The role of similarities for air connection choice

Nadine Schüssler

IVT
ETH
Zürich

Dezember 2007
Agenda

Background and Modelling Framework

Available Data

Specification

Results

Outlook
Problem and objectives

Two main objectives

- Modelling air transport connection choice

- Investigating a similarity measure for public transport connection choice
Basic assumptions:

Decision-maker \( n \) chooses from choice set \( C_n \) alternative \( i \) with the highest utility \( U_{in} \).

\( U_{in} \) can be represented by a function

\[
U_{in} = V_{in} + \epsilon_{in}
\]

The MNL model estimates for each alternative the probability of being chosen:

\[
P(i|C_n) = P[U_{in} \geq U_{jn}, \forall j \in C_n]
\]

\[
P(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}}
\]
Independence of Irrelevant Alternatives (IIA) property

The relative ratio of the choice probabilities of two alternatives is not effected by the presence or the characteristics of any other alternatives.

\[
\frac{P(i|C_n)}{P(k|C_n)} = \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}} = e^{\mu(V_{in}-V_{kn})}
\]

Similarity between alternatives are not accounted for, even though:

- Travelers are likely not able to distinguish between similar alternatives
- Error terms are not independently distributed when alternatives are similar
Overcoming the IIA property

In general, there are three different ways:

• **Probit models and Mixed Multinomial Logit models**

• **Nested Approaches**

• **Similarity measures**
Similarity Measures – Basic idea

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

\[
U_{in} = V_{in}' + g_{Cin} + \varepsilon_{in}
\]

The similarity of an alternative with other alternatives
- influences its perception as separate alternative
- influences its inclusion in the individual choice set

=> Thus influences its probability to be chosen

\[
P(i|C_n) = \frac{e^{\mu(V_{in}' + \ln g_{Cin})}}{\sum_j e^{\mu(V_{jn}' + \ln g_{Cjn})}}
\]
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
- Independence of a Connection Friedrich et al. (2001)

Spatial Correlations
- Competing Destinations Fotheringham (1988)
- Spatial Dependency Parameter Mohammadian et al. (2005)
- Field Effect Variable Dugundji and Walker (2005)
- Concept of Dominance Cascetta and Papola (2005)
Independence of a Connection

Choice probability depending on the independence of a connection

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

Independence of a connection as reciprocal sum of similarities:

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

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
Case study – 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 led to almost 19,000 observed choices
Observed markets
Duration of stay distribution

Mean = 1.72
Std. Dev. = 2.33
N = 98,966
Duration of stay per weekday

- Monday: 0% 1 day, 15% > 6 days
- Tuesday: 0% 1 day, 15% > 6 days
- Wednesday: 0% 1 day, 15% > 6 days
- Thursday: 0% 1 day, 15% > 6 days
- Friday: 0% 1 day, 15% > 6 days
- Saturday: 0% 1 day, 15% > 6 days
- Sunday: 0% 1 day, 15% > 6 days
Share of bookings

Days before departure
Share of bookings per days before departure

Days before departure
Choice set size

![Graph showing cumulative percent (%) of choice set size over different windows (1 hour, 2 hours, all, 4 hours).](image-url)
Comparison choice set fare

![Graph showing fare comparison](image-url)

- **Chosen Fare**
- **Average Fare Same Day**
  - Lowerbound
  - Upperbound
- **Average Fare Window 1 hour**
  - Lowerbound
  - Upperbound

**Axes**
- **X-axis**: Chosen Fare [€]
- **Y-axis**: Average Available Fare [€]

**Legend**
- Red dashed line: Chosen Fare
- Green dashed line: Average Fare Same Day Lowerbound
- Yellow dashed line: Average Fare Same Day Upperbound
- Red solid line: Average Fare Window 1 hour Lowerbound
- Light blue solid line: Average Fare Window 1 hour Upperbound
Departure time distribution – same day return
Specification of the independency measure

Choice probability depending on the independence of a connection

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

Independence of a connection as reciprocal sum of similarities:

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

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
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symmetric setting</th>
<th>Extreme values</th>
<th>$\gamma = 0.25$</th>
<th>$\gamma = 0.75$</th>
<th>$\gamma = 0.5$</th>
<th>Final model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_x$</td>
<td>720</td>
<td>360</td>
<td>720</td>
<td>720</td>
<td>720</td>
<td>720</td>
</tr>
<tr>
<td>$s_y^+$</td>
<td>780</td>
<td>180</td>
<td>780</td>
<td>780</td>
<td>780</td>
<td>780</td>
</tr>
<tr>
<td>$s_y^-$</td>
<td>780</td>
<td>120</td>
<td>540</td>
<td>540</td>
<td>540</td>
<td>540</td>
</tr>
<tr>
<td>$s_z^+$</td>
<td>1.7</td>
<td>0.9</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>$s_z^-$</td>
<td>1.7</td>
<td>0.9</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.25</td>
<td>0.75</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>
Effect of varying $s_x$, $s_y$ and $s_z$
Effect of varying $\gamma$
## Valuation of service attributes

<table>
<thead>
<tr>
<th>Valuation</th>
<th>Same day</th>
<th>Next day</th>
<th>&gt; 6 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code share [€]</td>
<td>-119.39</td>
<td>-165.03</td>
<td>-110.84</td>
</tr>
<tr>
<td>Regional aircraft* [€]</td>
<td>-19.82</td>
<td>-27.40</td>
<td>-18.40</td>
</tr>
<tr>
<td>Propeller aircraft* [€]</td>
<td>-201.04</td>
<td>-277.90</td>
<td>-186.64</td>
</tr>
<tr>
<td>Travel time [€/h]</td>
<td>-90.50</td>
<td>-125.09</td>
<td>-84.02</td>
</tr>
<tr>
<td>Transfer [€]</td>
<td>-602.52</td>
<td>-832.92</td>
<td>-559.41</td>
</tr>
<tr>
<td>Transfer [min]</td>
<td>400.96</td>
<td>400.96</td>
<td>400.96</td>
</tr>
</tbody>
</table>

(*) base category: Mainline jet
Departure hour preferences – Same day return

![Graph showing utility vs. departure hour]

- Blue line: Outbound departure time - Dummy
- Orange line: Inbound departure time - Dummy
Departure hour preferences – Next day return

![Graph showing departure hour preferences with utility values for different hours of the day, indicating preferences for outbound and inbound departures.](Graph.png)
Returning after 6 days
Valuation of outbound departure hour

![Graph showing the valuation of outbound departure hour with different lines for duration of stay 0 days, duration of stay 1 day, and duration of stay > 6 days. The graph plots valuation in € against departure time with values ranging from -800 to 0.]
**Parameter estimates with similarity measure**

<table>
<thead>
<tr>
<th>Variable</th>
<th>MNL model Parameter</th>
<th>MNL with IND(c) Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code share</td>
<td>-0.9247 (-12.84)</td>
<td>-0.9674 (-13.54)</td>
</tr>
<tr>
<td>Regional aircraft*</td>
<td>-0.1328 (-4.68)</td>
<td>-0.1236 (-4.31)</td>
</tr>
<tr>
<td>Propeller aircraft*</td>
<td>1.5388 (-14.79)</td>
<td>-1.4719 (-12.19)</td>
</tr>
<tr>
<td>Total travel time out</td>
<td>-0.0114 (-5.47)</td>
<td>-0.0105 (-5.13)</td>
</tr>
<tr>
<td>Transfer</td>
<td>-4.6923 (-12.42)</td>
<td>-4.6494 (-12.19)</td>
</tr>
<tr>
<td>Fare</td>
<td>-0.0068 (-75.79)</td>
<td>-0.0066 (-70.67)</td>
</tr>
<tr>
<td>Independence Measure</td>
<td>--</td>
<td>0.5379 (13.59)</td>
</tr>
</tbody>
</table>

(*) base category: Mainline jet
Conclusions

- Similar alternatives are perceived as negative
- Alternatives departing in a window of one hour are perceived as most similar
- In direct itineraries, fare and departure time play a large role
- In non-direct itineraries, transfers play an important role
- Fare and departure time sensitivity varies per duration of stay
Outlook

- Testing the influence of the individual similarity components
- More advanced model structures to overcome similarity problem
- Investigate further choice situations
References (1)


References (2)


# Modelling performance 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>-46101.7</td>
<td>0.3308</td>
<td>--</td>
</tr>
<tr>
<td>(s_x=720, \gamma=0.25)</td>
<td>0.1608</td>
<td>0.0388</td>
<td>-46100.5</td>
<td>0.3308</td>
<td>0.00%</td>
</tr>
<tr>
<td>(s_x=720, \gamma=0.5)</td>
<td>0.2194</td>
<td>0.0458</td>
<td>-46099.1</td>
<td>0.3308</td>
<td>0.01%</td>
</tr>
<tr>
<td>(s_x=720, \gamma=0.75)</td>
<td>0.3992</td>
<td>0.0626</td>
<td>-46089.2</td>
<td>0.3309</td>
<td>0.03%</td>
</tr>
<tr>
<td>(s_x=120, \gamma=0.5)</td>
<td>0.5379</td>
<td>0.1660</td>
<td>-46004.2</td>
<td>0.3322</td>
<td>0.21%</td>
</tr>
<tr>
<td>(s_x=240, \gamma=0.5)</td>
<td>0.2210</td>
<td>0.0983</td>
<td>-46089.6</td>
<td>0.3309</td>
<td>0.03%</td>
</tr>
<tr>
<td>(s_x=360, \gamma=0.5)</td>
<td>0.1804</td>
<td>0.0732</td>
<td>-46095.2</td>
<td>0.3308</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

* Compared to MNL