Forecasting

Aggregation and simulation

Michel Bierlaire

Mathematical Modeling of Behavior

Motivation

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

Outline

Aggregation

Forecasting

Price optimization

Confidence intervals

Aggregation

- Disaggregate model:

$$P_n(i|x_n;\theta)$$

- Obtain x_n for each individual n in the population.

Aggregate market shares

Number of individuals choosing alternative i

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

Share of the population choosing alternative i

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

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	$P(1\|x_N;\theta)$	$P(2\|x_N;\theta)$	$\cdots$	$P(J\|x_N;\theta)$	1
Total	$N(1)$	$N(2)$	$\cdots$	$N(J)$	N

Large table

When the table has too many rows...

apply sample enumeration.

When the table has too many columns...

apply micro simulation.

Too many rows: large population

Issues

- Complete enumeration cannot be applied in practice.
- No full access to each x_n, or to their distribution.
- Practical methods are needed.

Practical methods

- Use a sample of size N_S.
- It may be the same sample as for estimation.
- Warning: It cannot be stated preference data, with x_n generated by experimental design.

Sample enumeration

Stratified sample

- Population is partitioned into G homogeneous segments.
- S_g observations are sampled from each segment g, with $S = \sum_g S_g$.
- Let ω_g be the weight of segment g, that is

$$\omega_g = \frac{N_g}{N}\frac{S}{S_g} = \frac{\text{share of persons in segment } g \text{ in population}}{\text{share of persons in segment } g \text{ in sample}}$$

- Weight for individual n:

$$\omega_n = \sum_{g=1}^{G} \delta_{ng}\omega_g,$$

where $\delta_{ng} = 1$ if n belongs to g, and 0 otherwise.

Sample enumeration

Weights

$$\omega_n = \sum_{g=1}^{G} \delta_{ng} \omega_g ,$$

where

$$\omega_g = \frac{N_g}{N} \frac{S}{S_g} .$$

Sum

$$\begin{aligned} \sum_{n=1}^{S} \omega_n &= \sum_g S_g \omega_g \\ &= \sum_g S_g \frac{N_g}{N} \frac{S}{S_g} \\ &= S \frac{\sum_g N_g}{N} = S . \end{aligned}$$

Sample enumeration

Predicted shares

$$\widehat{W}(i) = \frac{1}{S}\sum_{n=1}^{S} \omega_n P(i|x_n;\theta).$$

Comments

- $\sum_i \widehat{W}(i) = 1$.
- Consistent estimate.
- Estimate subject to sampling errors.
- Policy analysis: change the values of the explanatory variables, and apply the same procedure.

Example

Population $N = 2110K$

	Male	Female
Age $\leq$ 45	600K	590K
Age $>$ 45	450K	470K

Sample: $S = 800$

	Male	Female
Age $\leq$ 45	300	250
Age $>$ 45	150	100

N_g / N

	Male	Female
Age $\leq$ 45	0.284	0.280
Age $>$ 45	0.213	0.223

S_g / S

	Male	Female
Age $\leq$ 45	0.375	0.313
Age $>$ 45	0.188	0.125

Example

N_g / N

	Male	Female
Age ≤ 45	0.284	0.280
Age > 45	0.213	0.223

S_g / S

	Male	Female
Age ≤ 45	0.375	0.313
Age > 45	0.188	0.125

w_g

	Male	Female
Age ≤ 45	0.758	0.895
Age > 45	1.14	1.78

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: 1785.

Model: 3 alternatives

- Car,
- public transportation (PT),
- slow modes.

Example: interurban mode choice in Switzerland

Car: utility specification

- Constant: language (F [ref], G), subscription (no GA [ref], GA). [3]
- Travel time (min.): age, trip purpose (work [ref], other). [2+1]

$$T_{\text{car}}(\text{age}/100)^{\lambda}.$$

- Travel cost (CHF): no interaction. Coef. normalized to -1. [0]

Example: interurban mode choice in Switzerland

Public transport

- Constant: language (F [ref], G), subscription (no GA [ref], GA). [3]
- Travel time (min.): age, trip purpose (work [ref], other). [2+0]

$$T_{\mathrm{PT}}(\mathrm{age}/100)^{\lambda}.$$

- Travel cost (CHF): 0 if GA, no interaction. Coef. normalized to -1. [0]
- Waiting time (min.): profession (intellectual, manager, craftman, other [ref]). [4]

Example: interurban mode choice in Switzerland

Slow modes

Distance (km): education (university, other [ref]). [2]

Heteroscedasticity

With vs. without GA. [2]

Summary

19 parameters to estimate

Example: interurban mode choice in Switzerland

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	Cte. (Car)	34.3	12.9	2.66	0.00792
2	Cte., GA (Car)	-25.8	11.1	-2.32	0.0201
3	Cte., German (Car)	-22.9	8.2	-2.8	0.00518
4	Time (min.) $(\text{age}/100)^{\lambda}$, (Car)	-0.725	0.349	-2.08	0.0375
5	Time (min.) $(\text{age}/100)^{\lambda}$, other trip purp. (Car)	0.378	0.222	1.7	0.0891
6	Cte., (PT)	-17.3	11.9	-1.46	0.145
7	Cte., GA (PT)	3.62	11.0	0.33	0.741
8	Cte., German (PT)	5.05	7.92	0.638	0.524
9	Time (min.) $(\text{age}/100)^{\lambda}$, (PT)	-0.169	0.112	-1.51	0.13
10	Time (min.) $(\text{age}/100)^{\lambda}$, other trip purp. (PT)	0.0749	0.0742	1.01	0.312
11	Waiting time (min.), (PT)	-0.641	0.238	-2.7	0.00696
12	Waiting time (min.), craftsman (PT)	0.593	0.31	1.91	0.0562
13	Waiting time (min.), intellectual (PT)	0.551	0.268	2.06	0.0396
14	Waiting time (min.), manager (PT)	-0.881	0.405	-2.18	0.0295
15	Distance (km) (Slow modes)	-3.23	1.05	-3.08	0.0021
16	Distance (km), university (Slow modes)	-2.79	1.39	-2.01	0.0448
17	λ	-0.339	0.341	-0.994	0.32
18	μ (GA)	0.0796	0.0316	2.52	0.0117
19	μ (no GA)	0.0392	0.00962	4.08	0.0

Example: interurban mode choice in Switzerland

Estimation results

Number of estimated parameters	19
Sample size	1785
Excluded observations	121
Null log likelihood	-1922.909
Final log likelihood	-990.9459
Likelihood ratio test for the null model	1863.927
Rho-square for the null model	0.485
Rho-square-bar for the null model	0.475
Akaike Information Criterion	2019.892
Bayesian Information Criterion	2124.148

Example: interurban mode choice in Switzerland

	French speaking (25.6%)	German speaking (74.4%)	Population
Car	81.8%	55.3%	62.1%
PT	15.4%	37.1%	31.6%
Slow modes	4.26%	7.11%	6.38%

Too many columns: large choice sets

There are 80,000 ways to drink a Starbucks beverage, with fancy combinations such as a "tall, non-fat latte with caramel drizzle", a "grande, iced, sugar-free, vanilla latte with soy milk" or a "tall, half-caff, soy latte at 120 degrees". [HuffPost 2017]

Large choice sets

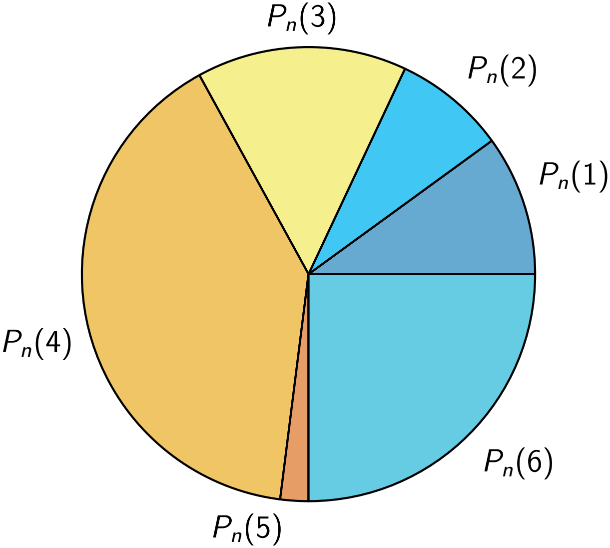
Combinatorial choice sets

- A choice is a combination of choices.
- Impossible to enumerate all possibilities.

Process sequentially

- First the size.
- Second the type of coffee.
- Third, additional options, if any.

Microsimulation



Microsimulation

Simulated choice

- For each observation, draw R times from the choice model.
- Define $\widehat{y}_{inr} = 1$ if alternative i has been generated by draw r, 0 otherwise.
- Approximation:

$$P_n(i|x_n;\theta) \approx \frac{1}{R}\sum_{r=1}^{R}\widehat{y}_{inr}.$$

Warning

It is invalid to select the alternative with the highest probability.

Aggregate market shares

Number of individuals choosing alternative i

$$\widehat{N}(i) = \frac{1}{R}\sum_{n=1}^{N}\sum_{i=1}^{R}\widehat{y}_{inr}.$$

Share of the population choosing alternative i

$$\widehat{W}(i) = \frac{1}{N}\frac{1}{R}\sum_{n=1}^{N}\sum_{i=1}^{R}\widehat{y}_{inr}.$$

Microsimulation

For each r

Population	Alternatives				Total
	1	2	$\cdots$	J	
1	$\widehat{y}_{11r}$	$\widehat{y}_{21r}$	$\cdots$	$\widehat{y}_{J1r}$	1
2	$\widehat{y}_{12r}$	$\widehat{y}_{22r}$	$\cdots$	$\widehat{y}_{J2r}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$\widehat{y}_{1Nr}$	$\widehat{y}_{2Nr}$	$\cdots$	$\widehat{y}_{JNr}$	1
Total	$\widehat{N}(1)$	$\widehat{N}(2)$	$\cdots$	$\widehat{N}(J)$	N

Microsimulation

In practice

Population	Draw			
	1	2	$\cdots$	R
1	i_{11}	i_{12}	$\cdots$	i_{1R}
2	i_{21}	i_{22}	$\cdots$	i_{2R}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	i_{N1}	i_{N2}	$\cdots$	i_{NR}

Outline

Scenarios

Definition

- Description of existing or foreseeable market conditions.
- Need for the values of each explanatory variable for each member of the population.
- If sample enumeration is used, values are needed only for the sample.

Typical process

- Collect revealed preference data from the population.
- Possibly the same data as for estimation.
- Consider it as the “base scenario”.
- Validate and calibrate the model on the base scenario.
- Define new scenarios by modifying the base scenario.

Scenarios

New alternatives

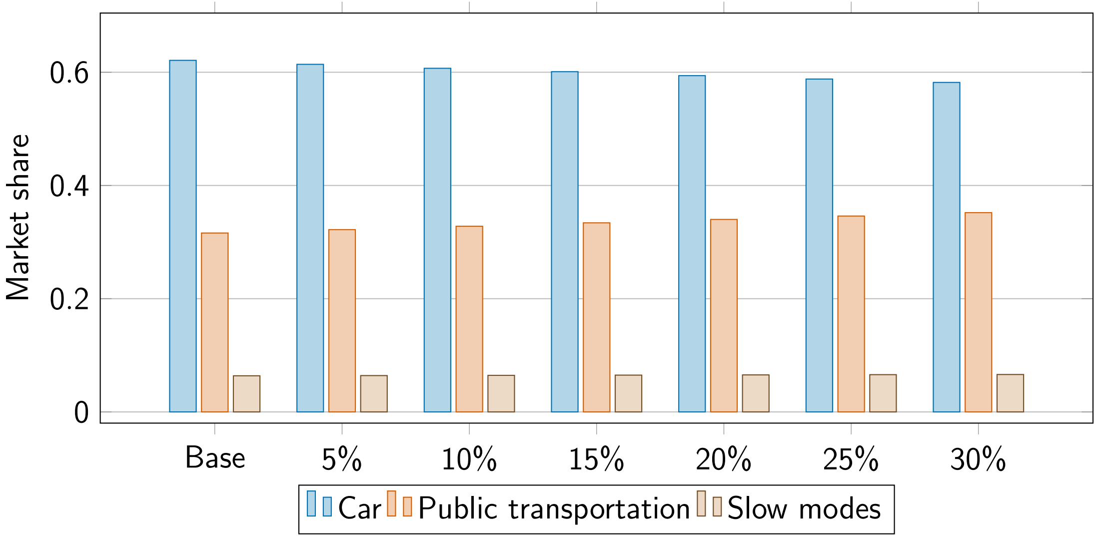
- Require SP data at the estimation stage.
- SP data cannot be used for the definition of scenarios.
- Values of the variables must correspond to a foreseeable situation.
- Good practice: during survey preparation, add one question corresponding to that foreseeable situation to the SP design. It is the only way to calibrate the alternative specific constant of the new alternative.

Forecasting

Market shares

		Increase of the car travel time					
	Now	5%	10%	15%	20%	25%	30%
Car	62.1%	61.4%	60.7%	60.1%	59.4%	58.8%	58.2%
PT	31.6%	32.2%	32.8%	33.4%	34.0%	34.6%	35.2%
Slow modes	6.38%	6.42%	6.46%	6.50%	6.54%	6.58%	6.61%

Forecasting



Outline

Price optimization

Optimizing the price of product i is solving the problem

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

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

$$\begin{aligned} V_1 &= \beta_p p_1 - 0.5, \\ V_2 &= \beta_p p_2, \end{aligned}$$

so that

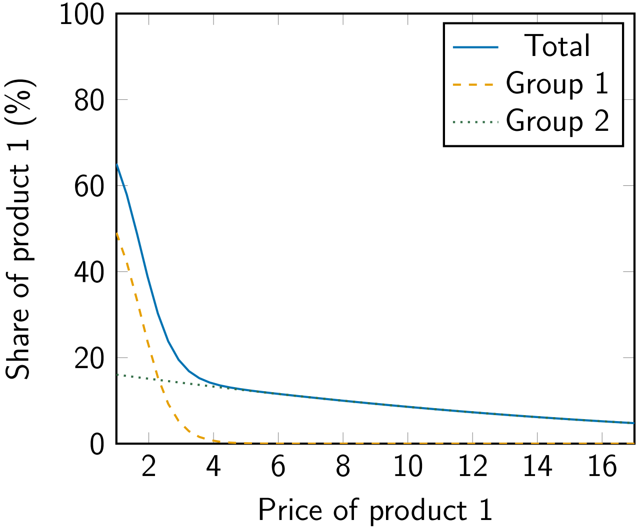
$$P(1|p) = \frac{e^{\beta_p p_1 - 0.5}}{e^{\beta_p p_1 - 0.5} + e^{\beta_p p_2}}.$$

Two groups in the population:

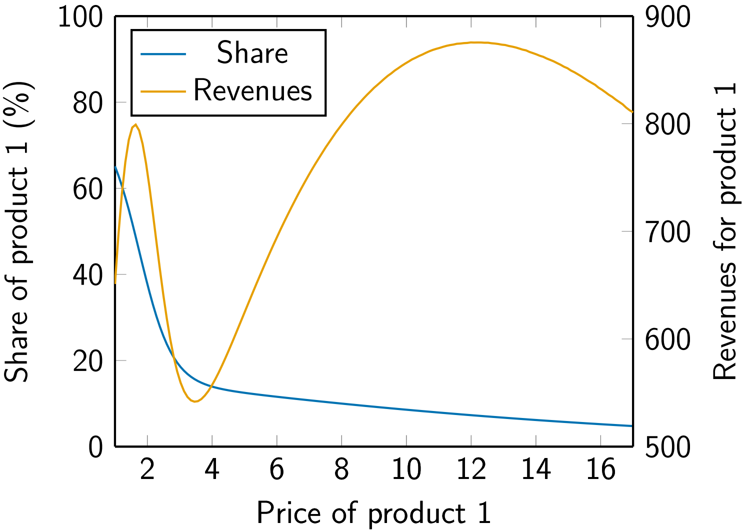
- Group 1: $\beta_p = -2$, $N_s = 600$.
- Group 2: $\beta_p = -0.1$, $N_s = 400$.

Assume that $p_2 = 2$.

Illustrative example



Illustrative example

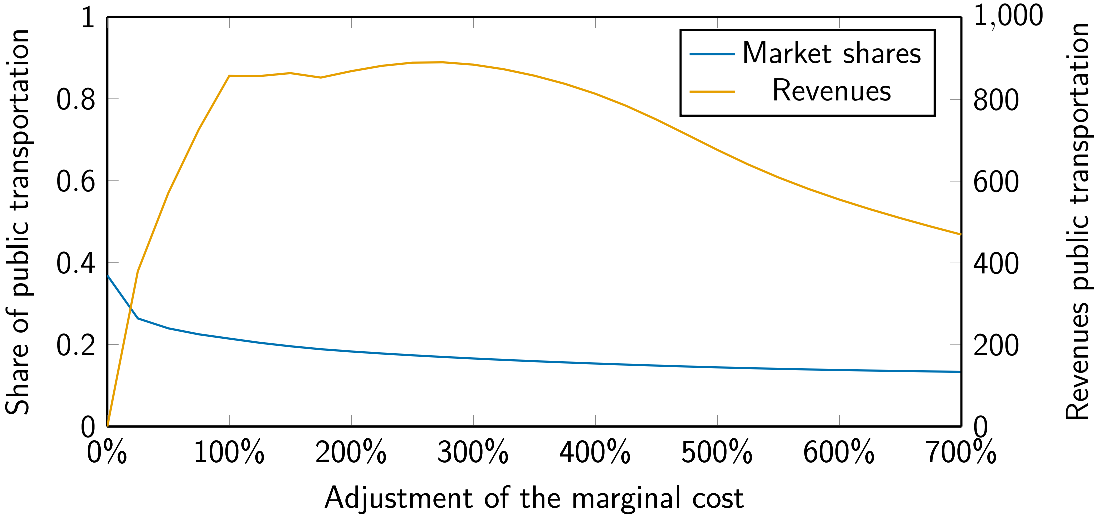


Case study: interurban mode choice in Switzerland

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



Price optimization

Comments

- In a competitive environment, competitors adjust their prices as well.
- 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

Confidence intervals

Model

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

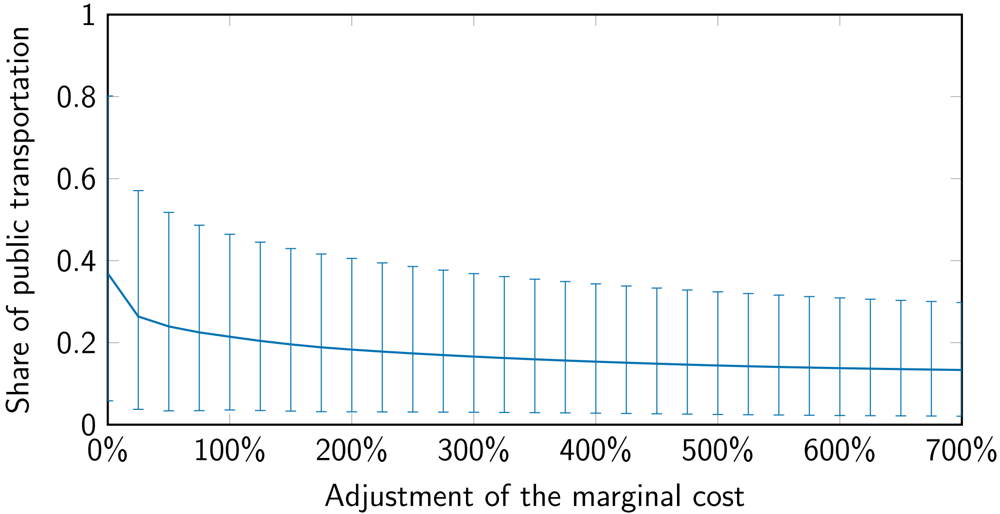
- In reality, we use $\widehat{\theta}$, the maximum likelihood estimate of θ.
- It is different from the true value θ due to sampling errors.
- Confidence intervals can be obtained using bootstrapping and simulation.

Confidence intervals

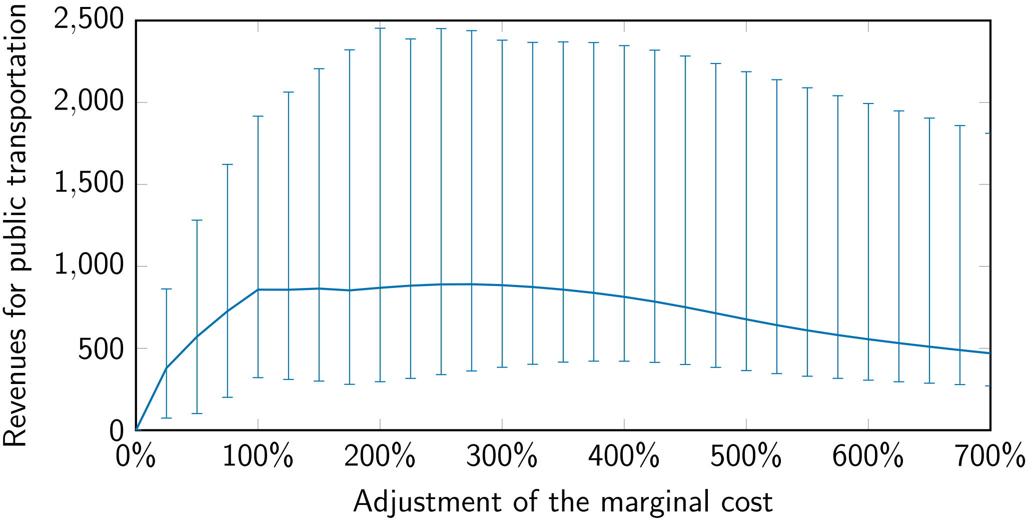
Calculating the confidence interval by bootstrapping

- Draw R bootstrap samples (draw from the data, with replacement).
- For each of them, re-estimate the parameters $\tilde{\theta}_r$.
- For each $\tilde{\theta}_r$, calculate the requested quantity (e.g. market share, revenue, etc.) using $P(i|x_n, p_i; \tilde{\theta}_r)$.
- Calculate the 5% and the 95% quantiles of the generated quantities.
- They define the 90% confidence interval.

Case study: confidence intervals (100 draws)



Case study: confidence intervals (100 draws)



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

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

Note: moneymetric utility function: $\beta_c = -1$, so that WTP$=\beta_x$.

Value of time

- Amount of money that an individual is willing to pay to **increase** travel time by one unit of time:

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

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

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

Value of time

Linear case

- If V is linear in the variables, we have

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

- Moneymetric utility function: $\beta_c = -1$.

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

Value of time

Average in the population

$$\text{VOT}_i = \sum_n \omega_n \text{VOT}_{in}.$$

Case study: average value of time

- $\text{VOT}_{\text{car}} = 40.2$ CHF/h.
- $\text{VOT}_{\text{PT}} = 9.22$ CHF/h (non zeros).

Case study: car drivers

Utility function

$$V_{\text{car},n} = -c_{\text{car},n} + \beta_{tn} \left(\frac{\text{age}}{100}\right)^{\lambda} t_{\text{car},n} + \cdots$$

Value of time for car drivers (CHF/h)

$$60 \frac{\beta_{tn} \left(\frac{\text{age}}{100}\right)^{\lambda}}{-1} = -60 \beta_{tn} \left(\frac{\text{age}}{100}\right)^{\lambda},$$

where $\beta_{tn} = -0.725$ if trip purpose for n is work, $\beta_{tn} = -0.725 + 0.378 = -0.347$ if trip purpose for n is not work, and $\lambda = -0.339$.

Case study: public transportation

Utility function

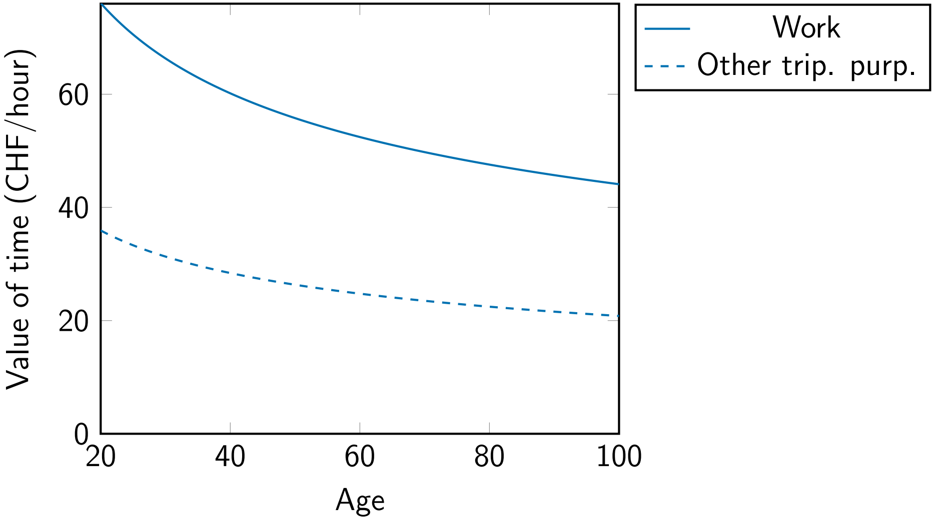
$$V_{\mathrm{PT},n} = -c_{\mathrm{PT},n} + \beta_{tn} \left(\frac{\mathrm{age}}{100}\right)^{\lambda} t_{\mathrm{PT},n} + \cdots$$

Value of time for public transportation (CHF/h)

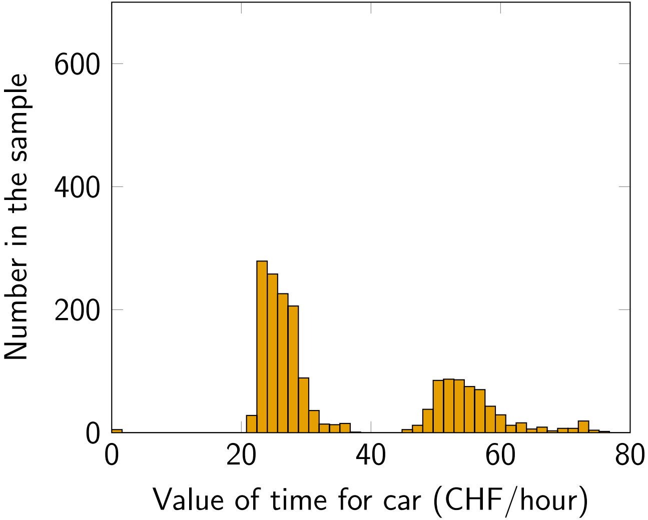
$$60 \frac{\beta_{tn} \left(\frac{\mathrm{age}}{100}\right)^{\lambda}}{-1} = -60\beta_{tn} \left(\frac{\mathrm{age}}{100}\right)^{\lambda},$$

where $\beta_{tn} = -0.169$ if trip purpose for n is work,
$\beta_{tn} = -0.169 + 0.0749 = -0.0941$ if trip purpose for n is not work, and
$\lambda = -0.339$.

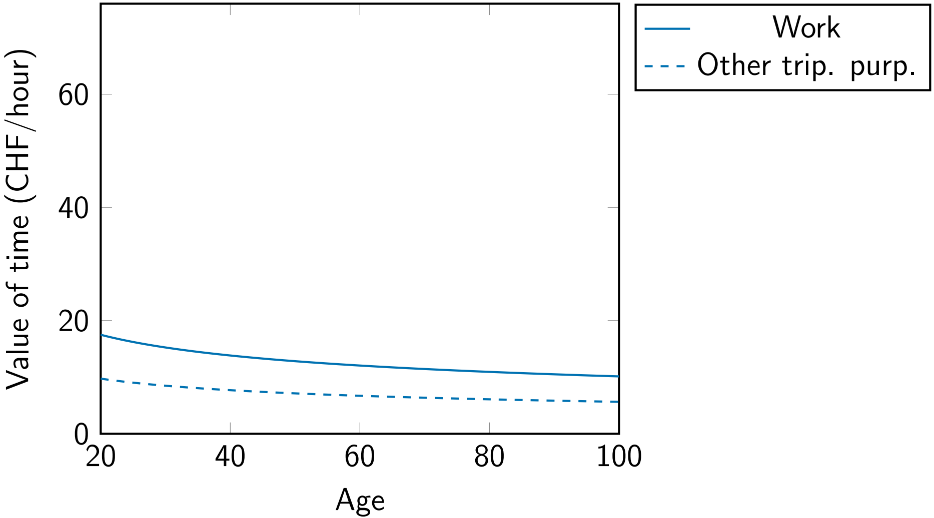
Case study: value of time for car drivers



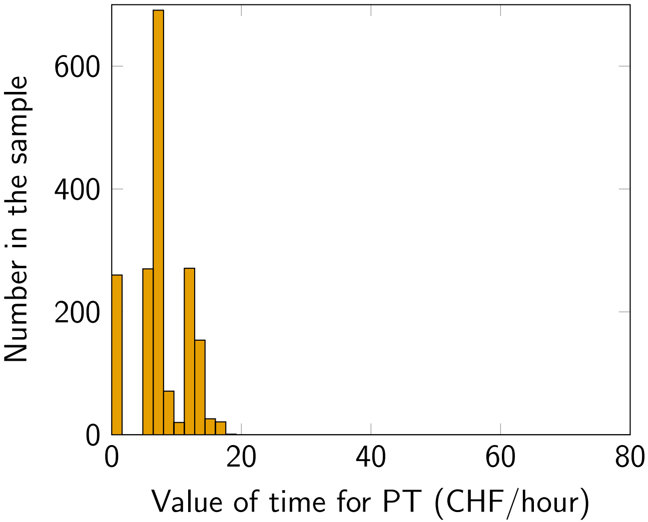
Case study: value of time for car drivers



Case study: value of time for public transportation



Case study: value of time for public transportation



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	$E_{x_{ink}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} \frac{x_{ink}}{P_n(i)}.$	$\frac{\Delta P_n(i)}{\Delta x_{ink}} \frac{x_{ink}}{P_n(i)}.$
Cross	$E_{x_{jnk}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_{jnk}} \frac{x_{jnk}}{P_n(i)}.$	$\frac{\Delta P_n(i)}{\Delta x_{jnk}} \frac{x_{jnk}}{P_n(i)}.$

Aggregate elasticities

Population share

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

Aggregate elasticity

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

Aggregate elasticities

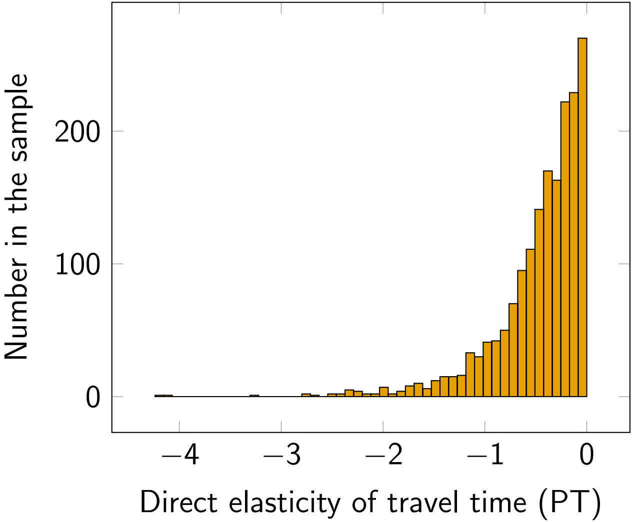
Sample enumeration

$$\widehat{W}(i) = \frac{1}{S}\sum_{n=1}^{S} \omega_n P(i|x_n).$$

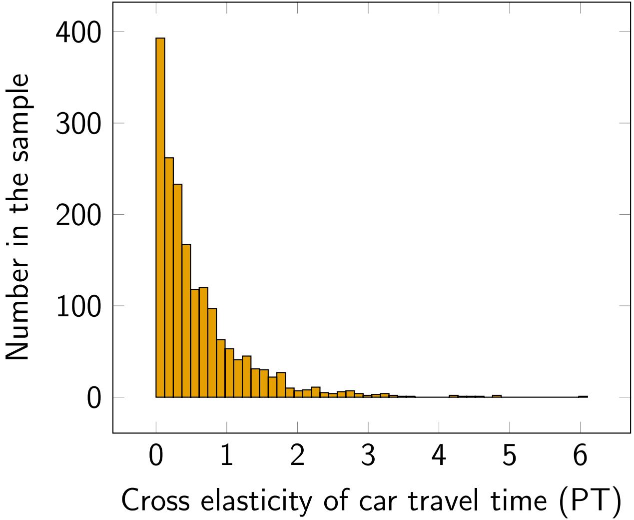
Aggregate elasticity

$$\begin{aligned}
E^{\widehat{W}(i)}_{x_{jk}} &= \frac{\partial \widehat{W}(i)}{\partial x_{jk}} \frac{x_{jk}}{\widehat{W}(i)} \\
&= \frac{1}{S}\sum_{n=1}^{S} \omega_n \frac{P_n(i)}{P_n(i)} \frac{\partial P_n(i)}{\partial x_{jk}} \frac{S x_{jk}}{\sum_{\ell=1}^{S} \omega_\ell P_\ell(i)} \\
&= \frac{1}{\sum_{\ell=1}^{S} \omega_\ell P_\ell(i)} \sum_{n=1}^{S} \omega_n P_n(i) E^{P_n(i)}_{x_{jnk}}.
\end{aligned}$$

Case study: direct elasticity of travel time (PT)



Case study: cross elasticity of car travel time (PT)



Case study: cross elasticity of car travel time (PT)

Comments

- 120 individuals with zero cross elasticity of car travel time (PT).
- 94 because car is not available.
- 29 because car cost $<=$ 0.32 CHF.

Consumer surplus

Definition

Difference between what a consumer is willing to pay for a good and what she actually pays for that good.

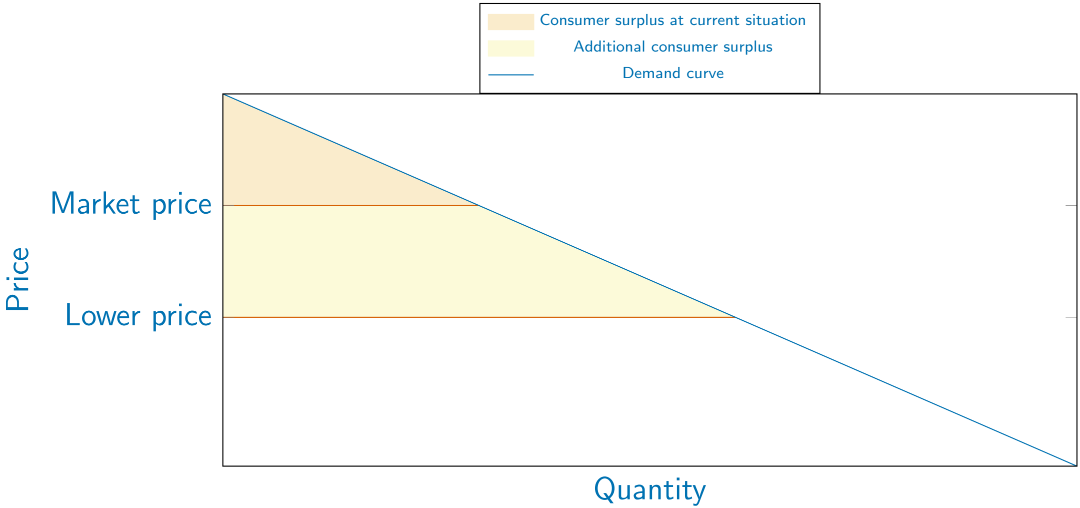
Calculation

Area under the demand curve and above the market price

Demand curve

- Plot of the inverse demand function
- Price as a function of quantity

Consumer surplus

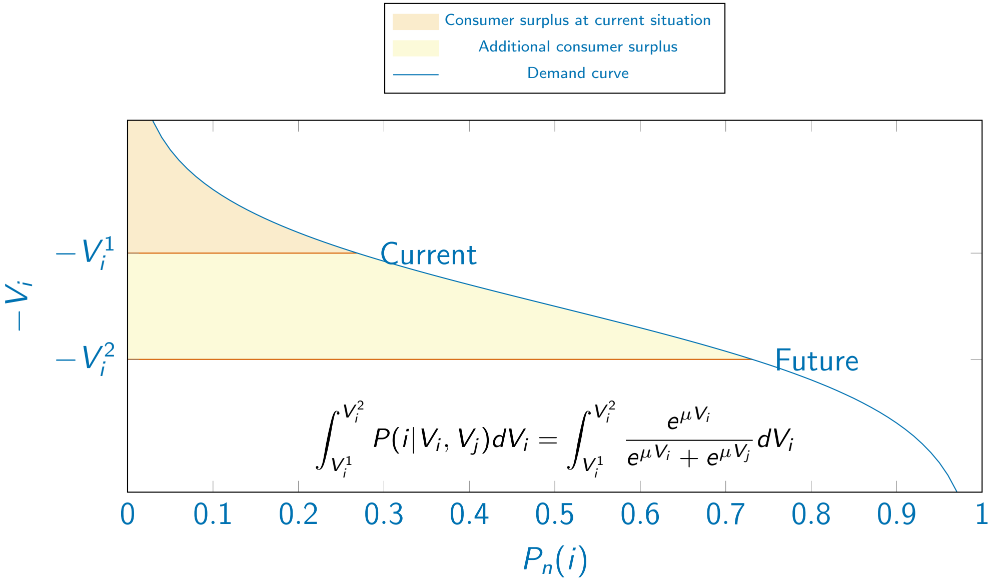


Consumer surplus

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

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

$$= \frac{1}{\mu} \ln\left(e^{\mu V_i^2} + e^{\mu V_j}\right) - \frac{1}{\mu} \ln\left(e^{\mu V_i^1} + e^{\mu V_j}\right).$$

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.
  - Too many individuals: sample enumeration.
  - Too many alternatives: micro simulation.
- Market shares.
- Price optimization.
- Confidence intervals.
- Willingness to pay.
- Elasticities.
- Consumer surplus.