Introduction to discrete choice theory

Stefanie Peer
stefanie.peer@wu.ac.at

Masaryk University, Brno
December 1, 2016

Table of Contents

1. Basics
   - Motivation
   - Modeling framework
   - Estimation
   - Specification & Interpretation
2. Data
   - RP data
   - SP data
   - Combining data sources

Table of Contents

3 Advanced
- Models
  - Nested and cross-nested logit
  - Mixed logit
  - Latent class models
- Alternative Modeling Approaches

4 Summary

About myself

- Bachelor in Economics in Innsbruck
- Master in Port, Transport & Urban Economics at EUR Rotterdam
- PhD in Transport Economics at the VU University Amsterdam
  - *The economics of trip scheduling, travel time variability and traffic information*
  - Modeling of travel-related choices (empirically and theoretically)
- Since 2014: Assistant Professor at the Vienna University of Economics and Business (Department of Socioeconomics)

What is discrete choice modeling?

- People make choices
  - Travel mode, work/ home location, etc.
- The choices imply certain preferences; discrete choice models aim at revealing them
  - Car vs. train
  - Time vs. costs
- Future choices can be predicted once preferences are known
  - Demand forecasts, policy impacts
  - Input to cost-benefit-analyses
    - Prediction of demand
    - Derivation of monetary valuations of attributes

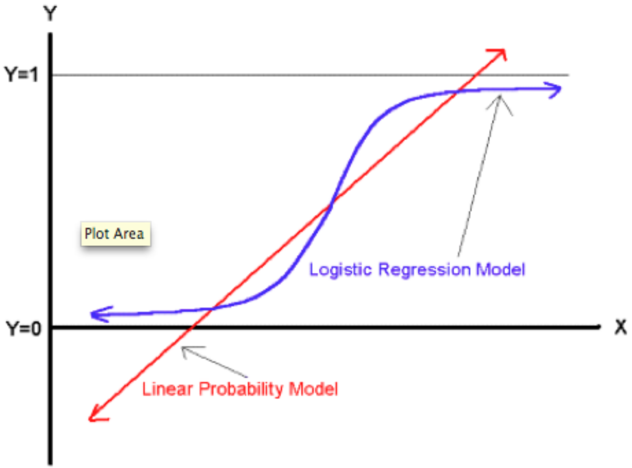
Scope

- Choice modeling is quite ‘math-heavy’
- Understanding of the main concepts is most important for today
- Mathematical notation is used to be precise

Motivation

An econometric perspective

- Many important research topics with 'discrete' dependent variables
  - Voting, product choice, etc.
- Example: 2 discrete alternatives
  - With OLS predicted probabilities can be smaller than 0 and larger than 1
  - Logistic regression constrains the estimated probabilities to lie between 0 and 1


Y
Y=1
Y=0
X
Plot Area
Logistic Regression Model
Linear Probability Model

Motivation

A choice modeling perspective I

Estimate latent preference structure from data on discrete choices in order to understand and forecast choices

- Observe choices (in a real-life or hypothetical choice situation)
- Infer trade-offs between choice alternatives
- Estimate preferences
- Forecast choices

Motivation

A choice modeling perspective II

- Discrete choice theory was developed only in the 70ies (McFadden: received Nobel Prize in 2000)
    - Closely related to traditional microeconomic theory of consumer behavior
    - A way to translate theoretical models into empirical settings
- However, while in theory the goods *per se* generate utility, in discrete choice modeling the *properties* of the goods generate the utility

Motivation

A choice modeling perspective III

Why choice modeling? (Or: why don't we ask directly?)

- Lack of ability for introspection
  - People are not used to reporting trade-offs
  - But they *are* used to make choices
  - Thus: choices as a unit of measurement tend to be more reliable

Motivation

A demand modeling perspective I

- Traditionally, aggregate approaches to measure demand are used
  - Aggregate data
  - Representative consumer approach
  - Aggregate demand is compatible with many forms of demand functions (which one is the "true"?)

Motivation

A demand modeling perspective II

- Discrete choice models as disaggregate approach to measure demand
  - Micro data (from individual decision-making units)
  - Larger number of observations
  - Well grounded in microeconomic theory
  - Explicit modeling of the choice making
    - Available alternatives and their attributes
    - Random disturbances
  - Aggregate demand can be derived from disaggregate choice data
    - Market shares can be derived from average choice probabilities

Transport applications I

In the context of:

- Demand forecasts (e.g. new public transport links, electric cars/bikes, self-driving cars)
- Modal shares
- Traffic flow
- Accessibility
- Environmental issues
- Land use
- etc.

Transport applications II

- **Choices**: routes, modes, car types, subscriptions for public transport/ car sharing/ bike sharing, purchase of traffic information etc. (sometimes decisions are discretized, e.g. departure time)
- **Relevant attributes**: costs, travel time, schedule delays, reliability, level of comfort, waiting time, number of interchanges, etc.
- Often **monetary valuations of the attributes** are derived: value of time, value of reliability, value of comfort, etc.
    - Ratio between marginal utilities
- Numerous applications also in environmental economics, health economics political economics, marketing, etc.

Transport applications III

- The results of discrete choice models are often used as an input for cost-benefit-analyses (CBA) of transport projects
  - Monetary valuations of attributes
  - Demand predictions
- CBA are compulsory in some countries

An example (very simplified)

- Route A: existent slow & cheap train connection
- Route B: new high-speed (& more expensive) train connection
- Trade-off between travel time and costs
- Several observations per person

	Route A	Route B
Travel time (min)	76	65
Costs (Euro)	1	2
Decision		

	Route A	Route B
Travel time (min)	70	40
Costs (Euro)	3	5
Decision		

Example II

	Route A	Route B
Travel time (min)	76	65
Costs (Euro)	1	2
Decision	x	

	Route A	Route B
Travel time (min)	70	40
Costs (Euro)	3	5
Decision		x

Left: B is 10 min faster and 1 Euro more expensive. Decision for A: Person is willing to pay less than 1 Euro for a travel time reduction of 10 min (or $<$ 6 Euro/hour)

Right: B is 30 min schneller and 2 Euro more expensive. Decision for B: Person is willing to pay more than 2 Euro for a travel time reduction of 30 min (or $>$ 4 Euro/hour)

Example III

Decisions can be predicted

- Forecast market share

	Route A	Route B
Travel time (min)	60	50
Costs (Euro)	1.5	4

	Route A	Route B
Travel time (min)	65	45
Costs(Euro)	3.5	5.5

- Assumption: "Value of travel time savings (VoTTS)" = 8 Euro/hour
- Left: VoTTS of 15 Euro/hour $\rightarrow$ A
- Right: VoTTS of 6 Euro/hour $\rightarrow$ B

Questions that can then be answered:

- Should the new connection be constructed?
  - Strongly depends on the travel time reduction and the (monetary) valuation of the reduction (value of travel time savings: VoTTS)
- Potential demand/market share?

Be aware of simplifications

In reality:

- Choice set consists of more than two alternatives
- Other factors play a role too (comfort, etc.)
- New transit service caters more to people with a high VoTTS
- Induced demand
- Etc.

Towards a statistical model

- Approach used in the simplified example is not very practical
  - Simulation by hand
  - Choices are assumed to be made deterministically

Develop statistical model that uses a large number of observations and allows for hypothesis testing

Terminology & Notation

- Decision-making units $n = 1, \ldots, N$
  - Individuals, households, or firms
- Alternatives $j, i = 1, \ldots, J$
  - Products, actions, timing etc.
- Choice set J
  - Set of alternatives
- Attributes z_{jn}
  - Set of characteristics describing a specific choice alternative j for a decision maker n

Set of alternatives

... must be

- Mutually exclusive
- Exhaustive
- The number of alternatives must be finite

Utility functions

- Decision makers maximize an indirect utility function
  - Depends on income and prices - budget constraint is considered indirectly
- Choice probability associated with alternative j depends on the utility associated with all other available alternatives
- Utility is probabilistic
  - Random utility model (RUM), McFadden (1974)
  - Measured variables do not include all relevant factors that determine decision

Utility formulation

- Most common: additive utility function
- However, also utility functions with multiplicative error terms exist
  - Fosgerau, M., Bierlaire, M. (2009) Discrete choice models with multiplicative error terms. Transportation Research Part B, 43 (5), pp. 494-505

Additive utility function

Utility of alternative j in choice by person n:

$$U_{jn} = V(z_{jn}, s_n, \alpha_j; \beta) + \epsilon_{jn},$$

where:

- $V(.)$ is a function known as *systematic* (or: representative) utility
- z_{jn} is a vector of attributes of the choice alternative j (as they apply to n)
- s_n is a vector of characteristics of the decision maker
- α_j is a vector of alternative-specific constants
- β is a vector of unknown parameters
- ϵ_{jn} is the *unobservable* (random) component of the utility function

Utility function: implications

Even if the systematic utility is highest for one alternative, that alternative might still not be chosen...

We can only predict choices up to a probability $\rightarrow$ a higher systematic utility implies a higher choice probability

Choice probability

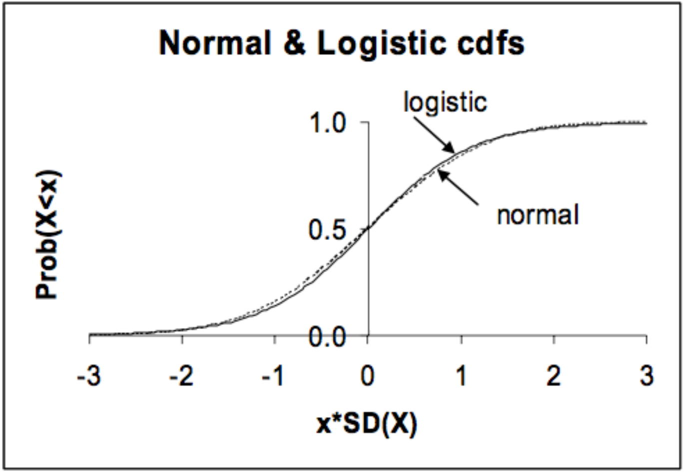
- Probability to choose alternative i:

$$\begin{aligned} P_{in} &= Prob[U_{in} > U_{jn} \text{ for all } j \neq i] \\ &= Prob[V_{in} + \epsilon_{in} > V_{jn} + \epsilon_{jn} \text{ for all } j \neq i] \\ &= Prob[V_{in} - V_{jn} > \epsilon_{jn} - \epsilon_{in} \text{ for all } j \neq i], \end{aligned}$$

  where V_{jn} is a shorthand for $V(z_{jn}, s_n, \alpha_j; \beta)$
- (Cumulative) distribution of random variable $\epsilon_{jn} - \epsilon_{in}$?
- The assumption on the cdf determines the type of model...
  - F is the cdf of the random variable $\epsilon_{2n} - \epsilon_{1n}$

1. Binary Probit
    - Assumption: $\epsilon_{2n} - \epsilon_{1n}$ is standard normal
    - Equivalent: $\epsilon_{2n}, \epsilon_{1n}$ are both normal with variance 0.5 and independent of each other
    - F is then the normal cumulative distribution function
2. Logit
    - Assumption: $\epsilon_{2n} - \epsilon_{1n}$ has a logistic distribution
    - Equivalent: $\epsilon_{2n}, \epsilon_{1n}$ are both Gumbel (also: double-exponential extreme value, Weibull) distributed with mean 0.58 (Euler's constant) and variance $\pi^2/6$
    - F is then the logistic cumulative distribution function

Little difference in the cdfs if scaled accordingly



For **probit** F cannot be expressed in closed form:

$$P_{1n} = \Phi \frac{V_{1n} - V_{2n}}{\sigma},$$

where Φ is the cumulative standard normal distribution function and σ is the standard deviation of $\epsilon_{2n} - \epsilon_{1n}$ (when iid distributed).

- σ cannot be distinguished from the scale of utility

For **logit** a closed form expression for F is available (again for iid distributed error terms):

$$F(x) = Prob[\epsilon_{2n} - \epsilon_{1n} < x] = \exp(-e^{-\mu x}),$$

where μ is a scale parameter (by convention $\mu = 1$). Then:

$$F(x) = \frac{1}{1 + \exp(-x)}$$

$$P_{1n} = F(V_{1n} - V_{2n}) = \frac{1}{1 + \exp(V_{2n} - V_{1n})} = \frac{\exp(V_{1n})}{\exp(V_{1n}) + \exp(V_{2n})}$$

Closed form allows for faster estimation!

Multinomial logit

Generalization of binary logit to J alternatives:

$$P_{in} = \frac{\exp(V_{in})}{\sum_{j=1}^{J} \exp(V_{jn})}$$

Odds ratio P_{in}/P_{jn} depends only on $V_{in} - V_{jn}$, not on the utilities associated with any other alternative: **Independence from irrelevant alternatives (IIA)**

IIA

- Adding new alternatives does not change relative proportions of choices for previously existing alternatives
- If attractiveness of one alternative is increased, the probabilities of all other alternatives being chosen will decrease by identical percentages

IIA violations

- When decision makers perceive alternatives to be close substitutes for each other
- When we omit variables that are common to two or more alternatives
- (Cross-) nested logit models can be used to avoid the restriction IIA imposes (or multinomial probit models)

Probit vs. logit

- Logit much more common, especially in multinomial form - mainly due to closed form properties of logit (no simulation of choice probabilities necessary)
- iid assumption (identically and independently distributed error terms) is restrictive in both models
- iid probit and logit can be generalized for non-iid distributions (to be discussed later)

Important:

- Only differences in utility matter
  - E.g. Adding or subtracting a constant from all utilities in a model has no impact
- Overall scale of utility is irrelevant
  - Normalizing the variance of the error terms is equivalent to normalizing the scale of utility
  - Parameter size and error variance cannot be estimated jointly

Variance

General

- Variance of the random utility term ϵ reflects randomness in behavior of the choice makers as well as unobserved heterogeneity between them
- Little randomness implies almost deterministic model
  - Sudden changes in behavior when (observable) characteristics of the alternatives change
- Much randomness means that behavior changes only gradually if the (observable) characteristics of the alternatives change
- Hence: variance important for prediction!

Variance

- Variance can be represented by the inverse of the scale of the systematic utility function
  - In MNL: $\sigma^2 = \pi^2/(6\lambda_i^2)$
  - $\rightarrow$ Models that fit well display larger scales (i.e. larger (absolute) β)
- Randomness in behavior also produces variety (*entropy*) in aggregate behavior
  - Link between aggregate and disaggregate models
  - Expected maximum utility from choice set increases with more alternatives (*love for variety*)

Estimation of coefficients

- Using data on observed choices (in real or hypothetical setting)
- Find set of parameters that best explain observed choices
- Required information
  - Choice set of each decision maker n
  - Attributes of *all* alternatives considered by decision maker n
    - Note difference to OLS!
  - The actual choice made by n: d_{in}
  - (Characteristics of decision maker n)
    - with $d_{in} = 1$ if i is the chosen alternative, 0 otherwise

Maximum likelihood estimation (MLE) I

Likelihood function (multiply over all observations (n) and all alternatives (i)):

$$L = \prod_{n=1}^{N} \left(P_{1n}(\beta)^{d_{1n}} \times P_{2n}(\beta)^{d_{2n}} \times \cdots \times P_{Jn}(\beta)^{d_{Jn}} \right)$$

Likelihood would become very small for non-trivial datasets. Maximize log-likelihood function instead:

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{J} d_{in} \log P_{in}(\beta)$$

Maximum likelihood estimation (MLE) II

- Derivatives of LL provide information about the preciseness of the estimated parameters
- Variance-covariance matrix Var(β)
  - Diagonal elements give variances of the individual parameters (sqrt is the standard error of the coefficients)
  - Off-diagonal elements give covariances
    - High correlation between two coefficients: difficult to explain variation in choices based on variation in βs (e.g. longer trips are also more expensive $\rightarrow$ difficult to assign variation in choices to either one of the attributes $\rightarrow$ large covariance between β_T and β_C $\rightarrow$ large standard errors for β_T and β_C)

Estimation

Models are estimated by iteratively finding combination of βs that make the observed data most likely.

E.g. Newton-Raphson-method

- First partial derivative of LL wrt to βs gives direction of step
- Second partial derivative of LL wrt to βs gives step size
  - Greater curvature $\rightarrow$ smaller step (maximum is near)

Log-likelihood and model fit

The log-likelihood can be used to assess a model's fit with the data McFadden's $\rho^2 = 1 - \frac{LL(\beta)}{LL(0)}$, where $LL(0)$ is the log-likelihood when all βs are 0

- If $\rho^2 = 0$: model does not do better in explaining than "throwing a dice"
- If $\rho^2 = 1$: perfect fit, deterministic model
- Not equal to R^2

Comparing model fit across models

- If Model A yields LL=-450 and Model B yields LL=-447, which one is better?
- What is the probability that B's fit is better due to coincidence? $\rightarrow$ Likelihood Ratio Test
  - Likelihood Ratio Statistic $LRS = -2(LL_A - LL_B)$
    - B has q more free parameters than A
    - LRS tests if B's better LL is due to coincidence (A being the better model)
    - LRS is distributed χ^2 with q degrees of freedom

Specification of the deterministic utility formulation

- Linear in parameters $\neq$ linear in variables
- With V linear in β, loglikelihood function is globally concave in β
- As usual: completeness vs. tractability
- Base empirical models on explicit behavioral theory
- Goal of transferability

Coefficients

- Different types of coefficients
  - Generic (e.g. cost-coefficient)
  - Alternative-specific (e.g. constants)
  - Interaction (e.g. income, education)
- Note: all person-specific variables s_n must be interacted with an alternative-specific variable or coefficient, otherwise they would cancel out when computing $V_{in} - V_{jn}$

Alternative-specific constants

$$V_{in} = \alpha_i + \beta' z_{in}$$

- α_i can be interpreted as average utility of the unobserved characteristics of alternative i (relative to base alternative)
    - Since only differences in utility count, one ASC must be normalized (usually to 0): "base alternative" (otherwise the model is unidentified)
    - Use of ASC render it difficult to predict the result of adding a new alternative (unless a-priori information on ASC is available)

Interpreting the coefficients

- β: units of utility gained loss by 1 unit increase of attribute
- Estimating β implies inferring the importance of the associated attribute relative to other observed attributes as well as relative to unobserved factors
- Having small βs (i.e. close to 0) is equivalent to saying that the variance of ϵ is large

Interpreting the coefficients
Marginal rates of substitution

- It's easier to interpret ratios of coefficients
- They represent the marginal rates of substitution between two attributes
- Famous example: "Value of travel time savings (VoTTS)" (or "Value of time" (VOT), "Willingness to pay for travel time savings")

$$VoTTS = \frac{\frac{\partial V}{\partial T}}{\frac{\partial V}{\partial C}} = \frac{\beta_T}{\beta_C}$$

The VoTTS is thus the ratio of the impact of a a (marginal) change in travel time on utility and the impact of a marginal change in travel cost on utility

VoTTS cont'd

- Most important measure of benefits in transport appraisals
- Depending on utility specification the VoTTS can vary
    - Across people
    - Across modes (self-selection?)
    - Across travel purposes
    - Across travel times
    - Etc.

Revisiting the example

Choice between two railway connections. Only travel time and costs matter.

- Determine market share of new high-speed line (Route B)

Revisiting the example II

- Assume logit model outcomes are $\beta_T = -0.1$ and $\beta_C = -0.5$, and:

	Route A	Route B
Travel time (min)	50	40
Costs (Euro)	2	3

$$P(B) = \frac{\exp(40 * -0.1 + 3 * -0.5)}{\exp(40 * -0.1 + 3 * -0.5) + \exp(50 * -0.1 + 2 * -0.5)} = 62\%$$

$$P(A) = 1 - P(B) = 38\%$$

Logsum-based consumer surplus I

- "Logsum": gives expected (maximum) utility of the choice set
  - By definition the maximum utility is associated with the chosen alternative
  - But analyst does not know which one is chosen; hence: "expected"
- Important metric
  - Can measure welfare impact of joint changes in multiple attributes of many alternatives
  - Can measure welfare impact of introducing or removing alternatives from the choice set

Logsum-based consumer surplus II

- Logsum can be translated into (expected) consumer surplus (benefits in monetary terms)
  - By dividing through the marginal utility of income (proxy: cost/reward coefficient is estimated: β_C)
  - Implies linear treatment of travel cost and absence of income effects

$$E(CS_n) = \frac{1}{|\beta_C|} E[\max_j (V_{jn} + \epsilon_{jn})]$$

Two data sources

- **Stated preference (SP) data**: hypothetical choices
- **Revealed preference (RP) data**: actual (real-life) choices

RP data

Main characteristics (I)

- Choice behavior in actual choice situation
- Preference information from observed choices (sometimes reported)
- Choice set ambiguous/unobservable in many cases
- Responses to non-existent alternatives cannot be measured
- Sometimes not feasible to observe multiple choices per person (i.e. no panel setting)

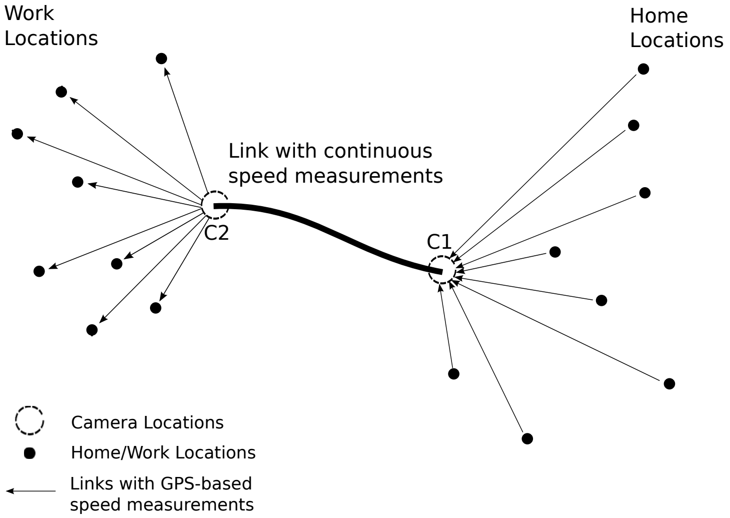
RP data
Main characteristics (II)

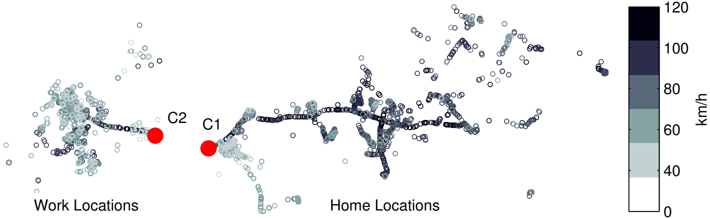
- Attributes
  - Often correlated
  - Limited ranges
  - Ambiguous/unobservable/biased $\rightarrow$ measurement errors, e.g.
    - Travel time expectations: definition? learning from past experience? traffic information? person-specific?
    - Schedule delays: w.r.t. which preferred arrival time? usual arrival time? arrival time without (recurrent) congestion?
  - Note: attributes must be known for chosen as well as unchosen alternatives
    - Engineering values?
    - Perceived values?
- Generally difficult & expensive to collect

An example from...

Peer, S., Knockaert, J., Koster, P., Tseng, Y.-Y., Verhoef, E. 2013. *Door-to-door travel times in RP departure time choice models: An approximation method using GPS data.* Transportation Research. Part B: Methodological 58, pp. 134-150

Attributes for non-chosen alternatives, using geographically weighted regression to predict person-specific, time-of-day-specific and day-specific travel times


Work
Locations
Home
Locations
Link with continuous
speed measurements
C2
C1
Camera Locations
Home/Work Locations
Links with GPS-based
speed measurements


C2
C1
Work Locations
Home Locations
120
100
80
60
40
0
km/h

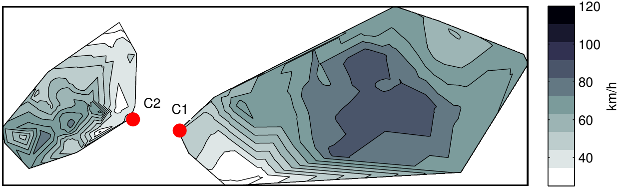


Figure: Predictions: C1–C2 speed = 50 km/h

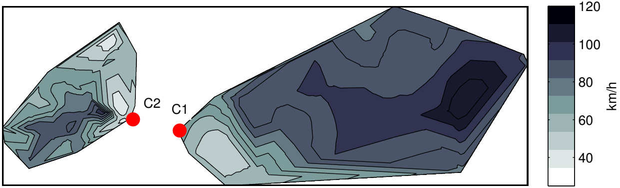


Figure: Predictions: C1–C2 speed = 100 km/h

SP data

Main characteristics (I)

- Choice behavior in hypothetical choice situation
- Various types of preference information feasible (choice, ranking, rating, matching, etc.)
- Choice set specified by researcher
- Preferences for non-existent alternatives can be measured
- Panel setup can be easily achieved

SP data

Main characteristics (II)

- Attributes
  - Multicollinearity can be avoided by choice design
  - Ranges determined by researcher
  - No measurement errors
- Usually fairly convenient & cheap to collect

Hence, compared to RP data, SP data...

- Tend to be "cleaner" (i.e. more controlled, well-defined attributes and choice sets, little correlation between attribute values)
- Can be used to investigate choice alternatives that are not present in reality (e.g. to predict structural, long-run changes such as a new route that reduces travel time substantially)

However, SP estimates might be biased...

- Choices might be incongruent with actual behavior
- Strategical interests (e.g. in order to affect future implementation of policies)
- Range of attribute values presented matters
- Difficulties to understand choice task
- Format of the choice task (e.g. representation of reliability or comfort not straightforward)

An example from...

Tseng, Y.-Y. et al. (2007) A pilot study into the perception of unreliability of travel times using in-depth interviews. Journal of Choice Modelling, 2(1), pp. 8-28

Different representations of travel time variability in SP...

In this version we show you the 5 possible travel times below each other.

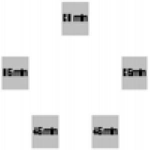
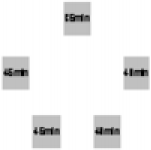
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?

Trip A	*Trip B*
Mean travel time: **40 min**	Mean travel time: **41 min**
You have an equal probability of each of these 5 travel times:	You have an equal probability of each of these 5 travel times:
35 min	**30 min**
40 min	**35 min**
40 min	**45 min**
40 min	**45 min**
45 min	**50 min**
Cost: **€ 3,80**	Cost: **€ 2,80**

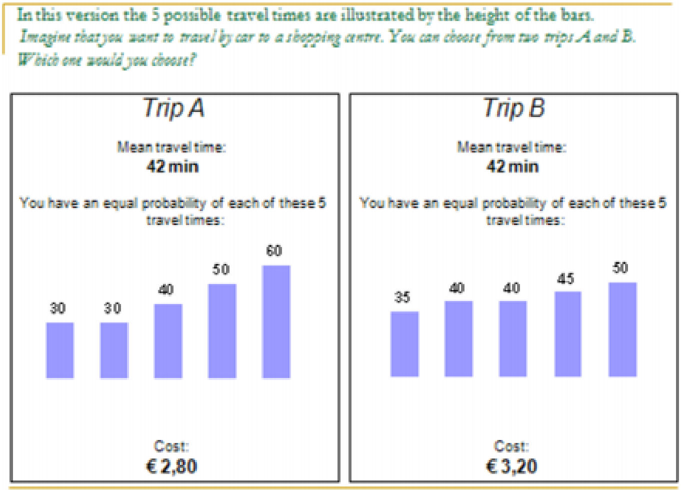
A

In this version we show you the 5 possible travel times as points on a circle.

Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?

Trip A	Trip B
Mean travel time: **44 min**	Mean travel time: **41 min**
You have an equal probability of each of these 5 travel times: 	You have an equal probability of each of these 5 travel times: 
Cost: **€ 2,80**	Cost: **€ 3,60**

B


In this version the 5 possible travel times are illustrated by the height of the bars.
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?
Trip A
Mean travel time:
42 min
You have an equal probability of each of these 5 travel times:
30
30
40
50
60
Cost:
€ 2,80
Trip B
Mean travel time:
42 min
You have an equal probability of each of these 5 travel times:
35
40
40
45
50
Cost:
€ 3,20

C

Combining SP and RP data

What can be gained?

- **Traditional view**: SP data should be used to enrich RP data
  - Based on the notion that RP data are *true* data source and therefore superior
  - Use SP data to correct for deficiencies of RP data (e.g. correlation between attribute values)
- **(More) recent view**: No superior data source
  - Each data source captures those aspects of the choice process for which it is superior
  - Hence: Stronger role of SP, probably as a consequence of advancements in research (e.g. pivoting of SP-attributes around status-quo: Hensher, 2010)

Benefits from combining (pooling) SP and RP...

... can be expected if:

- Common theoretical model underlying both datasets
- Similar structural form of the data (similar attribute definitions)
- Ratios of SP and RP parameters similar across attributes (when estimated separately)

Scale

- Scale may differ between between SP and RP
- Scale of one data source must be fixed to 1, otherwise identification is not possible
  - Usually variance is expected to be larger in RP data because of unobserved factors (SP more controlled)
  - However, no a priori theoretical basis for assuming that one of the variances is larger than the other

Example: Brownstone & Small, 2005 (I)

Valuing time and reliability: assessing the evidence from road pricing demonstrations
(Transportation Research-Part A)

- Probably most influential SP–RP paper in transport economics
- They review various studies, mainly covering two express-lane projects in the US (SP, RP, SP–RP data): focus on route choice
- Frequent outcome that RP estimates of the VOT are higher than SP estimates, by roughly a factor 2
  - E.g. Brownstone and Small, 2005; Ghosh, 2001; Hensher, 2001; Isacsson, 2007; Small et.al., 2005

Example: Brownstone & Small, 2005 (II)

- Suggest 2 possible explanations
  1. Time inconsistency: React more strongly to cost in laboratory setting
  2. Travel time misperception in reality
     - If in real life an individual perceives a 10-minute delay as 20 minutes, he probably reacts to a 20-minute delay in an SP setting in the same way as he would to a 10-minute delay in reality ($\rightarrow$ SP-based VOT half of RP-based VOT)
     - RP results correspond to what planners need to know in order to evaluate transportation projects

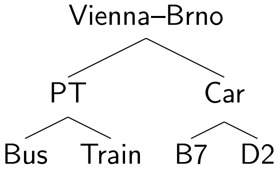
Main limitations of standard (multinomial) logit models

- Cannot represent random taste variation (differences in taste that cannot be linked to observed characteristics)
- Cannot represent unobserved categories of alternatives in a choice set ("nests")
    - E.g. dislike of all public transport alternatives
- Imply proportional substitution patterns (IIA)
- Cannot capture the dynamics of repeated choice (unobserved factors are correlated over choices/time)

Nested logit
Idea

- Allows for intra-choice correlation in preferences for a subset (a "nest") of choice alternatives (i.e. correlated random terms)
- It groups alternatives that are similar to each other in unobserved ways ("nests" are determined by researcher, preferably following some theoretical intuition)
- Relieves IIA assumption
- IIA holds within nests but not across nests

Example: nested logit



Note: It does not necessarily represent a sequential choice!

Cross-nested logit

Idea

- Generalization of the nested logit
- Alternatives can belong to more than one nest
- Allocation parameter that describes the proportion of membership of alternative j to nest k can be:
  - fixed
  - estimated

Mixed logit (error component models)

- **Allow coefficient(s) β to have any distribution**
  - Allow for random taste variation
  - Allow for flexible substitution patterns
  - Allow for correlations over time
- No closed form
  - Outer integration (over the distribution defining random parameters) using simulation methods
  - Inner integration (over remaining additive errors ϵ_{jn}) yields logit formula (no simulation needed)
  - Higher number of draws leads to a better representation of the probability density function, but also to (very) high computation times

Latent class models
Idea

- 2 or more classes
- Within each class: MNL
- Probabilistic (usually (multinomial) logit) model for class membership (with or without explanatory variables)
- Possible to fix coefficients across classes
- In contrast to mixed logit models, which assume a continuous distribution of (some) parameters, latent class models do not require any assumptions regarding the shape of the distribution of a given parameter (hence, no simulation needed)
- Panel setup possible
- Increasingly popular

Maximum score estimation

- Maximize the number of correct predictions (Manski, 1975, Econometrica)
- Advantages
  - Simple implementation (grid search)
  - Robust to heteroskedasticity, serial correlation and generally to mis-specifications of the distribution of ϵ_{jn}
- Disadvantages
  - Gradient-based methods are not feasible (hence: standard errors only via bootstrapping)
  - Slow convergence

Regret minimization (instead of utility maximization)

- Especially propagated by the group of Caspar Chorus (TU Delft)
- Core assumptions:
  - People choose alternative with minimum regret: avoiding (relatively) weak performance is more important than attaining (relatively) strong performance
  - Losses (relative to reference point) loom larger than gains of equal magnitude
  - Relative popularity of two alternatives depends on availability and performance of other alternatives in the choice set (choice set dependency)
- Performs sometimes (but not always) better than utility maximization
- More complex than utility maximization

Estimation software

- The estimation of probit and logit models is possible in all standard econometrics packages
  - E.g. STATA, Eviews, SPSS
- Many dedicated packages in R and Matlab
- Dedicated software: **Biogeme**, Alogit
  - `http://biogeme.epfl.ch/`
  - Standard Bison version (with GUI)
  - Python-based version
  - Find out more at the workshop tomorrow!

To sum up...

- Discrete choice approaches widely used
- SP and RP data with source-specific advantages and disadvantages
- Nested & mixed logit, as well as panel latent class models as extensions to the basic MNL
- Various new developments due to increase in computing power availability (supercomputers)

Main references

- Train, K. (2002) *Discrete Choice Methods with Simulation*, Cambridge University Press Kenneth E. Train (available online for free!)
- Louviere, J., Hensher, D., Swait, J. (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press
- Small, K., Verhoef, E. (2007) *The Economics of Urban Transportation*, Routledge

Thank you for your attention!

Questions? Comments?