Mathematical models of behavior for the prediction of demand in transportation and energy

Michel Bierlaire

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

July 11, 2017
Demand models in transportation

- Supply = infrastructure
- Demand = behavior, choices
- Congestion = mismatch
Aggregate demand

- Homogeneous population
- Identical behavior
- Price ($P$) and quantity ($Q$)
- Demand functions: $P = f(Q)$
- Inverse demand: $Q = f^{-1}(P)$
Disaggregate demand

- Heterogeneous population
- Different behaviors
- Many variables:
  - Attributes: price, travel time, reliability, frequency, etc.
  - Characteristics: age, income, education, etc.
- Complex demand/inverse demand functions.
Outline

1. Demand and supply
2. Disaggregate demand models
3. Data
4. Prediction
5. Energy
Choice models

Behavioral models

- Demand = sequence of choices
- Choosing means trade-offs
- In practice: derive trade-offs from choice models
Choice models

Theoretical foundations

- Random utility theory
- Choice set: $C_n$
- $y_{in} = 1$ if $i \in C_n$, 0 if not
- Logit model:

$$P(i|C_n) = \frac{y_{in}e^{V_{in}}}{\sum_{j \in C} y_{jn}e^{V_{jn}}}$$
Logit model

Utility

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

- Decision-maker \( n \)
- Alternative \( i \in C_n \)

Choice probability

\[ P_n(i|C_n) = \frac{y_{in}e^{V_{in}}}{\sum_{j \in C} y_{jn}e^{V_{jn}}} \]
Variables: $x_{in} = (z_{in}, s_n)$

Attributes of alternative $i$: $z_{in}$
- Cost / price
- Travel time
- Waiting time
- Level of comfort
- Number of transfers
- Late/early arrival
- etc.

Characteristics of decision-maker $n$: $s_n$
- Income
- Age
- Sex
- Trip purpose
- Car ownership
- Education
- Profession
- etc.
Disaggregate demand models

Demand curve

Disaggregate model

\[ P_n(i|c_{in}, z_{in}, s_n) \]

Total demand

\[ D(i) = \sum_n P_n(i|c_{in}, z_{in}, s_n) \]

Difficulty

Non linear and non convex in \( c_{in} \) and \( z_{in} \)
Outline

1. Demand and supply
2. Disaggregate demand models
3. Data
4. Prediction
5. Energy
Revealed preference

Observe actual behavior

- Representative sample (possibly biased)
- Collect
  - socio-economic characteristics
  - attributes of the alternatives
  - choice
- At one point in time: cross-sectional data
- Several times: panel data
Revealed preference

Advantages
- Real life choices
- Possibility to replicate market shares
- Decision-makers have to assume their choice
- Real constraints involved

Disadvantages
- Limited to existing alternatives and variables
- Lack of variability of some attributes
- No info on non chosen alternatives
- High level of correlation
- Data collection cost
- In general, one individual = one observation
## Stated preference

**Pick Your Preferred Flight**

Three flight options are described for your trip from Chicago to San Diego. These are options that might be available on this route or might be new options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that everything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Non-Stop (Option 1)</th>
<th>1 Stop (Option 2)</th>
<th>1 Stop (Option 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Departure time (local)</td>
<td>6:00 PM</td>
<td>4:30 PM</td>
<td>6:00 PM</td>
</tr>
<tr>
<td>Arrival time (local)</td>
<td>8:14 PM</td>
<td>8:44 PM</td>
<td>9:44 PM</td>
</tr>
<tr>
<td>Total time in air</td>
<td>4 hr 14 min</td>
<td>4 hr 44 min</td>
<td>4 hr 44 min</td>
</tr>
<tr>
<td>Total trip time</td>
<td>4 hr 14 min</td>
<td>6 hr 14 min</td>
<td>5 hr 44 min</td>
</tr>
<tr>
<td>Legroom</td>
<td>typical legroom</td>
<td>2-in more of legroom</td>
<td>4-in more of legroom</td>
</tr>
<tr>
<td>Fare</td>
<td>$565</td>
<td>$485</td>
<td>$620</td>
</tr>
</tbody>
</table>

1. Which is MOST attractive?  
   - Option 1  
   - Option 2  
   - Option 3

2. Which is LEAST attractive?  
   - Option 1  
   - Option 2  
   - Option 3

3. If these were the ONLY three options available, I would NOT take this trip by air.  
   - Yes  
   - No
Stated preference

Advantages

- Exploring new alternatives, attributes and attributes levels
- Control of the attributes variability
- Control on all alternatives
- Control on the level of correlation
- One individual can answer several questions
**Stated preference**

**Disadvantages**

- Hypothetical situations
- Cannot be used for market shares
- Decision-makers do not have to assume their choice
- “A bike or a Ferrari?” — “A Ferrari, of course!”
- Real constraints not involved
- Credibility
- Valid within the range of the experimental design
- Policy bias (“every body else should take the bus”)
- Justification bias (or inertia)
- Framing: phrasing of the question matters
- Anchoring: one variable explains it all
- Fatigue effect
Outline

1. Demand and supply
2. Disaggregate demand models
3. Data
4. Prediction
5. Energy
Market shares

Sample

- Revealed preference data
- Survey conducted between 2009 and 2010 for PostBus
- Questionnaires sent to people living in rural areas
- Each observation corresponds to a sequence of trips from home to home.
- Sample size: 1723

Model: 3 alternatives

- Car
- Public transportation (PT)
- Slow mode
### Example: interurban mode choice in Switzerland

<table>
<thead>
<tr>
<th>Parameter number</th>
<th>Description</th>
<th>Coeff. estimate</th>
<th>Robust Asympt. std. error</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cte. (PT)</td>
<td>0.977</td>
<td>0.605</td>
<td>1.61</td>
<td>0.11</td>
</tr>
<tr>
<td>2</td>
<td>Income 4-6 KCHF (PT)</td>
<td>-0.934</td>
<td>0.255</td>
<td>-3.67</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>Income 8-10 KCHF (PT)</td>
<td>-0.123</td>
<td>0.175</td>
<td>-0.70</td>
<td>0.48</td>
</tr>
<tr>
<td>4</td>
<td>Age 0-45 (PT)</td>
<td>-0.0218</td>
<td>0.00977</td>
<td>-0.70</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>Age 45-65 (PT)</td>
<td>0.0303</td>
<td>0.0124</td>
<td>2.44</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>Male dummy (PT)</td>
<td>-0.351</td>
<td>0.260</td>
<td>-1.35</td>
<td>0.18</td>
</tr>
<tr>
<td>7</td>
<td>Marginal cost [CHF] (PT)</td>
<td>-0.0105</td>
<td>0.0104</td>
<td>-1.01</td>
<td>0.31</td>
</tr>
<tr>
<td>8</td>
<td>Waiting time [min], if full time job (PT)</td>
<td>-0.0440</td>
<td>0.0117</td>
<td>-3.76</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>Waiting time [min], if part time job or other occupation (PT)</td>
<td>-0.0268</td>
<td>0.00742</td>
<td>-3.62</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>Travel time [min] × \log(1+ distance[km]) / 1000, if full time job</td>
<td>-1.52</td>
<td>0.510</td>
<td>-2.98</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>Travel time [min] × \log(1+ distance[km]) / 1000, if part time job</td>
<td>-1.14</td>
<td>0.671</td>
<td>-1.69</td>
<td>0.09</td>
</tr>
<tr>
<td>12</td>
<td>Season ticket dummy (PT)</td>
<td>2.89</td>
<td>0.346</td>
<td>8.33</td>
<td>0.00</td>
</tr>
<tr>
<td>13</td>
<td>Half fare travelcard dummy (PT)</td>
<td>0.360</td>
<td>0.177</td>
<td>2.04</td>
<td>0.04</td>
</tr>
<tr>
<td>14</td>
<td>Line related travelcard dummy (PT)</td>
<td>2.11</td>
<td>0.281</td>
<td>7.51</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>Area related travelcard (PT)</td>
<td>2.78</td>
<td>0.266</td>
<td>10.46</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>Other travel cards dummy (PT)</td>
<td>1.25</td>
<td>0.303</td>
<td>4.14</td>
<td>0.00</td>
</tr>
</tbody>
</table>
**Example: interurban mode choice in Switzerland**

<table>
<thead>
<tr>
<th>Parameter number</th>
<th>Description</th>
<th>Coeff. estimate</th>
<th>Robust Asympt. std. error</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Cte. (Car)</td>
<td>0.792</td>
<td>0.512</td>
<td>1.55</td>
<td>0.12</td>
</tr>
<tr>
<td>18</td>
<td>Income 4-6 KCHF (Car)</td>
<td>-1.02</td>
<td>0.251</td>
<td>-4.05</td>
<td>0.00</td>
</tr>
<tr>
<td>19</td>
<td>Income 8-10 KCHF (Car)</td>
<td>-0.422</td>
<td>0.223</td>
<td>-1.90</td>
<td>0.06</td>
</tr>
<tr>
<td>20</td>
<td>Income 10 KCHF and more (Car)</td>
<td>0.126</td>
<td>0.0697</td>
<td>1.81</td>
<td>0.07</td>
</tr>
<tr>
<td>21</td>
<td>Male dummy (Car)</td>
<td>0.291</td>
<td>0.229</td>
<td>1.27</td>
<td>0.20</td>
</tr>
<tr>
<td>22</td>
<td>Number of cars in household (Car)</td>
<td>0.939</td>
<td>0.135</td>
<td>6.93</td>
<td>0.00</td>
</tr>
<tr>
<td>23</td>
<td>Gasoline cost [CHF], if trip purpose HWH (Car)</td>
<td>-0.164</td>
<td>0.0369</td>
<td>-4.45</td>
<td>0.00</td>
</tr>
<tr>
<td>24</td>
<td>Gasoline cost [CHF], if trip purpose other (Car)</td>
<td>-0.0727</td>
<td>0.0224</td>
<td>-3.24</td>
<td>0.00</td>
</tr>
<tr>
<td>25</td>
<td>Gasoline cost [CHF], if male (Car)</td>
<td>-0.0683</td>
<td>0.0240</td>
<td>-2.84</td>
<td>0.00</td>
</tr>
<tr>
<td>26</td>
<td>French speaking (Car)</td>
<td>0.926</td>
<td>0.190</td>
<td>4.88</td>
<td>0.00</td>
</tr>
<tr>
<td>27</td>
<td>Distance [km] (Slow modes)</td>
<td>-0.184</td>
<td>0.0473</td>
<td>-3.90</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Summary statistics**

Number of observations = 1723

Number of estimated parameters = 27

\[
\begin{align*}
\mathcal{L}(\beta_0) &= -1858.039 \\
\mathcal{L}(\hat{\beta}) &= -792.931 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 2130.215 \\
\rho^2 &= 0.573 \\
\bar{\rho}^2 &= 0.559
\end{align*}
\]
Forecast

Disaggregate model for individual $n$

- $P_n(\text{Car}|z_{\text{Auto},n}, S_n)$
- $P_n(\text{PT}|z_{\text{PT},n}, S_n)$
- $P_n(\text{SM}|z_{\text{MD},n}, S_n)$

Total number of passengers in public transportation

$$\sum_n P_n(\text{PT}|z_{\text{PT},n}, S_n)$$

Market shares of public transportation (population size: $N$)

$$\frac{\sum_n P_n(\text{PT}|z_{\text{PT},n}, S_n)}{N}$$
Switzerland: current shares from the model
Forecast

Scenario: increase of fuel price

Market shares

Current 5% 10% 15% 20% 25% 30%

Car
Slow modes
PT
Other indicators

**CO₂ emissions**

\[ \sum_n P_n(\text{Car}|z_{\text{Car},n}, S_n)\text{CO₂(\text{Car type}_n)} \]

**Total travel time**

\[ \sum_n P_n(\text{Car}|z_{\text{Car},n}, S_n)\text{Time}_{\text{Car},n} + P_n(\text{PT}|z_{\text{PT},n}, S_n)\text{Time}_{\text{PT},n} \]

**Total income**

\[ \sum_n P_n(\text{PT}|z_{\text{PT},n}, S_n)\text{Cost}_{\text{PT},n} \]
# Willingness to pay

## Per trip purpose

<table>
<thead>
<tr>
<th></th>
<th>Business</th>
<th>Commute</th>
<th>Leasure</th>
<th>Shopping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transport (CHF/h)</td>
<td>49.57</td>
<td>27.81</td>
<td>21.84</td>
<td>17.73</td>
</tr>
<tr>
<td>Car (CHF/h)</td>
<td>50.23</td>
<td>30.64</td>
<td>29.20</td>
<td>24.32</td>
</tr>
</tbody>
</table>

[Axhausen et al., 2008]
Outline

1. Demand and supply
2. Disaggregate demand models
3. Data
4. Prediction
5. Energy
Discrete choice in energy

Transportation related

- [Bunch et al., 1993]
- [Horne et al., 2005]
- [Ziegler, 2012]

Not transportation

- Washing machines [Sammer and Wüstenhagen, 2006]
- Household demand [Vaage, 2000]
- Space heating [Nesbakken, 2001], [Michelsen and Madlener, 2012]
- Willingness to pay for energy saving (Filippini, ETHZ) [Banfi et al., 2008]
- Fuel choice (Filippini, ETHZ)[Farsi et al., 2007]
Electricity

- Demand = supply
- Time series
Renewable energy

- Uncertain supply
- Necessity to understand demand
- Causal effects instead of time series.


Bibliography II


