Market Analysis with Econometrics and Machine Learning

3a Discrete Choice Models

Uni Ulm

Prof. Dr. Sebastian Kranz

SoSe 2020

Discrete Choice Experiments

- In marketing it is popular to elicit consumers' preferences by discrete choice experiments.
- Assume, for example, a car company wants to better understand how car choices depend on the price and other attributes.
- In each round a study participant sees a set of different cars with different attributes and shall choose her preferred alternative. E.g.

	Alternative 1	**Alternative 2**	**Alternative 3**
model	VW Golf	VW ID3	Mercedes A
price	25000	30000	32000
fuel type	petrol	electricity	petrol
horse power	...	...	...
cost per 100km	...	...	...
...			

- This is typically repeated several rounds with modified alternatives (e.g. changing the price). All choices are collected and statistically analysed by estimating discrete choice models.

Discrete Choice Data

- One can also analyse non-experimental discrete choice data. E.g. studying which car, electricity provider or telecommunication company a consumer has chosen.
    - With such field data it is often much more complicated to construct the choice set (all relevant alternatives) than in a choice experiment, however.
    - With field data we also have to be more careful about possible endogeneity problems.
- Discrete choice data sets have been analysed for many applications, e.g. for public transport planning (choice of transportation mode), or climate policies for studying how subsidies affect the choice of more energy efficient heating systems.

Discrete Choice as Utility Maximization

- Assume we have N consumers indexed by n who each can each choose one of J different products.
- n's utility from choosing product $j \in \{1, \ldots, J\}$ shall be U_{nj}
- We assume n chooses that product j that gives highest utility U_{nj}.
  - usually also option to buy nothing: “not buying” is just a special product $j = 0$

Linear Random Utility Specification

We consider utility functions of the following form

$$U_{nj} = \beta_0 + \beta_1 x_{1nj} + \beta_2 x_{2nj} + \ldots + \varepsilon_{nj}$$

ε_{nj} is a random variable with zero mean, called *random utility shock*.

An explanatory variable x_{knj} could be ...

1. an observed attribute of product j
   - e.g. price, brand, horse power, speed, ...
   - possibly just a dummy for each product (alternative specific constants)
2. a variable that depend on both product attributes and consumer characteristics
   - e.g. “gender dummy * horse power”, “product dummy * income”, “fuel cost to drive to work”
     - Our goal will be to estimate the β.

Utilities are not observed in reality

- Our goal will be to estimate the parameters β of the utility function:

$$U_{nj} = \beta_0 + \beta_1 x_{1nj} + \beta_2 x_{2nj} + \ldots + \varepsilon_{nj}$$

- Could we in principle directly estimate the utility function (i.e. its coefficients) as stated by running a linear regression (OLS) on a discrete choice data set?
- No. We only observe choices but not utility levels U_{nj}.
- Instead one typically uses Maximum Likelihood estimation (explained further below). This requires to exactly specify the distribution of the error term (aka random utility shock) $\varepsilon_n j$

Conditional Logit Models for Discrete Choice

- Conditional logit models are a 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.
- Note that naming is often sloppy for discrete choice models: conditional logit models are also called *multinominal logit* models, or just simply *logit* models. Sometimes those alternative names refer to special cases of a conditional logit model, however.
- Conditional logit models assume a random utility specification as above and assume that the random utility shocks ε_{nj} are i.i.d. extreme value distributed with variance $\frac{\pi^2}{6}\sigma^2$.
    - Synonym: Gumbel distribution
    - The scale parameter σ typically normalized to 1
- This distribution is similar to a normal distribution but yields a nice functional for the probability P_{nj} that consumer n chooses product j.

Choice probabilities of conditional logit model

Let $V_{nj} = \beta_0 + \beta_1 x_{1nj} + \beta_2 x_{2nj} + ... + \beta_K x_{Knj}$ be the *deterministic* part of U_{nj} (i.e. everything except for ε_{nj}).

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 the random utility shock ε_{nj} is distributed i.i.d. extreme value with variance $\sigma^2 \frac{\pi^2}{6}$.

R illustration: Choice probabilities of conditional logit model

1. Write a function in R with name `choice.prob.logit` that takes as argument a matrix of deterministic valuations 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.
2. Compute expected market shares for each product j given a matrix of V_{nj}
3. 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.
4. Draw a plot that illustrates for a given individual n and product j how, ceteris paribus, P_{nj} changes in V_{nj}.

- Intuitively, we cannot really identify a person's absolute level of utility just by observing product choices.
    - Indeed our utility function U_{nj} is only a cardinal utility function that specifies preference orderings over lotteries of choices (recall your basic microeconomics classes)
- One can show that the choice probabilities of the conditional logit model don't change if we perform an affine linear transformation of a subjects utility function:

$$\tilde{U}_{nj} = a_n + m_n \cdot U_{nj}$$

Optional: Proof

Consider the transformed utility level $\tilde{U}_{nj} = a_n + m_n \cdot U_{nj}$

If we write the transformed utility level as the sum of representative utility and random utility shock $\tilde{U}_{nj} = \tilde{V}_{nj} + \tilde{\varepsilon}_{nj}$

we have

$$\tilde{V}_{nj} = a_n + m_n \cdot V_{nj}$$

and

$$\tilde{\varepsilon}_{nj} = m_n \varepsilon_{nj}$$

where $\tilde{\varepsilon}_{nj}$ is also iid extreme value distributed with scale parameter $\tilde{\sigma}_{nj} = m_n \sigma$.

The corresponding choice probabilities after this affine linear transformation are therefore

$$\begin{aligned}
\tilde{P}_{nj} &= \frac{\exp(\tilde{V}_{nj}/\tilde{\sigma})}{\sum_{h=1}^{J} \exp(\tilde{V}_{nh}/\tilde{\sigma})} \\
&= \frac{\exp((a_n + m_n V_{nj})/(m_n \sigma))}{\sum_{h=1}^{J} \exp((a_n + m_n V_{nh})/m_n \sigma)} \\
&= \frac{\exp(a_n/(m_n \sigma)) \cdot \exp(V_{nj}/\sigma)}{\sum_{h=1}^{J} \exp(a_n/(m_n \sigma)) \cdot \exp(V_{nh}/\sigma)} \\
&= \frac{\exp(V_{nj}/\sigma)}{\sum_{h=1}^{J} \exp(V_{nh}/\sigma)} = P_{nj}
\end{aligned}$$

This means we have the same choice probabilities for $\tilde{U}_{nj}$ than for U_{nj}.

- For our estimations this means that we can normalize two values of the utility function.

Common normalizations

First normalization

- The `mlogit` package normalizes by default the constant β_0 to zero.
- Alternatively, one can set for every person the deterministic utility V_{nj} of some particular choice j to zero. This is typically done if we have a “buy nothing” option.

Second normalization

- `mlogit` normalizes by default the variance of error term ε_{nj} by setting the scale parameter $\sigma = 1$.
- Alternatively, if assume that the price of a product affects utility in a linear fashion, we could normalize the coefficient β_k before price to -1.
  - A price increase by 1 would then reduce utility by 1.
  - If we also normalize the utility of "buy nothing" to 0, we could then interpret utility levels given a price of 0 as a maximum willingness to pay for that product.

Likelihood Function of the Conditional Logit Models

- We can consider choice probabilities P_{nj} as function of an unknown vector of parameter β:

$$P_{nj}(\beta) = \frac{\exp(V_{nj}(\beta))}{\sum_{h=1}^{J} \exp(V_h(\beta))}$$

$$= \frac{\exp(\beta_0 + \beta_1 x_{1nj} + \ldots + \beta_K x_{Knj})}{\sum_{h=1}^{J} \exp(\beta_0 + \beta_1 x_{1nh} + \ldots + \beta_K x_{Knh})}$$

- The likelihood function measures the probability to exactly observe all the choices in our data set as function of β:

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

where $j(n)$ shall be the product chosen by consumer n in our data set.

- The maximum likelihood estimator $\hat{\beta}^{ML}$ maximizes the likelihood function for a given data set:

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

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

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

- One can solve this optimization problem using standard numerical optimization procedures like a Newton method.
- There are several R packages who implement this estimation for discrete choice models and also compute standard errors, e.g. `mlogit` .
    - These packages typically also provide to the numerical optimization function analytic solutions for the gradient $\frac{\partial \log L(\beta)}{\partial \beta}$, and possibly also for the Hesse matrix of second derivatives $\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta'}$ which can improve optimization performance substantially.