Empirical Industrial Organization & Consumer Choice

2a Differentiated Goods and Discrete Choice Models

Summer 2018

Sebastian Kranz / Ulm University

May 15, 2018

Differentiated Goods

- Many industries have a large variety of differentiated products
- Some questions of interest:
  - Cross price effects: e.g. how much do we expect a 1% price increase of a VW Golf to affect market shares of each other car model?
  - What would be the market share if a new product is introduced at a certain price? How would competitors react?
  - Can we estimate firms' production cost when just observing prices, sales and product characteristics?
  - Effects of horizontal mergers on market outcomes.
  - Effects of changes in trade policy.

Descriptive Analysis of a Typical Data Set

Exercise 2a.1: The panel data set cars.dta (Stata format) contains information on sales and attributes of different car models for several European markets (made available by Frank Verbooven).

(a) Load the data into R and take a look at it.

(b) Using the R package dplyr, perform some descriptive analysis and data aggregation to get some exercise in R

1. Compute the total sales of VW in the French market for the year 1990
2. Generate a data frame of total sales for every brand, in every market, for every year.
3. Compute market shares of each brand in every market and year and plot the development of market shares of Volkswagen, Opel, Peugeot and Renault in all markets.
4. Show for every market the fraction of domestic cars, EU imports, and Japanese imports for all years

Data Preparation in R

Common tasks in data preparation and helpful R functions / packages

- Summarizing / transforming data by groups
  - package `dplyr` (good performance, nice intuitive syntax)
    - `dplyrExtras` (from github) contains some extra functionality
  - package `data.table` (fast and powerful, harder syntax and pitfalls for R beginners)
  - functions in base: by, aggregate (slow, not very convenient)
- Joining different data sets
  - join (in dplyr) or merge (in base)
- Reshaping data between wide and long formats
  - reshape (in stats)
  - melt and cast (in reshape2), gather, spread (in tidyr)

(Log-)linear demand systems

Exercise 2a.2: Assume one models differentiated product markets by estimating for each product $j = 1 ... J$ a system of linear demand functions

$$q_j = \alpha_j + \sum_{k=1}^{J} \beta_{jk} p_k + \varepsilon_j$$

or a system of log-linear demand functions

$$\log q_j = \alpha_j + \sum_{k=1}^{J} \beta_{jk} \log p_k + \varepsilon_j.$$

What would be problems if one wanted to estimate such demand systems?

Main approaches to differentiated goods

- There are different approaches to modeling differentiated goods:
  (a) The representative consumer who has a taste for variety.
  (b) Discrete choice models with heterogeneous consumers; goods are bundles of characteristics

Representative Consumer

- A first tradition considers a representative consumer who has a taste for consuming a variety of products
- Demand curve derived from a well-specified utility function and the marginal utility from the consumption of each good is decreasing
- Representative consumer has the incentive to spread consumption across a variety of goods
- The reason that goods are differentiated is typically buried in the parameters of the utility function, not very explicit
- Main implementations:
  - Multi-budgeting approach coupled with the Almost Ideal Demand System (Hausman, 1997)
  - Distance-metric (DM) demand model (Pinske, Slade, and Brett, 2002)

Characteristic Space & Discrete Choice Models

- Locate goods in a space of product characteristics
  - e.g. for cars: horse power, size, fuel economy, price
- Consumers have heterogeneous tastes and place utility weights on different product characteristics
- A consumer typically can buy at most one good (and one unit) $\Rightarrow$"discrete choice" model
- Product variety is a response to preference heterogeneity across different consumers rather than a taste for variety of a representative consumer
- Aggregate (market level) demand is found by summing up the demands of individual consumers
- Seminal contributions:
  - individual choice data: McFadden, 1979
  - market data: Berry, 1994; Berry, Levinsohn and Pakes (1995)

Random Utility Specification of Discrete Choice

- Each consumer $n \in \{1, ..., N\}$ can choose one of J different products
- Utility from choosing product $j \in \{1, ..., J\}$

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

- ε_{nj} a random variable with zero mean, not observed by econometrician
- V_{nj} "representative utility" a function of product attributes & possibly consumer characteristics
- n chooses that product j that gives highest utility U_{nj}.
  - usually also option to buy nothing: "not buying" is just a special product $j = 0$

Common: linear specification of V_{nj}

$$V_{nj} = x_{nj}^{'}\beta$$

- x_{nj}: a vector that can contain...

(a) observed attributes of product j
  - e.g. price, brand, horse power, speed, ...
  - possibly just a dummy for each product (alternative specific constants)

(b) variables that depend on both product attributes and consumer characteristics
  - e.g. "gender dummy * horse power", "product dummy * income", fuel cost to drive to work

- Variations
  - Multinomial models: coefficients can depend on product: β_j (can be replicated in formulation above by product dummies)
  - Mixed models: coefficients differ randomly across individuals: β_n

Logit Models for Discrete Choice

- Very flexible and tractable class of models with several variants. Daniel McFadden won a Nobel Prize mainly for his work on logit models and extensions.

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

- with ε_{nj} i.i.d. extreme value distributed with variance $\frac{\pi^2}{6}\sigma^2$.
  - Synonyms: Gumbel distribution, extreme value type 1
  - Scale parameter σ typically normalized to 1

Theorem

The probability that consumer n chooses product j is given by

$$P_{nj} = \frac{\exp(V_{nj}/\sigma)}{\sum_{h=0}^{J} \exp(V_{nh}/\sigma)}.$$

if and only if ε_{nj} is distributed i.i.d. extreme value with variance $\sigma^2 \frac{\pi^2}{6}$.

Exercise 2a.3: Logit choice probabilities

(a) Write a function in R with name `choice.prob.logit` that takes as argument a matrix of evaluations V_{nj} for N consumers and J products and a scale parameter σ and returns a matrix of choice probabilities P_{nj} for each consumer and product.

(b) Compute expected market shares for each product j given a matrix of V_{nj}

(c) Write another function `sim.choice.logit` that takes a matrix of V_{nj} and scale σ. It draws ε_{nj} from an i.i.d. extreme value distribution, computes the corresponding U_{nj} and returns for each individual n the selected product. Compare the simulated market shares with the expected market shares.

(d) Draw a plot that illustrates for a given individual n and product j how, ceteris paribus, P_{nj} changes in V_{nj}.

(e) Compute and simplify the derivatives $\frac{\partial P_{nj}}{\partial V_{nj}}$ and $\frac{\partial P_{nj}}{\partial V_{ni}}$. When do these derivatives have their largest values? Are these results consistent with your plot from d)?

Non-identification of utility levels

Exercise 2a.4: Show that in the logit discrete choice model, choice probabilities do not change if

(a) a constant amount A is added to all U_{nj} or
(b) all U_{nj} are multiplied by a constant $M > 0$.

- Interpretation: From choice data, we cannot identify levels of utilities. We can only estimate cardinal utility functions that specify preference orderings up to positive affine transformations
    - compare with your basic micro classes
- We can therefore normalize two values of the utility functions
- Result extends to other discrete choice models with $U_{nj} = V_{nj} + \varepsilon_{nj}$

Common normalizations

First normalization

- Set representative utility of some choice j to zero. Typically the "buy nothing" option: $V_{n0} = 0$
- Alternatively, normalize one alternative specific constant to 0

Second normalization

- Normalize the variance of error term ε_{nj}
  - logit models: set scale $\sigma = 1$, i.e. $Var(\varepsilon_{nj}) = \frac{\pi^2}{6}$
  - probit models (ε_{nj} are normally distributed): set $Var(\varepsilon_{nj}) = 1$.
- Alternatively, if utility is linear in money, it can be useful to normalize the utility from money to 1 instead of normalizing σ
  - a price increase by 1€ would then reduce utility by 1

Outlook of Discrete-Choice Models Covered in Class

- Simple logit models of consumer choice
  - given individual choice data (Train, 2009, Ch. 2 + 3) ("conditional logit")
  - given market level data and possibly endogeniety problems (Berry, 1994)
- Nested logit models
  - individual choice data (Train, 2009, Ch. 4)
  - market level data (Berry, 1994)
- Mixed logit models
  - individual choice data (Train, 2009, Ch. 6)
  - market level data (Berry, 1994, Berry, Levinsohn & Pakes (BLP), 1995, Nevo 2001, Train, 2009 Ch 13, Dube, Fox & Su, 2011)

Exploring Individual Choice Data

Exercise 2a.5: Exploring an individual choice data set, long and wide format

(a) Load the dataset Heating in the R package "mlogit" and take a look at it and its description in the help file.

(b) Use the function mlogit.data to transform it from the original wide format (one row per choice situation) to a long format (one row per alternative in each choice situation), which will be suitable for maximum likelihood estimation.

(c) Which columns correspond to individual characteristics, product attributes, and cross effects of indiviudal characteristics and product attributes?

Maximum Likelihood Estimation

- Let y_{nj} be equal to 1 if n has chosen j and otherwise be 0.
- Choice probabilities P_{nj} shall depend on an unknown vector of parameter β
- Likelihood function

$$L(\beta) = \prod_{\forall n} \prod_{\forall j} P_{nj}(\beta)^{y_{nj}}$$

- Maximum likelihood estimator

$$\hat{\beta}^{ML} = \arg\max_{\beta} L(\beta)$$

Maximum Likelihood Estimation

- For numerical reasons, typically the log-likelihood function $\log L(\beta)$ is maximized.

$$\log L(\beta) = \sum_{\forall n} \sum_{\forall j} y_{nj} \log P_{nj}(\beta)$$

- Computation can be much faster and more robust if one provides analytical solutions for the gradient $\frac{\partial \log L(\beta)}{\partial \beta}$, and possibly also for the Hesse matrix of second derivatives $\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta'}$.

Exercise 2a.6: Maximum likelihood estimation of (conditional) logit models

(a) Write down the log-likelihood function for a logit model with $V_{nj} = x_{nj}^{'}\beta$.

(b) Write an R function logLik.logit that takes as parameter β, a matrix of attributes X and a vector of choices y and returns the value of the corresponding logit log-likelihood function.

(c) Simulate a data set for a simple logit model with two products and two product attributes.

(d) Perform maximum likelihood simulation on the simulated data set using your function logLik.logit and the R package maxLik and compare with the true estimated parameters.

(e) Perform in a similar fashion ML estimation using the heating data and a model in which V_{nj} is a linear function of investment and operation costs. Compare with a ML estimation of the same model and data using the function mlogit from the package mlogit.