td>9.79</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>ASC Leisure</td>
<td>15.3</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>ASC Shopping</td>
<td>12.5</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>ASC Work</td>
<td>18.5</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>
## Results

- **Model 1, Workers: penalties**
  - Flexible (F): errands, leisure, shopping
  - Non-Flexible (NF): work, education, escort

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Param. estimate</th>
<th>Rob. Std error</th>
<th>Statistical significance ($p &lt; 0.05$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F: early</td>
<td>-0.813</td>
<td>0.160</td>
<td>Significant</td>
</tr>
<tr>
<td>F: late</td>
<td>-1.12</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td>F: short*</td>
<td>0*</td>
<td>- *</td>
<td>Not significant</td>
</tr>
<tr>
<td>F: long</td>
<td>-0.569</td>
<td>0.165</td>
<td>Significant</td>
</tr>
<tr>
<td>NF: early</td>
<td>-0.827</td>
<td>0.160</td>
<td></td>
</tr>
<tr>
<td>NF: late</td>
<td>-1.26</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>NF: short</td>
<td>-3.24</td>
<td>0.555</td>
<td></td>
</tr>
<tr>
<td>NF: long</td>
<td>-0.789</td>
<td>0.229</td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Estimated ABM parameters using sampled choice sets
- Expected signs, ratios and significance

Further work:
- Complex model specifications (logit mixtures, non-linear parameters…)
- Investigate stability (increasing N)
- Application on synthetic population for validation
Thank you!

janody.pougala@epfl.ch
tim.hillel@ucl.ac.uk
michel.bierlaire@epfl.ch
References

- Schultheiss M., Spatial familiarity and mobility motifs, Bridging Transportation Researchers, August 2021