Introduction to Disaggregate Demand Models

Michel Bierlaire Virginie Lurkin

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

October 23, 2017

Outline

Demand

Demand = behavior = sequence of choices

Aggregate demand

Aggregate demand

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

Disaggregate demand

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.

Discrete choice models

Daniel L. McFadden

- UC Berkeley 1963, MIT 1977, UC Berkeley 1991
- Laureate of *The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2000*
- Owns a farm and vineyard in Napa Valley
- "Farm work clears the mind, and the vineyard is a great place to prove theorems"

Outline

Microeconomic consumer theory

Continuous choice set

- Consumption bundle

$$Q = \begin{pmatrix} q_1 \\ \vdots \\ q_L \end{pmatrix} ; p = \begin{pmatrix} p_1 \\ \vdots \\ p_L \end{pmatrix}$$

- Budget constraint

$$p^T Q = \sum_{\ell=1}^{L} p_\ell q_\ell \leq I.$$

- No attributes, just quantities

Preferences

Operators $\succ$, $\sim$, and $\succsim$

- $Q_a \succ Q_b$: Q_a is preferred to Q_b,
- $Q_a \sim Q_b$: indifference between Q_a and Q_b,
- $Q_a \succsim Q_b$: Q_a is at least as preferred as Q_b.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Preferences

Rationality

- Completeness: for all bundles a and b,

$$Q_a \succ Q_b \text{ or } Q_a \prec Q_b \text{ or } Q_a \sim Q_b.$$

- Transitivity: for all bundles a, b and c,

$$\text{if } Q_a \succsim Q_b \text{ and } Q_b \succsim Q_c \text{ then } Q_a \succsim Q_c.$$

- "Continuity": if Q_a is preferred to Q_b and Q_c is arbitrarily "close" to Q_a, then Q_c is preferred to Q_b.

Utility

Utility function

- Parametrized function:

$$\widetilde{U} = \widetilde{U}(q_1, \ldots, q_L; \theta) = \widetilde{U}(Q; \theta)$$

- Consistent with the preference indicator:

$$\widetilde{U}(Q_a; \theta) \geq \widetilde{U}(Q_b; \theta)$$

  is equivalent to

$$Q_a \succsim Q_b.$$

- Unique up to an order-preserving transformation

Optimization

Optimization problem

$$\max_{Q} \widetilde{U}(Q;\theta)$$

subject to

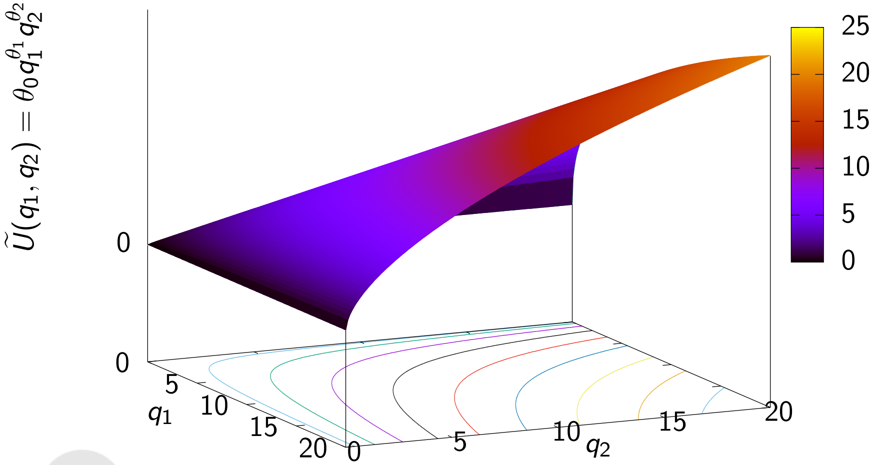
$$p^T Q \leq I, \;\; Q \geq 0$$

Demand function

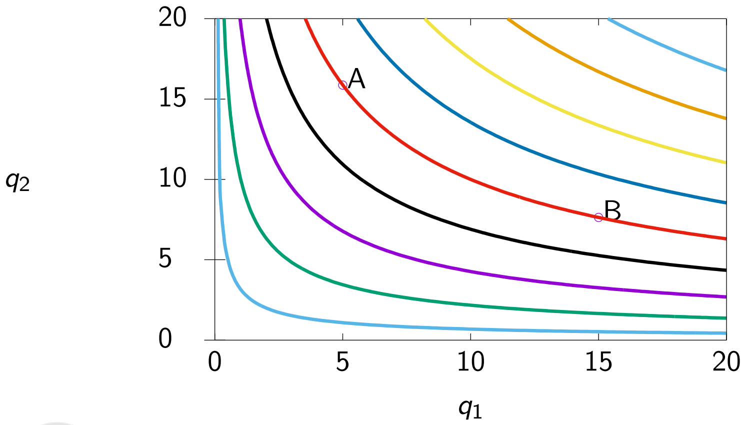
- Solution of the optimization problem
- Quantity as a function of prices p and budget I

$$Q^* = f(I, p; \theta)$$

Example: Cobb-Douglas



Example



Example

Optimization problem

$$\max_{q_1,q_2} \widetilde{U}(q_1, q_2; \theta_0, \theta_1, \theta_2) = \theta_0 q_1^{\theta_1} q_2^{\theta_2}$$

subject to

$$p_1 q_1 + p_2 q_2 = I$$

Lagrangian of the problem:

$$L(q_1, q_2, \lambda) = \theta_0 q_1^{\theta_1} q_2^{\theta_2} + \lambda(I - p_1 q_1 - p_2 q_2)$$

Necessary optimality condition

$$\nabla L(q_1, q_2, \lambda) = 0$$

Example

Necessary optimality conditions

$$
\begin{array}{rcrcll}
\theta_0\theta_1 q_1^{\theta_1-1} q_2^{\theta_2} & - & \lambda p_1 & = & 0 & (\times q_1) \\
\theta_0\theta_2 q_1^{\theta_1} q_2^{\theta_2-1} & - & \lambda p_2 & = & 0 & (\times q_2) \\
p_1 q_1 + p_2 q_2 & - & I & = & 0. &
\end{array}
$$

We have

$$
\begin{array}{rcrcl}
\theta_0\theta_1 q_1^{\theta_1} q_2^{\theta_2} & - & \lambda p_1 q_1 & = & 0 \\
\theta_0\theta_2 q_1^{\theta_1} q_2^{\theta_2} & - & \lambda p_2 q_2 & = & 0.
\end{array}
$$

Adding the two and using the third condition, we obtain

$$\lambda I = \theta_0 q_1^{\theta_1} q_2^{\theta_2}(\theta_1 + \theta_2)$$

or, equivalently,

$$\theta_0 q_1^{\theta_1} q_2^{\theta_2} = \frac{\lambda I}{(\theta_1 + \theta_2)}$$

Solution

From the previous derivation

$$\theta_0 q_1^{\theta_1} q_2^{\theta_2} = \frac{\lambda I}{(\theta_1 + \theta_2)}$$

First condition

$$\theta_0 \theta_1 q_1^{\theta_1} q_2^{\theta_2} = \lambda p_1 q_1.$$

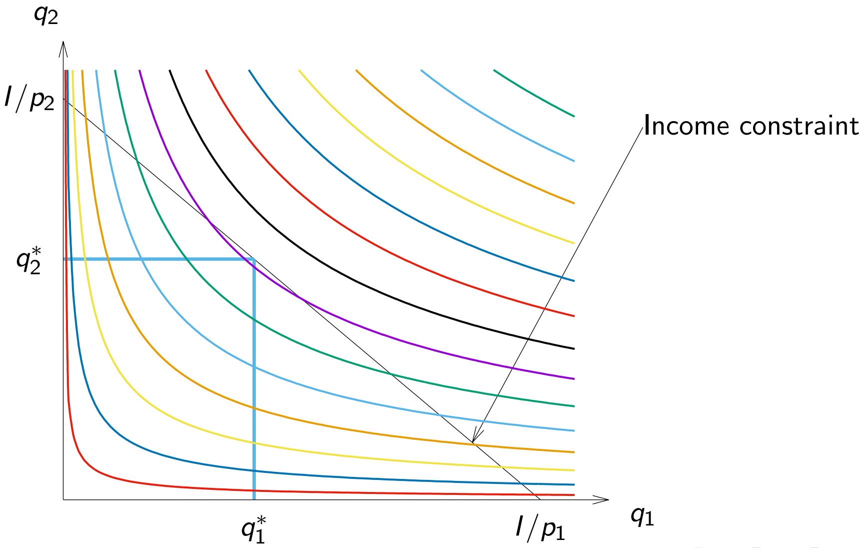
Solve for q_1

$$q_1^* = \frac{I\theta_1}{p_1(\theta_1 + \theta_2)}$$

Similarly, we obtain

$$q_2^* = \frac{I\theta_2}{p_2(\theta_1 + \theta_2)}$$

Optimization problem



Demand functions

Product 1

$$q_1^* = \frac{I}{p_1} \frac{\theta_1}{\theta_1 + \theta_2}$$

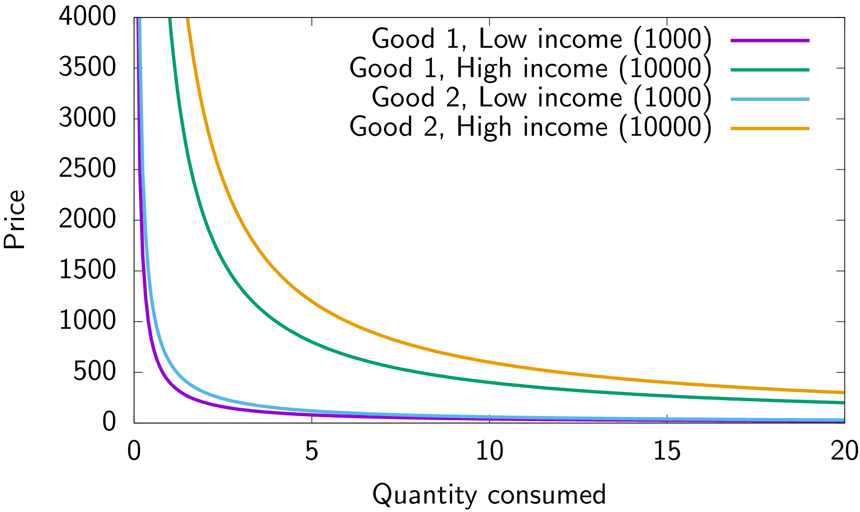
Product 2

$$q_2^* = \frac{I}{p_2} \frac{\theta_2}{\theta_1 + \theta_2}$$

Comments

- Demand decreases with price
- Demand increases with budget
- Demand independent of θ_0, which does not affect the ranking
- Property of Cobb Douglas: the demand for a good is only dependent on its own price and independent of the price of any other good.

Demand curve (inverse of demand function)



Indirect utility

Substitute the demand function into the utility

$$U(I, p; \theta) = \theta_0 \left(\frac{I}{p_1} \frac{\theta_1}{\theta_1 + \theta_2} \right)^{\theta_1} \left(\frac{I}{p_2} \frac{\theta_2}{\theta_1 + \theta_2} \right)^{\theta_2}$$

Indirect utility

Maximum utility that is achievable for a given set of prices and income

In discrete choice...

- only the indirect utility is used
- therefore, it is simply referred to as “utility”

Microeconomic theory of discrete goods

Car choice

- Discrete: what type of car?
- Continuous: how many kilometers per year?

Energy choice

- Discrete: electricity or gas for house heating?
- Continuous: what temperature for the house?

Holidays

- Discrete: what destination?
- Continuous: how long to stay?

Expanding the microeconomic framework

The consumer

- chooses the quantities of continuous goods: $Q = (q_1, \ldots, q_L)$
- chooses alternatives in a discrete choice set $i = 1, \ldots, j, \ldots, J$
- discrete decision vector: $(y_1, \ldots, y_J)$, $y_j \in \{0, 1\}$.

Utility maximization

Utility

$$\widetilde{U}(Q, y, \tilde{z}^T y; \theta)$$

- Q: quantities of the continuous good
- y: discrete choice
- $\tilde{z}^T = (\tilde{z}_1, \ldots, \tilde{z}_i, \ldots, \tilde{z}_J) \in \mathbb{R}^{K \times J}$: K attributes of the J alternatives
- θ: vector of parameters

Utility maximization

Optimization problem

$$\max_{Q,y} \widetilde{U}(Q, y, \tilde{z}^T y; \theta)$$

subject to

$$p^T Q + c^T y \leq I$$
$$y_j \in \{0, 1\}, \forall j$$

where $c^T = (c_1, \ldots, c_i, \ldots, c_J)$ contains the cost of each alternative.

Solving the problem

- Mixed integer optimization problem
- No optimality condition
- Impossible to directly derive demand functions

Solving the problem

Step 1: condition on the choice of the discrete goods

- Fix the discrete goods, that is select a feasible y.
- The problem becomes a continuous problem in Q.
- Conditional demand functions can be derived:

$$q_{\ell|y} = f(I - c^T y, p, \tilde{z}^T y; \theta),$$

- $I - c^T y$ is the income left for the continuous goods.
- If $I - c^T y < 0$, y is declared unfeasible.

Solving the problem

Conditional indirect utility functions

Substitute the demand functions into the utility:

$$U = U(I - c^T y, p, \tilde{z}; \theta).$$

Step 2: Choice of the discrete good

$$\max_y U(I - c^T y, p, \tilde{z}^T y; \theta)$$

subject to

$$c^T y \leq I$$

- Knapsack problem.
- In many practical case, it can be solved by enumeration.

Model for individual n

Choice set

Each feasible y is an alternative i

(Indirect) utility function

$$\max_y U(I_n - c_n^T y, p_n, \tilde{z}_n^T y; \theta_n)$$

simplifies to

$$\max_i U_{in} = U(z_{in}, S_n; \theta)$$

Simple example: mode choice

Attributes

	Attributes	
Alternatives	Travel time (t)	Travel cost (c)
Car (1)	t_1	c_1
Bus (2)	t_2	c_2

Utility

$$\widetilde{U} = \widetilde{U}(y_1, y_2),$$

where we impose the restrictions that, for $i = 1, 2$,

$$y_i = \begin{cases} 1 & \text{if travel alternative i is chosen,} \\ 0 & \text{otherwise;} \end{cases}$$

and that only one alternative is chosen: $y_1 + y_2 = 1$.

Simple example: mode choice

Utility functions

$$\begin{aligned} U_1 &= -\beta_t t_1 - \beta_c c_1 \\ U_2 &= -\beta_t t_2 - \beta_c c_2 \end{aligned}$$

where $\beta_t > 0$ and $\beta_c > 0$ are parameters.

Equivalent specification

$$\begin{aligned} U_1 &= -(\beta_t/\beta_c) t_1 - c_1 &= -\beta t_1 - c_1 \\ U_2 &= -(\beta_t/\beta_c) t_2 - c_2 &= -\beta t_2 - c_2 \end{aligned}$$

where $\beta > 0$ is a parameter.

Choice

- Alternative 1 is chosen if $U_1 \geq U_2$.
- Ties are ignored.

Simple example: mode choice

Choice

Alternative 1 is chosen if

$$-\beta t_1 - c_1 \geq -\beta t_2 - c_2$$

or

$$-\beta(t_1 - t_2) \geq c_1 - c_2$$

Alternative 2 is chosen if

$$-\beta t_1 - c_1 \leq -\beta t_2 - c_2$$

or

$$-\beta(t_1 - t_2) \leq c_1 - c_2$$

Dominated alternative

- If $c_2 > c_1$ and $t_2 > t_1$, $U_1 > U_2$ for any $\beta > 0$
- If $c_1 > c_2$ and $t_1 > t_2$, $U_2 > U_1$ for any $\beta > 0$

Simple example: mode choice

Trade-off

- Assume $c_2 > c_1$ and $t_1 > t_2$.
- Is the traveler willing to pay the extra cost $c_2 - c_1$ to save the extra time $t_1 - t_2$?
- Alternative 2 is chosen if

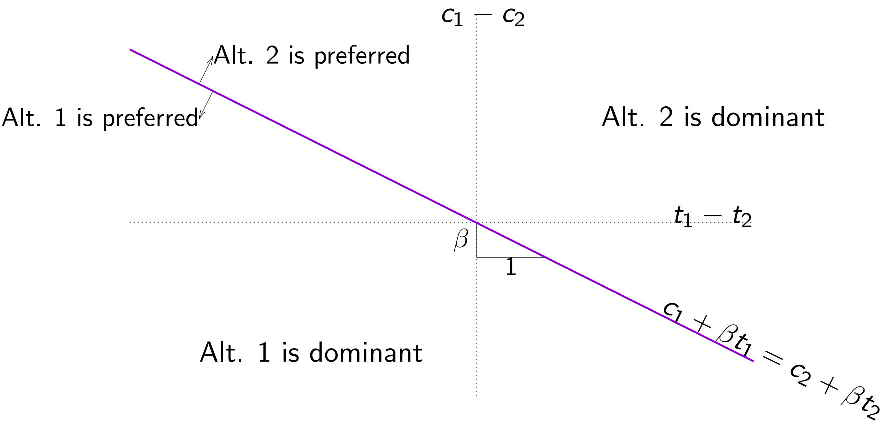
$$-\beta(t_1 - t_2) \leq c_1 - c_2$$

or

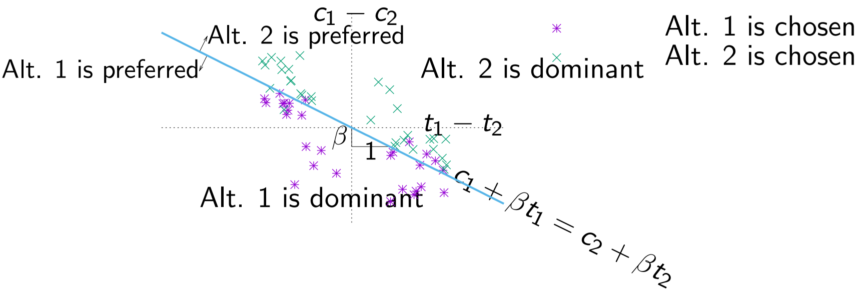
$$\beta \geq \frac{c_2 - c_1}{t_1 - t_2}$$

- β is called the *willingness to pay* or *value of time*

Simple example: mode choice



Simple example: mode choice



Outline

Behavioral validity of the utility maximization?

Assumptions

Decision-makers

- are able to process information
- have perfect discrimination power
- have transitive preferences
- are perfect maximizers
- are always consistent

Relax the assumptions

Use a probabilistic approach: what is the probability that alternative i is chosen?

Introducing probability

Constant utility

- Human behavior is inherently random
- Utility is deterministic
- Consumer does not maximize utility
- Probability to use inferior alternative is non zero

Random utility

- Decision-makers are rational maximizers
- Analysts have no access to the utility used by the decision-maker
- Utility becomes a random variable

Niels Bohr

Nature is stochastic

Albert Einstein

God does not throw dice

Random utility model

Probability model

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}, \forall j \in \mathcal{C}_n),$$

Random utility

$$U_{in} = V_{in} + \varepsilon_{in}.$$

Random utility model

$$P(i|\mathcal{C}_n) = \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in \mathcal{C}_n),$$

or

$$P(i|\mathcal{C}_n) = \Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in \mathcal{C}_n).$$

Concrete models

Model derivation

- Assume a distribution for ε_{in}.
- Derive the probability formula for the choice model.

Probit model

- Assumption: ε_{in} are normally distributed.
- Problem: CDF is involved in the model. No closed form.

Logit model

Assumption: ε_{in} are i.i.d. extreme value: $\text{EV}(0, \mu)$.

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{\mu V_{jn}}}.$$

Choice set

Choice set potentially different for each individual

$\mathcal{C}$={car, train, bus, metro }, $\mathcal{C}_n$={train, bus }

Binary variable for choice set membership: $z^c_{in} \in \{0, 1\}$

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}, j \in \mathcal{C}_n) = \Pr(U_{in} + \ln z^c_{in} \geq U_{jn} + \ln z^c_{jn}, j \in \mathcal{C}) = P(i|z^c, \mathcal{C})$$

Logit

$$P(i|z^c, \mathcal{C}) = \frac{z^c_{in} e^{V_{in}}}{\sum_{j \in \mathcal{C}} z^c_{jn} e^{V_{jn}}}$$

A concrete example: transportation mode choice

Binary choice

- Car
- Train

Utility function for car

$$
\begin{aligned}
V_{in} = {} & 3.04 \\
& - 0.0527 \cdot \text{cost}_{in} \\
& - 2.66 \cdot \text{travelTime}_{in} \cdot \text{work}_n \\
& - 2.22 \cdot \text{travelTime}_{in} \cdot (1 - \text{work}_n) \\
& - 0.850 \cdot \text{male}_n \\
& + 0.383 \cdot \text{mainEarner}_n \\
& - 0.624 \cdot \text{fixedArrivalTime}_n.
\end{aligned}
$$

A concrete example: transportation mode choice

Utility function for train

$$
\begin{aligned}
V_{jn} = & -0.0527 \cdot \text{cost}_{jn} \\
& -0.576 \cdot \text{travelTime}_{jn} \\
& +0.961 \cdot \text{firstClass}_{n}.
\end{aligned}
$$

A concrete example: transportation mode choice

Three individuals

	Individual 1	Individual 2	Individual 3
Train cost	40.00	7.80	40.00
Car cost	5.00	8.33	3.20
Train travel time	2.50	1.75	2.67
Car travel time	1.17	2.00	2.55
Gender	M	F	F
Trip purpose	Not work	Work	Not work
Class	Second	First	Second
Main earner	No	Yes	Yes
Arrival time	Variable	Fixed	Variable

A concrete example: transportation mode choice

		Individual 1	
Variables	Coef.	Car	Train
Car dummy	3.04	1	0
Cost	-0.0527	5.00	40.00
Tr. time by car (work)	-2.66	0	0
Tr. time by car (not work)	-2.22	1.17	0
Tr. time by train	-0.576	0	2.50
First class dummy	0.961	0	0
Male dummy	-0.850	1	0
Main earner dummy	0.383	0	0
Fixed arrival time dummy	-0.624	0	0
V_{in}		-0.6709	-3.5480
$P_n(i)$		0.947	0.0533

A concrete example: transportation mode choice

		Individual 2	
Variables	Coef.	Car	Train
Car dummy	3.04	1	0
Cost	-0.0527	8.33	7.80
Tr. time by car (work)	-2.66	2	0
Tr. time by car (not work)	-2.22	0	0
Tr. time by train	-0.576	0	1.75
First class dummy	0.961	0	1
Male dummy	-0.850	0	0
Main earner dummy	0.383	1	0
Fixed arrival time dummy	-0.624	1	0
V_{in}		-2.9600	-0.4581
$P_n(i)$		0.0757	0.924

A concrete example: transportation mode choice

		Individual 3	
Variables	Coef.	Car	Train
Car dummy	3.04	1	0
Cost	-0.0527	3.20	40.00
Tr. time by car (work)	-2.66	0	0
Tr. time by car (not work)	-2.22	2.55	0
Tr. time by train	-0.576	0	2.67
First class dummy	0.961	0	0
Male dummy	-0.850	0	0
Main earner dummy	0.383	1	0
Fixed arrival time dummy	-0.624	0	0
V_{in}		-2.4066	-3.6459
$P_n(i)$		0.775	0.225

Outline

Parameters

Utility function for train

$$V_{jn} = \bigcirc\!\!\!\!-0.0527 \cdot \text{cost}_{jn}$$

$$-0.576 \cdot \text{travelTime}_{jn}$$

$$+0.961 \cdot \text{firstClass}_{n}.$$

Data

Sample of individuals n

Stratified sampling

Independent variables: x_n

Travel time, travel cost, first class, income, etc.

Dependent variables: y_{in}

Choice: train or car.

Likelihood: one observation

$$P_n(\text{auto}; \beta)^{y_{\text{auto},n}} P_n(\text{train}; \beta)^{y_{\text{train},n}}$$

Maximum likelihood estimation

Estimators for the parameters

Parameters that achieve the maximum likelihood

$$\max_{\beta} \prod_{n} \left(P_n(\text{auto}; \beta)^{y_{\text{auto},n}} P_n(\text{train}; \beta)^{y_{\text{train},n}} \right)$$

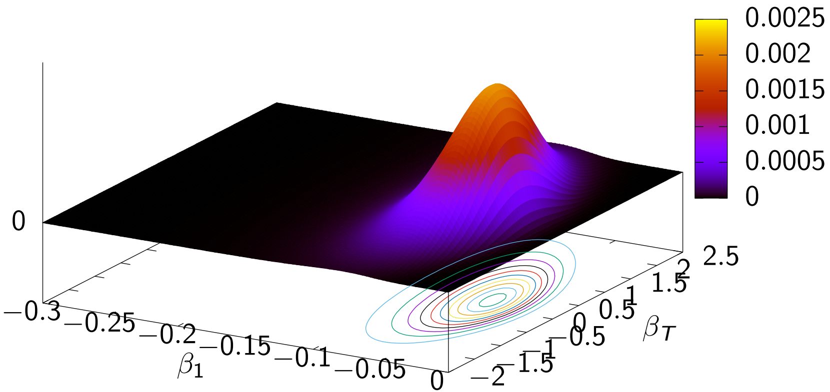
Log likelihood

Alternatively, we prefer to maximize the log likelihood

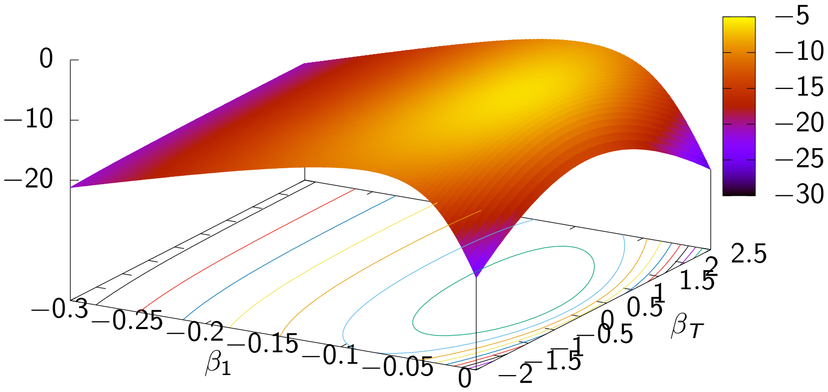
$$\max_{\beta} \ln \prod_{n} \left(P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{train})^{y_{\text{train},n}} \right) =$$

$$\max_{\beta} \sum_{n} y_{\text{auto},n} \ln P_n(\text{auto}) + y_{\text{train},n} \ln P_n(\text{train})$$

Likelihood


0.0025
0.002
0.0015
0.001
0.0005
0
0
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
β_T
−0.3
−0.25
−0.2
−0.15
−0.1
−0.05
0
β_1

Log likelihood



Outline

Using the model

Behavioral model

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

What do we do with it?

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 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

Large table

When the table has too many rows...

apply sample enumeration.

When the table has too many columns...

apply micro simulation.

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

$$\mathcal{L}(\beta_0) = -1858.039$$

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

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

$$\rho^2 = 0.573$$

$$\bar{\rho}^2 = 0.559$$

Example: interurban mode choice in Switzerland

	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%

Forecasting

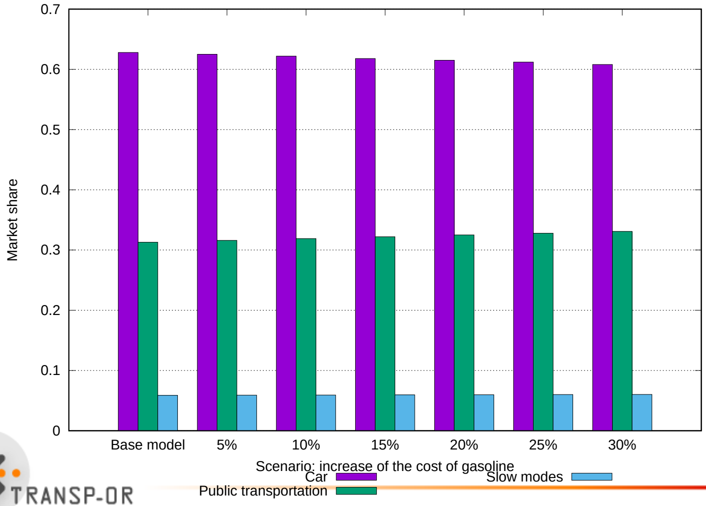
Procedure

- 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



Price optimization

Expected market share

$$W(i) = \frac{1}{N_T} \sum_{n=1}^{N_T} P(i|p_{in}, x_n; \theta).$$

Expected revenue

$$R(i; p_i) = \frac{1}{N_T} \sum_{n=1}^{N_T} p_{in} P(i|p_{in}, x_n; \theta).$$

Price optimization

$$\max_{p_i} R(i; p_i) = \frac{1}{N_T} \sum_{n=1}^{N_T} p_{in} P(i|p_{in}, x_n; \theta).$$

A simple example

Context

- $\mathcal{C}$: set of movies
- Population of N individuals
- Competition: staying home watching TV

One theater – homogenous population

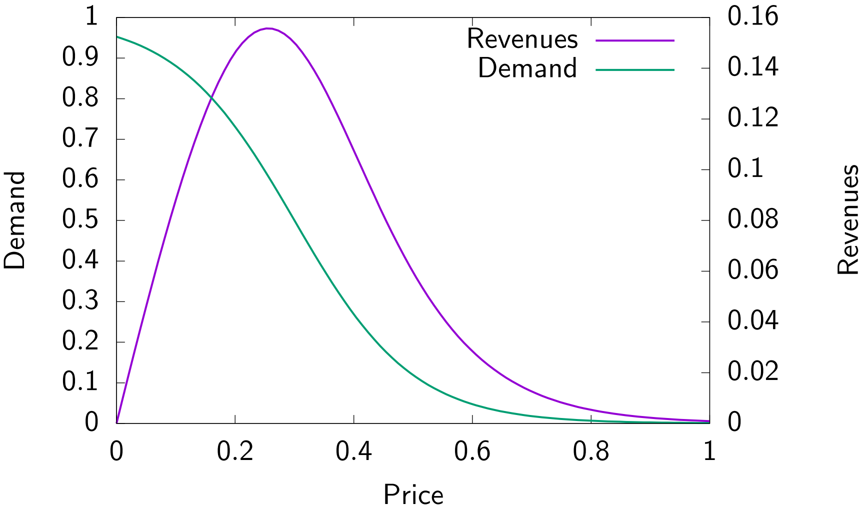
Alternatives

- Staying home: $U_{cn} = 0 + \varepsilon_{cn}$
- My theater: $U_{mn} = -10.0 p_m + 3 + \varepsilon_{mn}$

Logit model

ε_m i.i.d. EV(0,1)

Demand and revenues



Heterogeneous population

Two groups in the population

$$U_{mn} = -\beta_n p_m + c_n$$

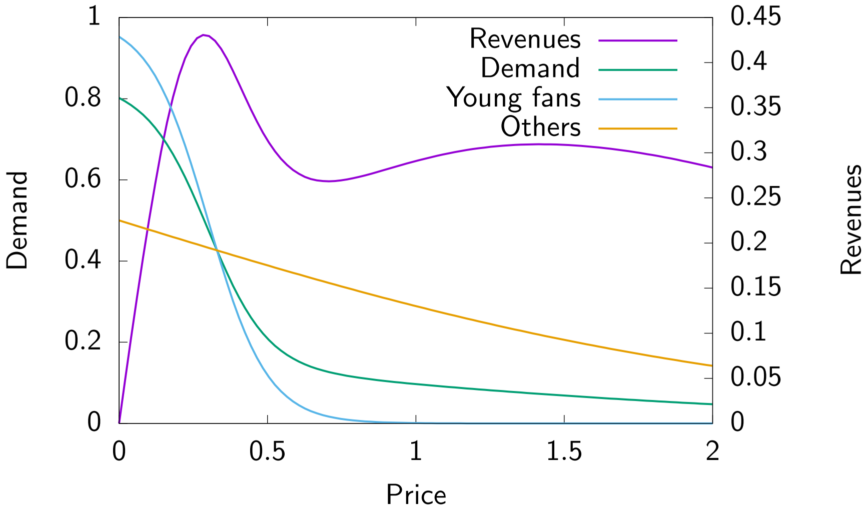
Young fans: 2/3

$\beta_1 = -10, \; c_1 = 3$

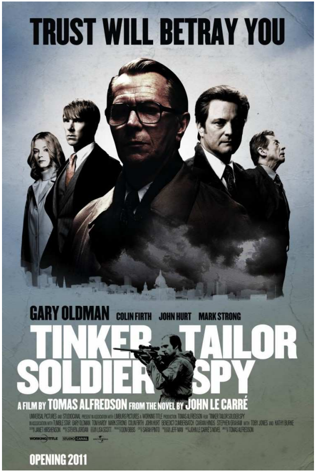
Others: 1/3

$\beta_2 = -0.9, \; c_2 = 0$

Demand and revenues



Two theaters, different types of films



Two theaters, different types of films

Theater m

- Attractive for young people
- Star Wars Episode VII

Theater k

- Not particularly attractive for young people
- Tinker Tailor Soldier Spy

Heterogeneous demand

- Two third of the population is young (price sensitive)
- One third of the population is not (less price sensitive)

Two theaters, different types of films

Data

- Theaters m and k
- $U_{mn} = -10p_m + \textcircled{4}$, $n =$young
- $U_{mn} = -0.9p_m$, $n =$others
- $U_{kn} = -10p_k + \textcircled{0}$, $n =$young
- $U_{kn} = -0.9p_k$, $n =$others

Theater m

- Optimum price m: 0.390
- Young customers: 58%
- Other customers: 36%
- Total demand: 51%
- Revenues: 1.779

Theater k

- Optimum price k: 1.728
- Young customers: 0%
- Other customers: 13%
- Demand: 4%
- Revenues: 0.581

Two theaters, same type of films

Theater m

- Expensive
- Star Wars Episode VII

Theater k

- Cheap (half price)
- Star Wars Episode VIII

Heterogeneous demand

- Two third of the population is young (price sensitive)
- One third of the population is not (less price sensitive)

Two theaters, same type of films

Data

- Theaters m and k
- $N = 9$
- $R = 50$
- $U_{mn} = -10p + \textcircled{4}$, $n =$young
- $U_{mn} = -0.9p$, $n =$others
- $U_{kn} = -10p/2 + \textcircled{4}$, $n =$young
- $U_{kn} = -0.9p/2$, $n =$others

Theater m

- Optimum price m: 3.582
- Young customers: 0%
- Other customers: 63%
- Total demand: 21%
- Revenues: 3.42

Theater k

Closed

Outline

Conclusion

Demand

Demand is a sequence of choices

Choice

Choice is the result of an optimization problem: utility

Operational choice models

Random utility — logit

Parameter estimation

Maximum likelihood estimation

Applications

Market shares prediction — Revenue optimization