Preferred citation style

Similarities in air transport connection choice

Nadine Schuessler

IVT
ETH
Zurich

September 2007
Problem and objectives

Two main objectives

- Modelling air transport connection choice

- Investigate similarity measures for public transport connection choice
Accounting for the IIA property of the MNL

In general, there are three different ways:


- Similarity measures
Similarity measures

Based on the Implicit Availability and Perception (IAP) model by Cascatta et al. (1996)

The utility of an alternative is extended by a term $q_{Cin}$ representing its degree of membership in the individual choice set $C_n$:

$$U_{in} = V_{in}' + q_{Cin} + \varepsilon_{in}$$

The similarity of an alternative with other alternatives
- decreases its probability to be perceived as separate alternative
- decreases its probability to be included in the individual choice set

=>$>$ Thus decreases its probability to be chosen
Existing formulations for similarity measures

**Private transport route choice**
- C-Logit Cascetta et al. (1996)
- Path Size Logit Bierlaire and Ben-Akiva (1999), Ramming (2002)

**Multimodal route choice**
- Path Size Logit Hoogendoorn-Lanser and Bovy (2007)

**Public transport Connection Choice**
- Autonomy of a connection Friedrich et al. (2000)

**Spatial Correlations**
- Competing Destinations Fotheringham (1988)
- Spatial Dependency Parameter Mohammadian et al. (2005)
- Field Effect Variable Dugundji and Walker (2005)
Autonomy of a connection

Choice probability depending on the independence of a connection

\[ P_{in} = \frac{e^{V_{in} \cdot \text{IND}(i)}}{\sum_{j \in C} e^{V_{in} \cdot \text{IND}(j)}} = \frac{e^{V_{in} + \ln(\text{IND}(i))}}{\sum_{j \in C} e^{V_{in} + k \ln(\text{IND}(j))}} \]

Independence of a connection as reciprocal sum of similarities:

\[ \text{IND}(c) = \frac{1}{\sum_{c' \in C} f_c(c')} = \frac{1}{1 + \sum_{c' \in C, c' \neq c} f_c(c')} \]

Similarity takes into account similarities in departure and arrival times \( x_c(c') \), perceived journey time \( y_c(c') \) and fare \( z_c(c') \)

\[ f_c(c') = \left(1 - \frac{x_c(c')}{s_x}\right)^+ \cdot \left(1 - \gamma \cdot \min\left(1, \frac{y_c(c')}{s_y} + \frac{z_c(c')}{s_z}\right)\right) \]

\( s_x, s_y \) and \( s_z \) set the range of influence of \( x_c(c'), y_c(c') \) and \( z_c(c') \)
\( \gamma \) weights the right part of the formula
### Parameter settings for the independence measure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>sx, sy, sz tests</th>
<th>γ tests</th>
<th>Final model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Symmetric setting</td>
<td>Extreme values</td>
<td>γ = 0.25</td>
</tr>
<tr>
<td>sx</td>
<td>720</td>
<td>360</td>
<td>720</td>
</tr>
<tr>
<td>sy⁺</td>
<td>780</td>
<td>180</td>
<td>780</td>
</tr>
<tr>
<td>sy⁻</td>
<td>780</td>
<td>120</td>
<td>540</td>
</tr>
<tr>
<td>sz⁺</td>
<td>1.7</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td>sz⁻</td>
<td>1.7</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td>γ</td>
<td>0.5</td>
<td>0.5</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Effect of varying $s_x$, $s_y$ and $s_z$
Effect of varying $\gamma$
Available Data – Data Sources

Marketing Data Information Transfer (MIDT) Dataset – booked tickets in November 2006
~ 200,000 bookings

Expedia Dataset observed prices on 70 origin-destination pairs between September and November 2006
~ 12,000,000 observations

Official Airline Guide (OAG) November – information on code share, type of aircraft, detailed waiting times can be extracted

Matching these datasets on several criteria leads to almost 19,000 observed choices
Variables for connection choice model

Alternatives are return flights with several characteristics:

- Carrier
- Code share
- Aircraft type (mainline jet, regional aircraft, propeller aircraft)
- Total travel time
- Number of transfers
- Fare without tax
- Combination of time of day of out- and inbound flight and return period
- Independence of the connection
## Modelling results

<table>
<thead>
<tr>
<th>Model</th>
<th>MNL Parameter</th>
<th>MNL Robust t-test</th>
<th>MNL with IND(c) Parameter</th>
<th>MNL with IND(c) Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier const.</td>
<td>-0.7465</td>
<td>-12.81</td>
<td>-0.7622</td>
<td>-12.29</td>
</tr>
<tr>
<td>Code share</td>
<td>-0.0006</td>
<td>-5.86</td>
<td>0.0002</td>
<td>1.25</td>
</tr>
<tr>
<td>Travel time</td>
<td>-6.2404</td>
<td>-40.42</td>
<td>-6.6154</td>
<td>-37.75</td>
</tr>
<tr>
<td>N.o. Transfers</td>
<td>-0.0057</td>
<td>-73.71</td>
<td>-0.0063</td>
<td>-66.68</td>
</tr>
<tr>
<td>Similarity</td>
<td>0.0000</td>
<td>0.00</td>
<td>-1.0362</td>
<td>-12.97</td>
</tr>
<tr>
<td>Time of day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aircraft type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional</td>
<td>-0.2170</td>
<td>-8.55</td>
<td>-0.2087</td>
<td>-8.65</td>
</tr>
<tr>
<td>Propeller</td>
<td>-1.6427</td>
<td>-19.04</td>
<td>-1.6711</td>
<td>-18.40</td>
</tr>
</tbody>
</table>
Modelling results for different parameter settings

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter for IND(c)</th>
<th>Average for IND(c)</th>
<th>Final Log-Likelihood</th>
<th>Adjusted r-square</th>
<th>Difference r-square*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNL</td>
<td>--</td>
<td>--</td>
<td>-48816</td>
<td>0.2948</td>
<td>--</td>
</tr>
<tr>
<td>$s_x=720,\gamma=0.25$</td>
<td>-1.1966</td>
<td>0.0386</td>
<td>-48712</td>
<td>0.2963</td>
<td>0.5%</td>
</tr>
<tr>
<td>$s_x=720,\gamma=0.5$</td>
<td>-1.0362</td>
<td>0.0473</td>
<td>-48684</td>
<td>0.2967</td>
<td>0.6%</td>
</tr>
<tr>
<td>$s_x=720,\gamma=0.75$</td>
<td>-0.7210</td>
<td>0.0621</td>
<td>-48752</td>
<td>0.2957</td>
<td>0.3%</td>
</tr>
<tr>
<td>$s_x=120,\gamma=0.5$</td>
<td>-0.1835</td>
<td>0.1650</td>
<td>-48798</td>
<td>0.2950</td>
<td>0.1%</td>
</tr>
<tr>
<td>$s_x=240,\gamma=0.5$</td>
<td>-0.6776</td>
<td>0.0984</td>
<td>-48622</td>
<td>0.2976</td>
<td>0.9%</td>
</tr>
<tr>
<td>$s_x=360,\gamma=0.5$</td>
<td>-0.7666</td>
<td>0.0734</td>
<td>-48640</td>
<td>0.2973</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

* Compared to MNL
Conclusion

Estimated parameters for the independence measure are highly significant and the model performance increases.

Traveller’s perception of similarity between air transport connections is multi-dimensional.

The inclusion of multiple dimensions can lead to more insight in preferences without penalising other alternatives too strong.

Setting of the parameters defining the range of influence has to be determined for specific data set.

The sign of the parameter indicates that tra perceive similar alternatives as positive.
Outlook

For the air connection choice data set:
  • Testing the influence of the individual similarity components
  • Comparison with other model structures, such as (cross)-nested model structures
  • Closer look to the formation of the choice

Investigation of similarity measures based on a GPS data set:
  • Investigation of other choice situations
  • Testing different similarity measures
  • Comparison to complex model structures
The GPS data set

- On-person GPS data for travellers living in Zurich, Winterthur and Geneva
- Complete multi-modal urban and inter-urban trip chains
- GPS records of 4878 persons comprising on average 6.64 days per persons
- No additional information about socio-demographics, trip details etc.
Modelling the GPS data

Motorised private transport route choice

Public transport connection choice

Combined route and mode choice

Destination choice

Combined route, mode and destination choice
Post-processing the GPS records

Adaptation of the method by Tsui and Shalaby (2006) for
- Trip and activity detection
- Identification of single-mode segments
- Mode detection

Map-matching with method by Marchal, Hackney and Axhausen (2006)

Activity type and trip purpose classification

Validation with Swiss Microcensus 2005

Choice set generation
References (1)


References (2)


Reference (3)


Appendices
Available Data – Duration of Stay

Mean = 1.72
Std. Dev. = 2.33
N = 98,065
Available Data – Choice Set Size

Choice set size [#] vs Cumulative percent [%]

- Window 1 hour
- Window 2 hour
- All
- Window 4 hour
Available Data – Comparison Choice Set Fare

[Graph showing the comparison of chosen and non-chosen fares across different fare ranges (0% to 100%) with fare values ranging from 0€ to 1400€.]
Available Data – Comparison Choice Set Fare

![Graph showing fare comparison](image)

- **Chosen Fare**
- **Average Fare Same Day**
  - Lowerbound
  - Upper Bound
- **Average Fare Window 1 hour**
  - Lowerbound
  - Upperbound
Available data – Comparison Choice Set Fare
## Modelling results (2)

<table>
<thead>
<tr>
<th></th>
<th>MNL Parameter</th>
<th>MNL Robust t-test</th>
<th>MNL with IND(c) Parameter</th>
<th>MNL with IND(c) Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Same day return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afternoon Out</td>
<td>-1.6057</td>
<td>-63.39</td>
<td>-1.5867</td>
<td>-61.10</td>
</tr>
<tr>
<td><strong>Overnight stay</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afternoon Out</td>
<td>-0.7807</td>
<td>-24.92</td>
<td>-1.0462</td>
<td>-26.70</td>
</tr>
<tr>
<td>Evening Out</td>
<td>-0.8573</td>
<td>-26.15</td>
<td>-2.0964</td>
<td>-14.97</td>
</tr>
<tr>
<td><strong>Fortnight stay</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afternoon Out</td>
<td>-0.2829</td>
<td>-2.75</td>
<td>-0.5711</td>
<td>-4.85</td>
</tr>
<tr>
<td>Evening Out</td>
<td></td>
<td>not significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Same day return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afternoon In</td>
<td></td>
<td>not significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evening In</td>
<td>0.1907</td>
<td>2.04</td>
<td>0.1958</td>
<td>1.88</td>
</tr>
<tr>
<td><strong>Overnight stay</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afternoon In</td>
<td>1.1967</td>
<td>12.90</td>
<td>1.2171</td>
<td>12.56</td>
</tr>
<tr>
<td>Evening In</td>
<td>1.5915</td>
<td>17.28</td>
<td>1.6055</td>
<td>16.63</td>
</tr>
<tr>
<td><strong>Fortnight stay In</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>