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

Outline

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

Introduction

Demand = choices

- Travel or not?
- Destination
- Mode
- Route
- Departure time

Introduction

Choice model

$$P(i|x_n, \mathcal{C}_n; \theta)$$

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

$$P(i|x_n, \mathcal{C}; \theta) = P(i|x_n; \theta)$$

Example: logit

$$P(i|x_n, \mathcal{C}_n; \theta) = \frac{e^{V_{in}(x_n;\theta)}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}(x_n;\theta)}} \qquad P(i|x_n; \theta) = \frac{y_{in} e^{V_{in}(x_n;\theta)}}{\sum_{j \in \mathcal{C}} y_{jn} e^{V_{jn}(x_n;\theta)}}$$

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.

Aggregation

Population

- Identify the population T of interest (in general, already done during the phase of the model specification and estimation).
- Obtain x_n and $\mathcal{C}_n$ 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

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

Aggregation

Population	Alternatives				Total
	1	2	$\cdots$	J	
1	$P(1\|x_1;\theta)$	$P(2\|x_1;\theta)$	$\cdots$	$P(J\|x_1;\theta)$	1
2	$P(1\|x_2;\theta)$	$P(2\|x_2;\theta)$	$\cdots$	$P(J\|x_2;\theta)$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N_T	$P(1\|x_{N_T};\theta)$	$P(2\|x_{N_T};\theta)$	$\cdots$	$P(J\|x_{N_T};\theta)$	1
Total	$N_T(1)$	$N_T(2)$	$\cdots$	$N_T(J)$	N_T

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

$$W(i) = \sum_{x^D} \int_{x^C} P_n(i|x^C, x^D) p^C(x^C) p^D(x^d) dx^C = E\left[P_n(i|x_n;\theta)\right],$$

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 ω_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

$$\widehat{N}(i) = \sum_{n \in \text{sample}} P(i|x_n;\theta) \sum_g \omega_g I_{ng} = \sum_n \omega_{g(n)} P(i|x_n;\theta)$$

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

Sample enumeration

Predicted shares

$$\widehat{W}(i) = \sum_{n \in \text{sample}} P(i|x_n;\theta) \sum_g \frac{N_g}{N_T} \frac{1}{S_g} I_{ng} = \frac{1}{N_T} \sum_n \omega_{g(n)} P(i|x_n;\theta)$$

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

$$\widehat{N}_h(i) = \sum_n \omega_{g(n)} P(i|x_n;\theta) I_{nh}$$

- Market share of alternative i in segment h

$$\widehat{W}_h(i) = \frac{\sum_n \omega_{g(n)} P(i|x_n;\theta) I_{nh}}{\sum_n \omega_{g(n)} I_{nh}}.$$

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

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	Cte. (PT)	0.977	0.605	1.61	0.11
2	Income 4-6 KCHF (PT)	-0.934	0.255	-3.67	0.00
3	Income 8-10 KCHF (PT)	-0.123	0.175	-0.70	0.48
4	Age 0-45 (PT)	-0.0218	0.00977	-2.23	0.03
5	Age 45-65 (PT)	0.0303	0.0124	2.44	0.01
6	Male dummy (PT)	-0.351	0.260	-1.35	0.18
7	Marginal cost [CHF] (PT)	-0.0105	0.0104	-1.01	0.31
8	Waiting time [min], if full time job (PT)	-0.0440	0.0117	-3.76	0.00
9	Waiting time [min], if part time job or other occupation (PT)	-0.0268	0.00742	-3.62	0.00
10	Travel time [min] $\times$ log(1+ distance[km]) / 1000, if full time job	-1.52	0.510	-2.98	0.00
11	Travel time [min] $\times$ log(1+ distance[km]) / 1000, if part time job	-1.14	0.671	-1.69	0.09
12	Season ticket dummy (PT)	2.89	0.346	8.33	0.00
13	Half fare travelcard dummy (PT)	0.360	0.177	2.04	0.04
14	Line related travelcard dummy (PT)	2.11	0.281	7.51	0.00
15	Area related travelcard (PT)	2.78	0.266	10.46	0.00
16	Other travel cards dummy (PT)	1.25	0.303	4.14	0.00

Example: interurban mode choice in Switzerland

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

Summary statistics

Number of observations = 1723

Number of estimated parameters = 27

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

Example: interurban mode choice in Switzerland

	Male	Female	Unknown gender	Population
Car	64.96%	60.51%	70.88%	62.8%
PT	30.20%	32.52%	25.59%	31.3%
Slow modes	4.83%	6.96%	3.53%	5.88%

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

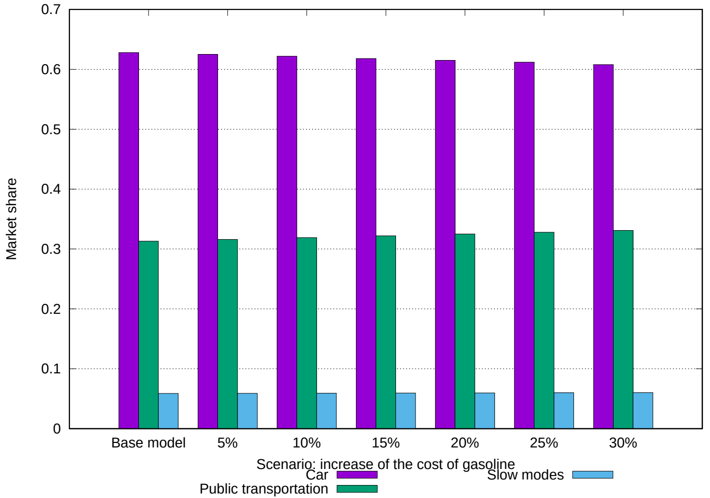
Procedure

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

Market shares

		Increase of the cost of gasoline					
	Now	5%	10%	15%	20%	25%	30%
Car	62.8%	62.5%	62.2%	61.8%	61.5%	61.2%	60.8%
PT	31.3%	31.6%	31.9%	32.2%	32.5%	32.8%	33.1%
Slow modes	5.88%	5.90%	5.92%	5.95%	5.97%	6.00%	6.02%

Forecasting



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

Price optimization

Optimizing the price of product i is solving the problem

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

Notes:

- It assumes that everything else is equal
- In practice, it is likely that the competition will also adjust the prices

Illustrative example

A binary logit model with

$$\begin{array}{rcl} V_1 & = & \beta_p p_1 - 0.5 \\ V_2 & = & \beta_p p_2 \end{array}$$

so that

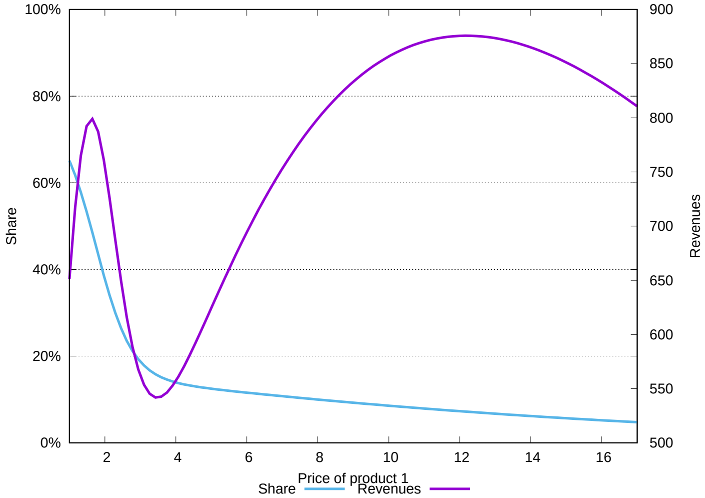
$$P(1|p) = \frac{e^{\beta_p p_1 - 0.5}}{e^{\beta_p p_1 - 0.5} + e^{\beta_p p_2}}$$

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

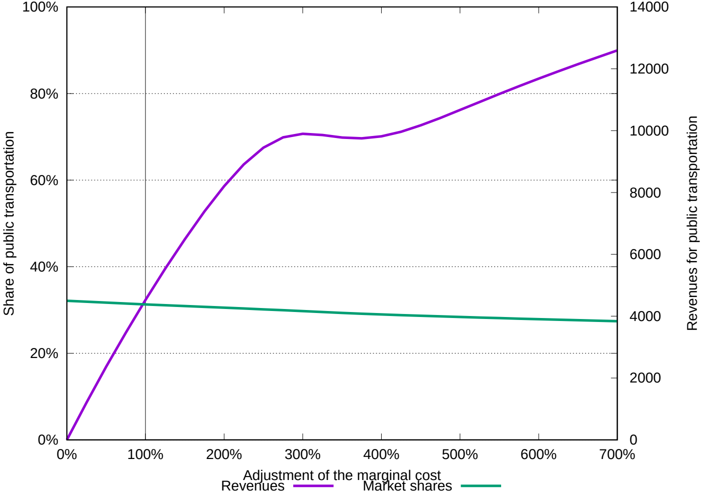


Case study: interurban mode choice in Switzerland

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



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

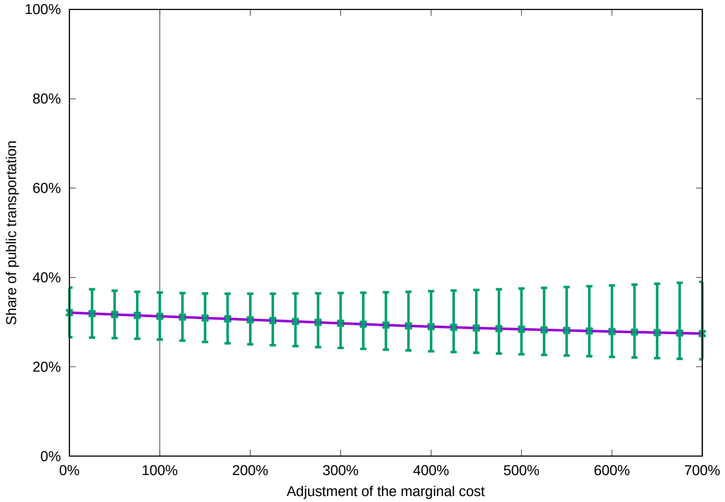
$$P(i|x_n, p_i; \theta)$$

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

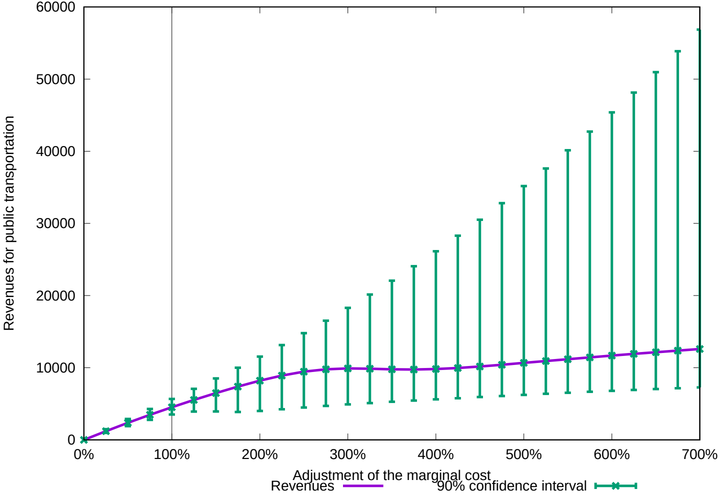
Calculating the confidence interval by simulation

- Draw R times $\tilde{\theta}$ from $N(\widehat{\theta}, \widehat{\Sigma})$.
- 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

$$P(i|x_n, p_i; \widehat{\theta})$$

- There are also errors in the x_n.
- If the distribution of x_n is known, draw from both x_n and θ.
- Apply the same procedure.

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

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 $x'_{in} = x_{in} + \delta^x_{in}$.
- We denote by δ^c_{in} the additional cost that would achieve the same utility, that is

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

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

$$\delta^c_{in}/\delta^x_{in}$$

Willingness to pay

Continuous variable

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

$$\begin{aligned} V_{in}(c_{in}, x_{in}) &= V_{in}(c_{in} + \delta^c_{in}, x_{in} + \delta^x_{in}) \\ &\approx V_{in}(c_{in}, x_{in}) + \delta^c_{in} \frac{\partial V_{in}}{\partial c_{in}}(c_{in}, x_{in}) + \delta^x_{in} \frac{\partial V_{in}}{\partial x_{in}}(c_{in}, x_{in}) \end{aligned}$$

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

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

$$\frac{\delta^c_{in}}{\delta^x_{in}} = -\frac{(\partial V_{in}/\partial x_{in})(c_{in}, x_{in})}{(\partial V_{in}/\partial c_{in})(c_{in}, x_{in})} = -\frac{\beta_x}{\beta_c}$$

Value of time

- The value of time is defined as the amount of money that an individual is willing to pay to save one unit of time

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

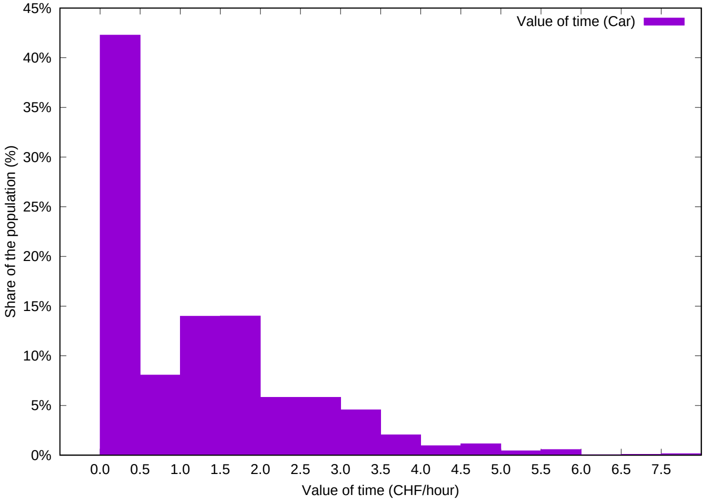
- Therefore

$$\mathrm{VOT}_{in} = \delta^c_{in}/(-\delta^t_{in}) = \frac{(\partial V_{in}/\partial t_{in})(c_{in}, t_{in})}{(\partial V_{in}/\partial c_{in})(c_{in}, t_{in})}$$

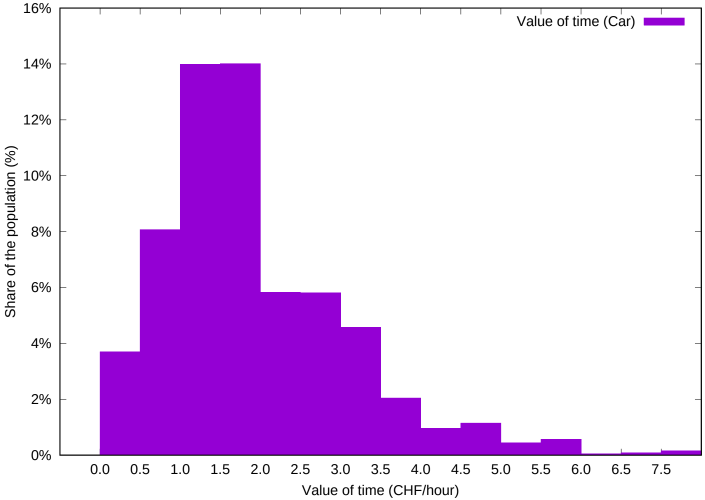
- If V is linear in these variables, we have

$$\mathrm{VOT}_{in} = \delta^c_{in}/(-\delta^t_{in}) = \frac{\beta_t}{\beta_c}.$$

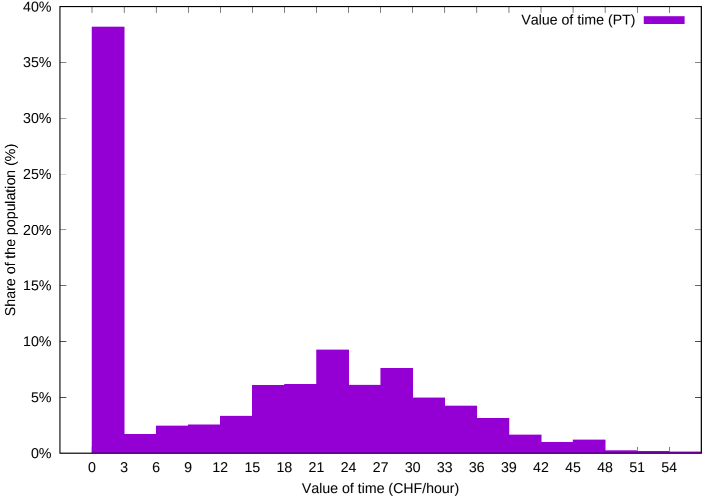
Case study: value of time for car drivers



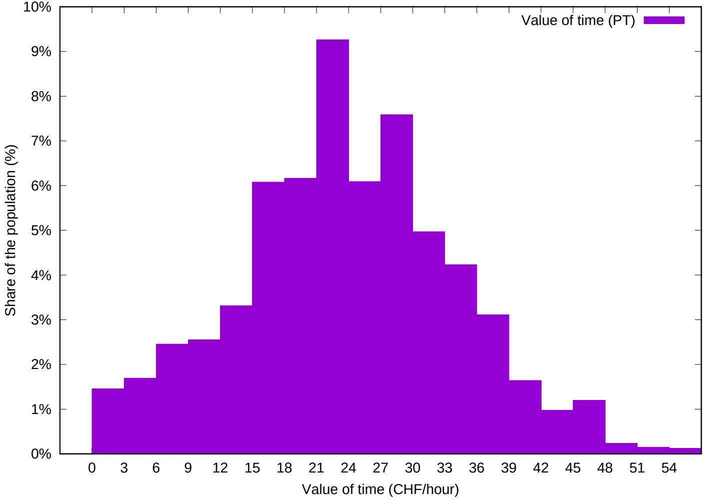
Case study: value of time for car drivers (nonzero)



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

	Point	Arc
Direct	$E^{P_n(i)}_{x_{ink}} = \frac{\partial P_n(i)}{\partial x_{ink}} \frac{x_{ink}}{P_n(i)}$	$\frac{\Delta P_n(i)}{\Delta x_{ink}} \frac{x_{ink}}{P_n(i)}$
Cross	$E^{P_n(i)}_{x_{jnk}} = \frac{\partial P_n(i)}{\partial x_{jnk}} \frac{x_{jnk}}{P_n(i)}$	$\frac{\Delta P_n(i)}{\Delta x_{jnk}} \frac{x_{jnk}}{P_n(i)}$

Aggregate elasticities

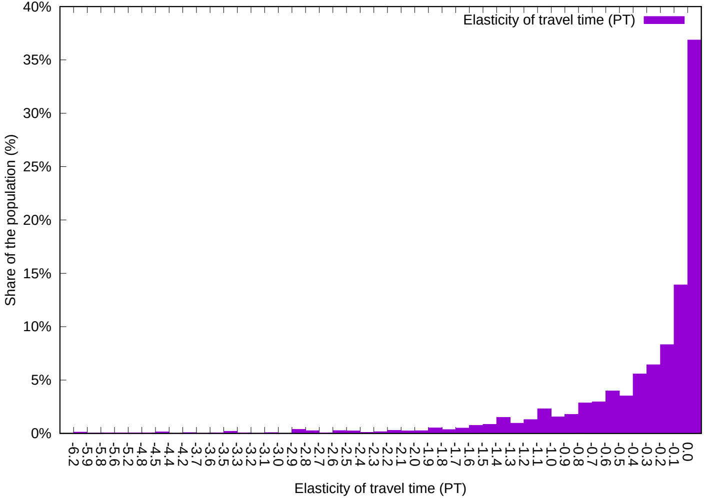
Population share

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

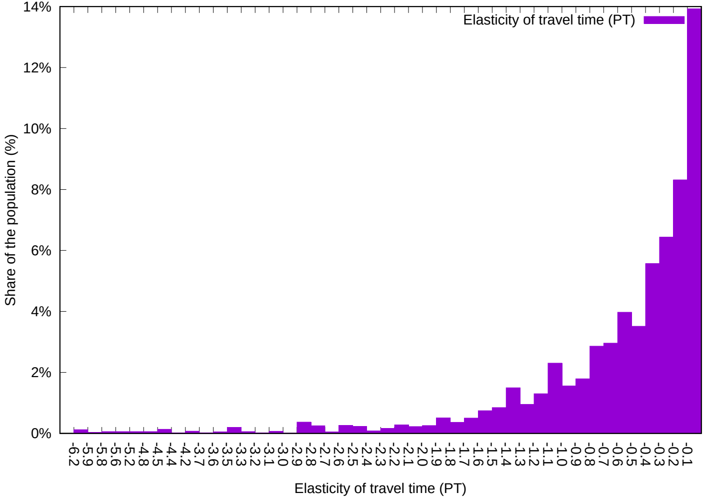
Aggregate elasticity

$$\begin{aligned} E_{x_{jk}}^{W(i)} &= \frac{\partial W(i)}{\partial x_{jk}} \frac{x_{jk}}{W(i)} \\ &= \sum_{n=1}^{N_T} \frac{P_n(i)}{P_n(i)} \frac{\partial P_n(i)}{\partial x_{jk}} \frac{x_{jk}}{\sum_{n=1}^{N_T} P_n(i)} \\ &= \sum_{n=1}^{N_T} \frac{P_n(i)}{\sum_{n=1}^{N_T} P_n(i)} E_{x_{jnk}}^{P_n(i)}. \end{aligned}$$

Case study: elasticity of travel time (PT)



Case study: elasticity of travel time (PT, non zero)



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

Consumer surplus

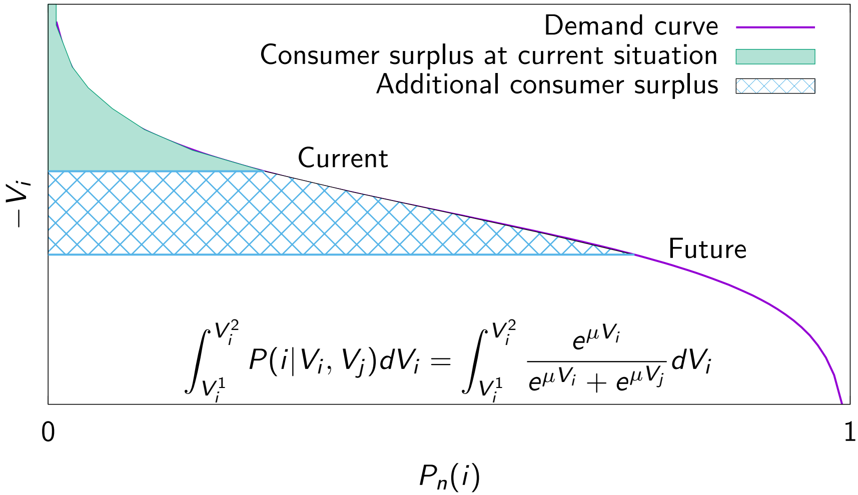
Concept

- 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



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

Consumer surplus

Binary logit

$$
\begin{aligned}
\int_{V_i^1}^{V_i^2} P(i|V_i, V_j) dV_i &= \int_{V_i^1}^{V_i^2} \frac{e^{\mu V_i}}{e^{\mu V_i} + e^{\mu V_j}} dV_i \\
&= \frac{1}{\mu} \ln(e^{\mu V_i^2} + e^{\mu V_j}) - \frac{1}{\mu} \ln(e^{\mu V_i^1} + e^{\mu V_j}).
\end{aligned}
$$

Consumer surplus

Generalization

$$\sum_{i \in \mathcal{C}} \int_{V^1}^{V^2} P(i|V) dV_i.$$

If the choice model has equal cross derivatives, that is

$$\frac{\partial P(i|V,\mathcal{C})}{\partial V_j} = \frac{\partial P(j|V,\mathcal{C})}{\partial V_i}, \ \forall i,j \in \mathcal{C},$$

the integral is path independent.

Logit

$$\sum_{i \in \mathcal{C}} \int_{V^1}^{V^2} P(i|V) dV_i = \frac{1}{\mu} \ln \sum_{j \in \mathcal{C}^2} e^{\mu V_j^2} - \frac{1}{\mu} \ln \sum_{j \in \mathcal{C}^1} e^{\mu V_j^1}.$$

Summary

Aggregation

- Sample enumeration

Precision

- Confidence intervals
- Calculated by simulation

Indicators

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