Demand forecasting with discrete choice models

Michel Bierlaire

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

January 3, 2017
Introduction

Demand = choices
- Travel or not?
- Destination
- Mode
- Route
- Departure time
Introduction

Choice model

\[ P(i|x_n, C_n; \theta) \]

What do we do with it?

Note

It is always possible to characterize the choice set using availability variables, included into \( x_n \). So the model can be written

\[ P(i|x_n, C; \theta) = P(i|x_n; \theta) \]

Example: logit

\[
\begin{align*}
P(i|x_n, C_n; \theta) &= \frac{e^{V_{in}(x_n;\theta)}}{\sum_{j \in C_n} e^{V_{jn}(x_n;\theta)}} \\
P(i|x_n; \theta) &= \frac{y_{in} e^{V_{in}(x_n;\theta)}}{\sum_{j \in C} y_{jn} e^{V_{jn}(x_n;\theta)}}
\end{align*}
\]
Outline

1. Introduction
2. Aggregation
   - Sample enumeration
3. Forecasting
4. Price optimization
5. Confidence intervals
6. Willingness to pay
   - Value of time
7. Elasticities
8. Consumer surplus
9. Summary
Aggregation

Aggregate shares

- Prediction about a single individual is of little use in practice.
- Need for indicators about aggregate demand.
- Typical application: aggregate market shares.
Population

- Identify the population $T$ of interest (in general, already done during the phase of the model specification and estimation).
- Obtain $x_n$ and $C_n$ for each individual $n$ in the population.
- The number of individuals choosing alternative $i$ is

$$N_T(i) = \sum_{n=1}^{N_T} P_n(i|x_n; \theta).$$

- The share of the population choosing alternative $i$ is

$$W(i) = \frac{1}{N_T} \sum_{n=1}^{N_T} P(i|x_n; \theta) = \mathbb{E} \left[ P(i|x_n; \theta) \right].$$
### Aggregation

<table>
<thead>
<tr>
<th>Population</th>
<th>Alternatives</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>$P(1</td>
<td>x_1; \theta)$</td>
</tr>
<tr>
<td>2</td>
<td>$P(1</td>
<td>x_2; \theta)$</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>$N_T$</td>
<td>$P(1</td>
<td>x_{N_T}; \theta)$</td>
</tr>
<tr>
<td>Total</td>
<td>$N_T(1)$</td>
<td>$N_T(2)$</td>
</tr>
</tbody>
</table>
Distribution

Data

- Assume the distribution of $x_n$ is available.
- $x_n = (x_n^C, x_n^D)$ is composed of discrete and continuous variables.
- $x_n^C$ distributed with pdf $p^C(x)$.
- $x_n^D$ distributed with pmf $p^D(x)$.

Market shares

$$W(i) = \sum x_n^D \int x_n^C P_n(i|x_n^C, x_n^D) p^C(x_n^C) p^D(x_n^D) dx_n^C = E [P_n(i|x_n; \theta)],$$
Aggregation methods

Issues
- None of the above formulas can be applied in practice.
- No full access to each $x_n$, or to their distribution.
- Practical methods are needed.

Practical methods
- Use a sample.
- It must be revealed preference data.
  - $x_n$ must reflect a specific, well identified, real scenario
  - Observed market shares are used to calibrate the constants
- It may be the same sample as for estimation.
Sample enumeration

Stratified sample

- Population is partitioned into homogenous segments.
- Each segment is randomly sampled.
- Let $n$ be an observation in the sample belonging to segment $g$
- Let $\omega_g$ be the weight of segment $g$, that is
  \[
  \omega_g = \frac{N_g}{S_g} = \frac{\text{# persons in segment } g \text{ in population}}{\text{# persons in segment } g \text{ in sample}}
  \]

- The number of persons choosing alt. $i$ is estimated by
  \[
  \hat{N}(i) = \sum_{n \in \text{sample}} P(i|x_n; \theta) \sum_g \omega_g l_{ng} = \sum_n \omega_g(n) P(i|x_n; \theta)
  \]

  where $l_{ng} = 1$ if individual $n$ belongs to segment $g$, 0 otherwise, and $g(n)$ is the segment containing $n$. 
Sample enumeration

Predicted shares

\[ \hat{W}(i) = \sum_{n \in \text{sample}} P(i|x_n; \theta) \sum_{g} \frac{N_g}{N_T} \frac{1}{S_g} I_{ng} = \frac{1}{N_T} \sum_{n} \omega_g(n) P(i|x_n; \theta) \]

Comments

- Consistent estimate.
- Estimate subject to sampling errors.
- Policy analysis: change the values of the explanatory variables, and apply the same procedure.
Market shares per market segment

Let $h$ be a segment of the population.

Let $I_{nh} = 1$ if individual $n$ belongs to this segment, 0 otherwise.

Number of persons of segment $h$ choosing alternative $i$

$$\hat{N}_h(i) = \sum_n \omega_g(n) P(i|x_n; \theta) I_{nh}$$

Market share of alternative $i$ in segment $h$

$$\hat{W}_h(i) = \frac{\sum_n \omega_g(n) P(i|x_n; \theta) I_{nh}}{\sum_n \omega_g(n) I_{nh}}.$$
Example: interurban mode choice in Switzerland

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. estimate</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

\[
\mathcal{L}(\beta_0) = -1858.039
\]

\[
\mathcal{L}(\hat{\beta}) = -792.931
\]

\[
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 2130.215
\]

\[
\rho^2 = 0.573
\]

\[
\bar{\rho}^2 = 0.559
\]
### Example: interurban mode choice in Switzerland

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Unknown gender</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>64.96%</td>
<td>60.51%</td>
<td>70.88%</td>
<td>62.8%</td>
</tr>
<tr>
<td>PT</td>
<td>30.20%</td>
<td>32.52%</td>
<td>25.59%</td>
<td>31.3%</td>
</tr>
<tr>
<td>Slow modes</td>
<td>4.83%</td>
<td>6.96%</td>
<td>3.53%</td>
<td>5.88%</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Aggregation
   - Sample enumeration
3. Forecasting
4. Price optimization
5. Confidence intervals
6. Willingness to pay
   - Value of time
7. Elasticities
8. Consumer surplus
9. Summary
Forecasting

Procedure

- Scenarios: specify future values of the variables of the model.
- Recalculate the market shares.

Market shares

<table>
<thead>
<tr>
<th></th>
<th>Increase of the cost of gasoline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Now</td>
</tr>
<tr>
<td>Car</td>
<td>62.8%</td>
</tr>
<tr>
<td>PT</td>
<td>31.3%</td>
</tr>
<tr>
<td>Slow modes</td>
<td>5.88%</td>
</tr>
</tbody>
</table>
Forecasting

Scenario: increase of the cost of gasoline

- Car
- Public transportation
- Slow modes

Market share

Base model, 5%, 10%, 15%, 20%, 25%, 30%
Price optimization

Optimizing the price of product $i$ is solving the problem

$$\max_{p_i} \sum_{n \in \text{sample}} \omega_g(n) P(i|x_n, p_i; \theta)$$

Notes:

- It assumes that everything else is equal
- In practice, it is likely that the competition will also adjust the prices
A binary logit model with

\[ V_1 = \beta_p p_1 - 0.5 \]
\[ V_2 = \beta_p p_2 \]

so that

\[ P(1|p) = \frac{e^{\beta_p p_1 - 0.5}}{e^{\beta_p p_1 - 0.5} + e^{\beta_p p_2}} \]

Two groups in the population:

- Group 1: \( \beta_p = -2, N_s = 600 \)
- Group 2: \( \beta_p = -0.1, N_s = 400 \)

Assume that \( p_2 = 2 \).
Illustrative example
Scenario

- A uniform adjustment of the marginal cost of public transportation is investigated.
- The analysis ranges from 0% to 700%.
- What is the impact on the market share of public transportation?
- What is the impact of the revenues for public transportation operators?
Case study: interurban mode choice in Switzerland

Adjustment of the marginal cost

Revenues

Market shares

Share of public transportation

Revenues for public transportation

0%
20%
40%
60%
80%
100%

0%
100%
200%
300%
400%
500%
600%
700%
800%
900%
1000%
1100%
1200%
1300%
1400%

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Case study: interurban mode choice in Switzerland

Comments

- Typical non concavity of the revenue function due to taste heterogeneity.
- In general, decision making is more complex than optimizing revenues.
- Applying the model with values of $x$ very different from estimation data may be highly unreliable.
Outline

1. Introduction
2. Aggregation
   - Sample enumeration
3. Forecasting
4. Price optimization
5. Confidence intervals
6. Willingness to pay
   - Value of time
7. Elasticities
8. Consumer surplus
9. Summary
Confidence intervals

Model

\[ P(i|x_n, p_i; \theta) \]

- In reality, we use \( \hat{\theta} \), the maximum likelihood estimate of \( \theta \)
- Property: the estimator is normally distributed \( N(\hat{\theta}, \hat{\Sigma}) \)

Calculating the confidence interval by simulation

- Draw \( R \) times \( \tilde{\theta} \) from \( N(\hat{\theta}, \hat{\Sigma}) \).
- For each \( \tilde{\theta} \), calculate the requested quantity (e.g. market share, revenue, etc.) using \( P(i|x_n, p_i; \tilde{\theta}) \)
- Calculate the 5% and the 95% quantiles of the generated quantities.
- They define the 90% confidence interval.
Case study: confidence intervals (500 draws)
Case study: confidence intervals (500 draws)
Confidence interval

Model

\[ P(i|x_n, p_i; \hat{\theta}) \]

- There are also errors in the \( x_n \).
- If the distribution of \( x_n \) is known, draw from both \( x_n \) and \( \theta \).
- Apply the same procedure.
Willingness to pay

Context
- If the model contains a cost or price variable,
- it is possible to analyze the trade-off between any variable and money.
- It reflects the willingness of the decision maker to pay for a modification of another variable of the model.
- Typical example in transportation: value of time

Value of time
Price that travelers are willing to pay to decrease the travel time.
Willingness to pay

Definition

- Let $c_{in}$ be the cost of alternative $i$ for individual $n$.
- Let $x_{in}$ be the value of another variable of the model (travel time, say).
- Let $V_{in}(c_{in}, x_{in})$ be the value of the utility function.
- Consider a scenario where the variable under interest takes the value $x_{in}' = x_{in} + \delta_{in}^x$.
- We denote by $\delta_{in}^c$ the additional cost that would achieve the same utility, that is

$$V_{in}(c_{in} + \delta_{in}^c, x_{in} + \delta_{in}^x) = V_{in}(c_{in}, x_{in}).$$

- The willingness to pay is the additional cost per unit of $x$, that is

$$\frac{\delta_{in}^c}{\delta_{in}^x}.$$
Continuous variable

- If $x_{in}$ is continuous,
- if $V_{in}$ is differentiable in $x_{in}$ and $c_{in}$,
- invoke Taylor’s theorem:

$$V_{in}(c_{in}, x_{in}) = V_{in}(c_{in} + \delta^c_{in}, x_{in} + \delta^x_{in})$$

$$\approx V_{in}(c_{in}, x_{in}) + \delta^c_{in} \frac{\partial V_{in}}{\partial c_{in}}(c_{in}, x_{in}) + \delta^x_{in} \frac{\partial V_{in}}{\partial x_{in}}(c_{in}, x_{in})$$

$$\frac{\delta^c_{in}}{\delta^x_{in}} = -\frac{\frac{\partial V_{in}}{\partial x_{in}}(c_{in}, x_{in})}{\frac{\partial V_{in}}{\partial c_{in}}(c_{in}, x_{in})}$$
Willingness to pay

Linear utility function

- If $x_{in}$ and $c_{in}$ appear linearly in the utility function, that is
  \[
  V_{in}(c_{in}, x_{in}) = \beta_c c_{in} + \beta_x x_{in} + \cdots
  \]

- then the willingness to pay is
  \[
  \frac{\delta c}{\delta x} = -\frac{(\partial V_{in}/\partial x_{in})(c_{in}, x_{in})}{(\partial V_{in}/\partial c_{in})(c_{in}, x_{in})} = -\frac{\beta_x}{\beta_c}
  \]
The value of time is defined as the amount of money that an individual is willing to pay to save one unit of time

\[
\frac{\delta^c_{in}}{\delta^t_{in}} = \frac{\delta^c_{in}}{-1} = \frac{\partial V_{in}/\partial t_{in}}{\partial V_{in}/\partial c_{in}}(c_{in}, t_{in}) = -\frac{\beta_t}{\beta_c}
\]

Therefore

\[
VOT_{in} = \frac{\delta^c_{in}}{(-\delta^t_{in})} = \frac{\partial V_{in}/\partial t_{in}}{\partial V_{in}/\partial c_{in}}(c_{in}, t_{in})
\]

If \( V \) is linear in these variables, we have

\[
VOT_{in} = \frac{\delta^c_{in}}{(-\delta^t_{in})} = \frac{\beta_t}{\beta_c}.
\]
Willingness to pay

Case study: value of time for car drivers

M. Bierlaire (TRANSP-OR ENAC EPFL)
Case study: value of time for car drivers (nonzero)
Case study: value of time for public transportation
Case study: value of time for public transportation (nonzero)
Outline

1. Introduction
2. Aggregation
   - Sample enumeration
3. Forecasting
4. Price optimization
5. Confidence intervals
6. Willingness to pay
   - Value of time
7. Elasticities
8. Consumer surplus
9. Summary
### Disaggregate elasticities

#### Point vs. arc
- **Point**: marginal rate
- **Arc**: between two values

#### Direct vs. cross
- **Direct**: wrt attribute of the same alternative
- **Cross**: wrt attribute of another alternative

<table>
<thead>
<tr>
<th></th>
<th>Point</th>
<th>Arc</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct</strong></td>
<td>$E_{x_{ink}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} \frac{x_{ink}}{P_n(i)}$</td>
<td>$\frac{\Delta P_n(i)}{\Delta x_{ink}} \frac{x_{ink}}{P_n(i)}$</td>
</tr>
<tr>
<td><strong>Cross</strong></td>
<td>$E_{x_{jnk}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_{jnk}} \frac{x_{jnk}}{P_n(i)}$</td>
<td>$\frac{\Delta P_n(i)}{\Delta x_{jnk}} \frac{x_{jnk}}{P_n(i)}$</td>
</tr>
</tbody>
</table>
Aggregate elasticities

Population share

\[ W(i) = \frac{1}{N_T} \sum_{n=1}^{N_T} P(i|x_n) \]

Aggregate elasticity

\[ E_{x_{jk}}^{W(i)} = \frac{\partial W(i)}{\partial x_{jk}} \frac{x_{jk}}{W(i)} = \sum_{n=1}^{N_T} \frac{P_n(i)}{P_n(i)} \frac{\partial P_n(i)}{\partial x_{jk}} \frac{x_{jk}}{\sum_{n=1}^{N_T} P_n(i)} \]

\[ = \sum_{n=1}^{N_T} \frac{P_n(i)}{\sum_{n=1}^{N_T} P_n(i)} E_{x_{jk}}^{P_n(i)} \]
Case study: elasticity of travel time (PT)

Elasticities

Elasticity of travel time (PT)

Elasticity of travel time (PT)

Share of the population (%) vs. Elasticity of travel time (PT)

M. Bierlaire (TRANSP-OR ENAC EPFL)
Case study: elasticity of travel time (PT, non zero)
Consumer surplus

Concept
- Difference between what a consumer is willing to pay for a good and what she actually pays for the good
- Area under the demand curve and above the market price

Discrete choice
- Demand characterized by the choice probability
- Role of price taken by the utility
- Utility can always be transformed into monetary units
Consumer surplus

\[ \int_{V_i^1}^{V_i^2} P(i|V_i, V_j) dV_i = \int_{V_i^1}^{V_i^2} \frac{e^{\mu V_i}}{e^{\mu V_i} + e^{\mu V_j}} dV_i \]

Demand curve
Consumer surplus at current situation
Additional consumer surplus
Consumer surplus

Binary logit

\[
\int_{V_i^1}^{V_i^2} P(i|V_i, V_j) dV_i = \int_{V_i^1}^{V_i^2} \frac{e^{\mu V_i}}{e^{\mu V_i} + e^{\mu V_j}} dV_i = \frac{1}{\mu} \ln(e^{\mu V_i^2} + e^{\mu V_j}) - \frac{1}{\mu} \ln(e^{\mu V_i^1} + e^{\mu V_j}).
\]
Consumer surplus

Generalization

\[
\sum_{i \in C} \int_{V_1}^{V_2} P(i \mid V) dV_i.
\]

If the choice model has equal cross derivatives, that is

\[
\frac{\partial P(i \mid V, C)}{\partial V_j} = \frac{\partial P(j \mid V, C)}{\partial V_i}, \quad \forall i, j \in C,
\]

the integral is path independent.

Logit

\[
\sum_{i \in C} \int_{V_1}^{V_2} P(i \mid V) dV_i = \frac{1}{\mu} \ln \sum_{j \in C^2} e^{\mu V_j^2} - \frac{1}{\mu} \ln \sum_{j \in C^1} e^{\mu V_j^1}.
\]
Summary

Aggregation
- Sample enumeration

Precision
- Confidence intervals
- Calculated by simulation

Indicators
- Market shares
- Revenues
- Willingness to pay
- Elasticities
- Consumer surplus