Introduction to choice modelling and data requirements

Stephane Hess
stephane.hess@gmail.com

Outline

Why do we model choices?

Why do we model choices?

Humans make choices all the time

- Products
  - consumer goods
  - food
  - ...
- Services
  - travel
  - healthcare
  - ...
- Life-style
  - activities and time management
  - social interactions
  - ...

Why do we model choices?

Not all choices are the same

- Short-term *vs* long-term
- Frequency
- Complexity
- Consequentiality

Why do we model choices?

Choices and demand

- Choices drive demand
- Economic, societal and environmental impacts
- Policy-makers and industry need understanding and prediction of demand

Why do we model choices?

Two main reasons for modelling choice behaviour

"Understanding" current behaviour

- Why does somebody choose a specific product in a given setting?
- Used in appraisal, cost-benefit analysis, etc

Predicting future behaviour

- What would somebody choose in a future setting?
- Used in demand forecasting and scenario testing

Why do we model choices?

Choice modelling across disciplines

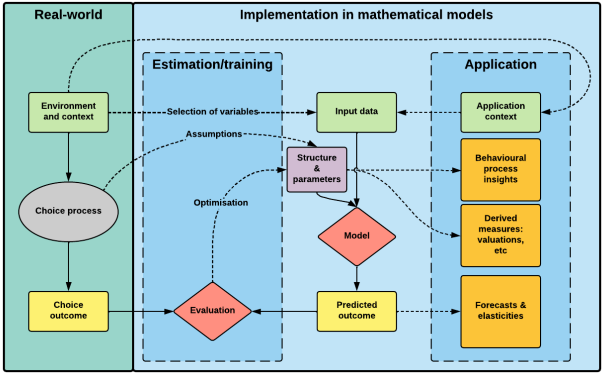
- Transport: mode choice, route choice, destination, car ownership, departure time
- Health decisions: treatment choices, insurance, lifestyle
- Environmental context: energy choices, donations to wildlife protection
- Location choices: migration, home location, rental *vs* ownership, work location
- Career and education: topics of study, full-time *vs* part-time, career area
- Consumption choices: consumer goods, products and services, food
- Investment choices: shares, pensions
- Social choices: interactions with other people, joint decisions
- ...

How do we model choices?

How do we model choices?

The whole story

- Inidividuals make choices
- Analyts represent choices in a model
  1. Define a structure
  2. Estimate parameters
  3. Apply model and produce outputs



How do we model choices?

Types of models

Aggregate *vs* disaggregate

- Aggregate
  - choice shares for people with specific characteristics
- Disaggregate
  - data points are individual choices
  - does not mean we model or explain individual choices

Discrete *vs* discrete-continuous

- Discrete choices
  - one product or service at a time
- Extension to discrete-continuous
  - people choosing different quantities of different products
  - substantially more complex

How do we model choices?

Key principles

Decision context

- what is the choice about?
- what environment/setting is choice made in?

Decision maker

- 'entity' that makes the choice
- individual, household/group, company/organisation

Choice set

- set of possible options/alternatives
- finite, but exhaustive
- mutually exclusive in discrete choice
- alternatives described by attributes

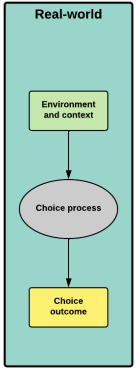
Decision rule

- representation of choice process
- key step, making many assumptions
- major trade-offs between realism, pragmatism and ease of use

How do we model choices?

Real-world process

- Decision-maker:
  - faces choice situation
  - uses an internal process
  - reaches outcome
- Analyst:
  - observes inputs (maybe in part)
  - observes outcome
  - does NOT observe process



How do we model choices?

From real-world process to mathematical model

Step 1: observe choice

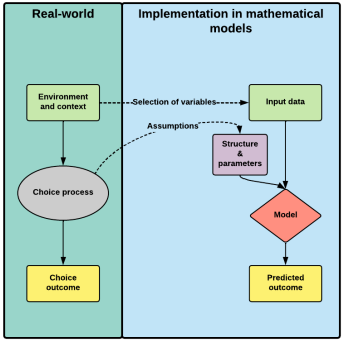
- dependent variable for our model (Y)

Step 2: identify factors influencing choices

- characteristics of alternatives (x), choice setting (w), and decision-maker (z)

Step 3: Build model

- $Y = m(\beta, x, w, z)$, where $m()$ reflects model structure and β are parameters



How do we model choices?

Estimation *vs* prediction

$$Y = m(\beta, x, w, z)$$

Estimation of parameters

- Step 1: make a decision on model structure (m)
- Step 2: estimate parameters (β)
  - find parameters that best explain observed choices given observed choices and assumed model
- Note that x, w and z are fixed

Model application/prediction

- Step 1: determine application context (changing x, w and/or z)
  - e.g. products are added or removed
- Step 2: apply previously estimated model
  - use previously estimated model to predict choices in new scenarios
- Note that now it is β that is fixed

Data types for choice modelling

Data types for choice modelling

Requirements for model estimation

- Sample with many observations
  - potentially with multiple observations per individual
- Variation across observations in attribute level combinations
  - increases chances of trade-off behaviour
- Information on all relevant attributes
  - for decision-maker, alternatives (including availability), and choice context
- Information on choice outcome
  - for discrete choice, only need information on chosen alternative
  - for other formats, need ranking, allocation, etc

Question:

Is information on past choices useful?

Data types for choice modelling

A data example: choice of travel mode

z			w		x								Y
ID & Socios			Context		Drive (1)		Bus (2)			Train (3)			
Person	age	female	weekday	commute	time	cost	time	walk	cost	time	walk	cost	choice
1	55	1	1	1	35m	£5	40m	10m	£3	20m	10m	£7.5	3
2	35	0	0	1	20m	£7.50	30m	5m	£4	25m	15m	£10	1
3	42	1	0	0	50m	£6	40m	5m	£3	35m	20m	£10	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Data types

- Two core possibilities
  - Revealed preference data
  - Stated preference data

Data types for choice modelling

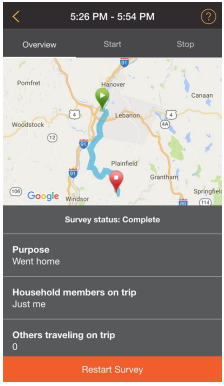
Revealed Preference data

- Mainly used in transport
- Transport use focusses on data relating to an actual trip
- Traditionally used recall data (reported trips)
  - details on single trip or travel diary
- Increasing move to automatically captured data (revealed trips)
  - road sensors, cameras, GPS tracking
- More and more possibilities also in other areas
  - loyalty card data, medical records, etc

Data types for choice modelling

Manual and automatic trip reporting

DAY	Sunday						
	Start		End		What were you doing (mode)	Where	
N°	Hour	Min	Hour	Min		Street 1	Street 2
1	10	0	11	20	Wake up, breakfast	Michimalongo 15	NA
2	11	20	14	0	Tidy up at home	Michimalongo 15	NA
3	14	0	14	5	Going to the shop (walk)	NA	NA
4	14	5	14	10	In the shop	Michimalongo central	NA
5	14	10	14	15	Going back home (walk)	NA	NA
6	14	15	15	30	Lunch	Michimalongo 15	NA
7	15	30	15	40	Going to a friend's home (walk)	NA	NA
8	15	40	15	50	At a friend's home	yerbas buenas alto	NA
9	15	50	16	0	Going back home (walk)	NA	NA
10	16	0	19	0	Stay at home	Michimalongo 15	NA
11	19	0	19	30	Going to the doctor (walk)	NA	NA
12	19	30	21	30	Staying at the doctor	Salas	O'Higgins
13	21	30	22	0	Going back home (walk)	NA	NA
14	22	0	0	0	Stay at home, sleep	Michimalongo 15	NA



Data types for choice modelling

Non-transport examples

NHS choices

Data types for choice modelling

Advantages of RP data

- RP data contains real choices
  - people actually made the trips, ate the meals, spent the money, etc
- No concern about the choice settings not being realistic
- These advantages cannot be overstated!
  - in many areas of economics, RP is seen as the only truth

Data types for choice modelling

Disadvantages of RP data

- Getting access can be difficult
- Cannot really study choice contexts that do not yet exist
- Potentially correlation between attributes and limited trade-off behaviour
  - affects ability to estimate sensitivites
- Errors/gaps in data
  - availabilities of alternatives
  - attributes of unchosen alternatives
  - respondent characteristics
  - difference between what was perceived by decision maker and recorded by analyst

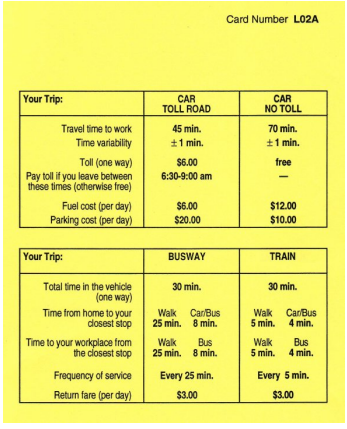
Data types for choice modelling

Stated Preference data

- A (shrinking) majority of DCM applications use Stated Preference (SP) data
- Different types of SP, but key principle is that setting is hypothetical
- Main focus on Stated Choice (SC)
  - hypothetical choice scenarios between mutually exclusive alternatives
  - respondents typically faced with multiple choice situations

Data types for choice modelling

Paper & pen survey



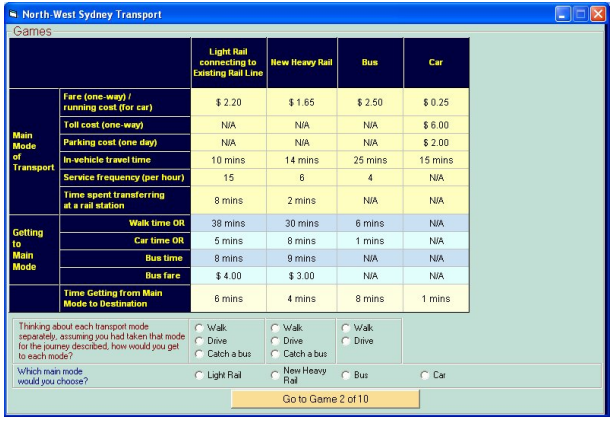
Card Number **L02A**

Your Trip:	CAR TOLL ROAD	CAR NO TOLL
Travel time to work	45 min.	70 min.
Time variability	±1 min.	±1 min.
Toll (one way)	$6.00	free
Pay toll if you leave between these times (otherwise free)	6:30-9:00 am	—
Fuel cost (per day)	$6.00	$12.00
Parking cost (per day)	$20.00	$10.00

Your Trip:	BUSWAY	TRAIN
Total time in the vehicle (one way)	30 min.	30 min.
Time from home to your closest stop	Walk 25 min. Car/Bus 8 min.	Walk 5 min. Car/Bus 4 min.
Time to your workplace from the closest stop	Walk 25 min. Bus 8 min.	Walk 5 min. Bus 4 min.
Frequency of service	Every 25 min.	Every 5 min.
Return fare (per day)	$3.00	$3.00

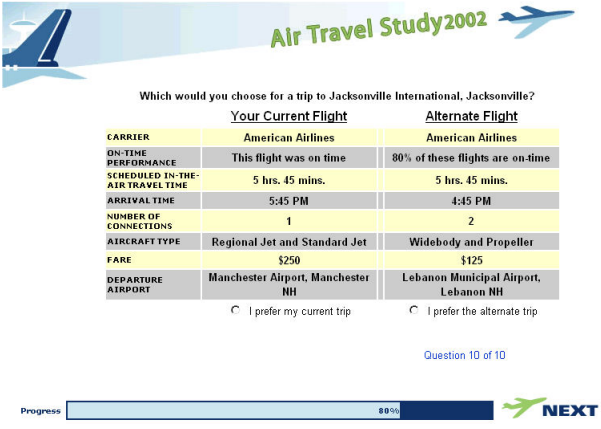
Data types for choice modelling

Computer aided personal interview (CAPI)



Data types for choice modelling

Internet-based survey



Data types for choice modelling

Some very simple choices

	Option A	Option B
One way fuel cost	£33.30	£35.00
One way travel time by car	4 hours 23 minutes	3 hours 30 minutes

○ Option A ◉ Option B

Data types for choice modelling

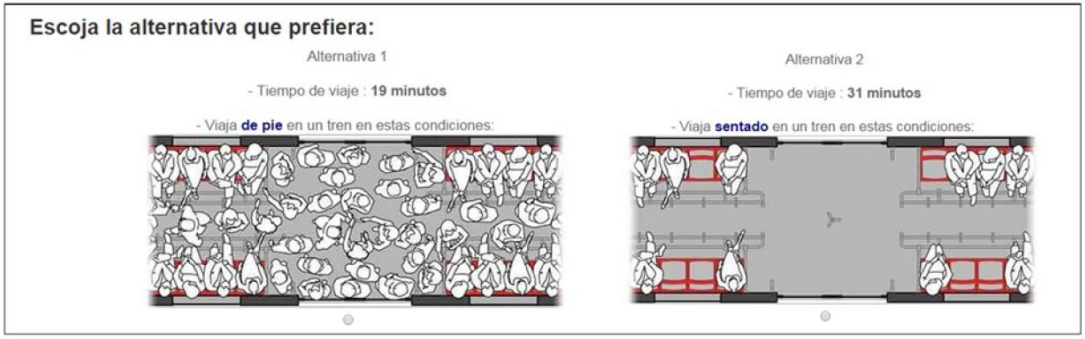
A little more difficult

	Phone A	Phone B	Phone C
Brand	iPhone	Samsung	HTC
OS	iOS	Android	Android
Minutes	150	200	300
SMS messages	unlimited	200	100
Internet	2GB	1.5GB	1GB
Cost per month	£45	£40	£38
Cost phone	£130	£100	£120
CHOICE	□	□	□

Table: Which phone do you prefer?

Data types for choice modelling

Crowding in public transport



Data types for choice modelling

An example in health

	Nurse gives IVs in your home	You have your IVs in hospital	You give IVs to yourself at home
Number of treatments each day	One	One	Two
Appointment times given	Daily appointment time given	Daily appointment time not given	No appointment needed
Who gives the IVs?	Specialist IV antibiotic nurse	Doctor	You give the IVs yourself after half a day of training
Communication between you and healthcare professionals (HCPs)	See a HCP who does not know you	See a HCP who knows you	Speak on the phone with a HCP who knows you
Aftercare from healthcare professionals after the end of treatment	None	Appointment at hospital with nurse	Appointment with your GP
Risk of a problem such as another infection or having to go into hospital	1 in 6 chance	1 in 10 chance	1 in 10 chance
Please tick which service you would prefer to have:	❑	❑	❑

Data types for choice modelling

Improvements in water quality

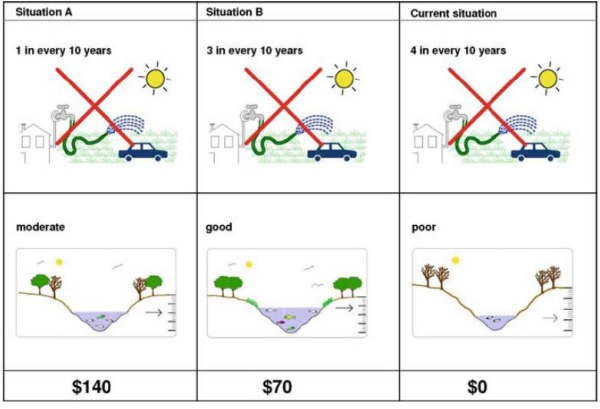
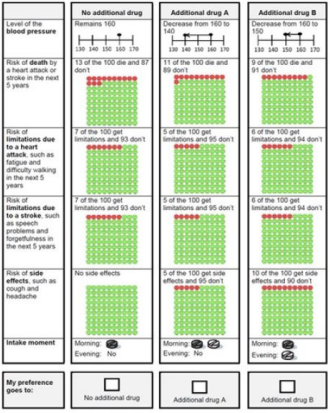


Fig. 1 Example choice card

Data types for choice modelling

Cholesterol drugs

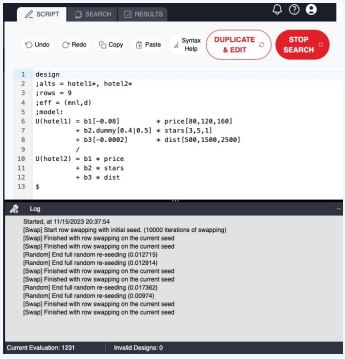


Data types for choice modelling

Survey design

- Asymptotically, random combinations of values would work
  - risk of illogical scenarios & bad data quality
- Aim of experimental design:
  - develop surveys so as to maximise information with limited number of observations
- Design determines values of various attributes in each choice situation
- Limited number oflevels for each attribute
- Often use variations around values for observed choice (e.g. trip)

NGene example



Data types for choice modelling

Advantages of SP

- Can look at hypothetical choice scenarios
  - alternatives that do not yet exist
- Analyst designs survey
  - perfect information on choice situation
  - design encourages trade-off behaviour
  - attributes and levels chosen by analyst

Covid vaccine choices before development

Please consider the following vaccination options and make your choice as if they happened in the current environment. Please remember there is no right or wrong answer.

	Vaccine A	Vaccine B	No vaccine
Risk of infection (out of 100,000 people coming in contact with infected person):	500 (0.5%)	4,000 (4%)	7,500 (7.5%)
Risk of serious illness (out of 100,000 people who become infected):	4,000 (4%)	2,000 (2%)	20,000 (20%)
Estimated protection duration:	two years	five years	
Risk of mild side effects (out of 100,000 vaccinated people):	5,000 (5%)	100 (0.1%)	
Risk of severe side effects (out of 100,000 vaccinated people):	1 (0.001%)	1 (0.001%)	
Population coverage:	40%		
Exemption from international travel restrictions:	exempt		restrictions apply
Waiting time (free vaccination):	6 months	6 months	
Fee (no waiting time):	£400	£400	

	Vaccine A free	Vaccine A paid	Vaccine B free	Vaccine B paid	No vaccine
Yourself:	○	○	○	○	○

Data types for choice modelling

Disadvantages of SP

- Need decisions on attributes, alternatives, segments, contexts, ...
  - blessing and curse
  - are all relevant attributes included?
- Realism
  - respondents can change their choices more easily than in real life
  - not spending real money, time, etc
  - no risk or consequences
  - temptation to think SP can be used for anything

Who would get into a flying AV?

Data types for choice modelling

How should we decide between RP and SP?

- Depends on:
  - Existence of real-world choice context
  - Intended use of model outputs

Data types for choice modelling

Case I: real-world choice exists

- RP should be a our first option
- Reasons to use SP:
  - access & cost for RP
  - lack of trade-offs in RP
- Should only use SP is necessary

Value of time: RP vs SP

	Option A	Option B
One way fuel cost	£33.30	£35.00
One way travel time by car	4 hours 23 minutes	3 hours 30 minutes

Option A Option B

Data types for choice modelling

Case II: real-world choice does not exist

- Generally, SP seen to be only option
- Insights might be possible from related RP choices
- Not if choice is completely new

Mode choice pre-BART

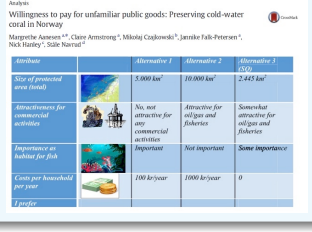
Flu vaccine pre-Covid

Data types for choice modelling

Ability to collect data does not ensure validity of insights

- Transport: value of statistical life, fundamentally new modes of transport (e.g. AV)
- Health: organ donation, Qaly work
- Environment: existence value

Things you don't know how to value



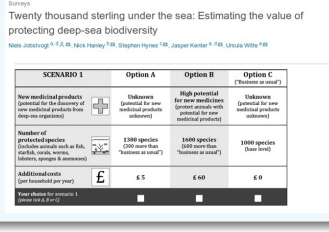
Analysis

Willingness to pay for unfamiliar public goods: Preserving cold-water coral in Norway

Margrethe Aanesen [a,b], Claire Armstrong [a], Mikolaj Czajkowski [b], Jannike Falk-Petersen [a], Nick Hanley [c], Ståle Navrud [d]

Attribute		Alternative 1	Alternative 2	Alternative 3 SQ
Size of protected area (total)		5.000 km²	10.000 km²	2.445 km²
Attractiveness for commercial activities		No, not attractive for any commercial activities	Attractive for oil/gas and fisheries	Somewhat attractive for oil/gas and fisheries
Importance as habitat for fish		Important	Not important	Some importance
Costs per household per year		100 kr/year	1000 kr/year	0
I prefer				

Valuing things you don't know exist



Surveys

Twenty thousand sterling under the sea: Estimating the value of protecting deep-sea biodiversity

Niels Jobstvogt [a, f, g], Nick Hanley [f,g], Stephen Hynes [f,g], Jasper Kenter [h, i,f,g], Ursula Witte [f,g]

SCENARIO 1	Option A	Option B	Option C ("business as usual")
New medicinal products (potential for the discovery of new medicinal products from deep-sea organisms)	Unknown (potential for new medicinal products unknown)	High potential for new medicines (protect animals with potential for new medicinal products)	Unknown (potential for new medicinal products unknown)
Number of protected species (includes animals such as fish, starfish, corals, worms, lobsters, sponges & anemones)	1300 species (300 more than "business as usual")	1600 species (600 more than "business as usual")	1000 species (base level)
Additional costs (per household per year)	£ 5	£ 60	£ 0
Your choice for scenario 1 (please tick A, B or C)	☐	☐	☐

Data types for choice modelling

Responsible use of SP

- SP great for understanding trade-offs, e.g. willingness-to-pay
  - still consider hypothetical bias, lack of consequentiality, etc
- SP less suitable for forecasting
  - choices made in isolated setting
  - overall response often overstated
  - at very least, need correction approaches for scale and market shares

Data types for choice modelling

Best of both worlds

- Using SP and RP jointly is a very appealing option
- Typical application 'stacks' SP and RP data
  - parameters can be generic or specific to one preference data source

Data types for choice modelling

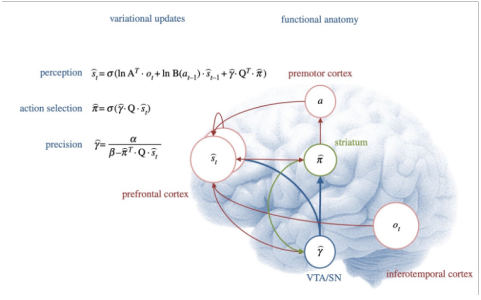
Additional material at end

- Non discrete choice data
- Terminology when working with SP
- Data collection and sampling of decision makers

Decision rules

Decision rules

- Mathematical representation of choice process
- Does not imply that people make choices according to the rules we use
- Simply *convenient* way of representing process
- Factors to consider
  - behavioural realism
  - tractability
  - properties of outputs



Decision rules

Two contrasting approaches

Compensatory models

- Changes in one attribute can be counteracted by changes in another attribute
- Theory: increases in cost can always be counteracted by reductions in time
- Practice: depends on size of change that is needed
- Key foundation of random utility maximisation (RUM)

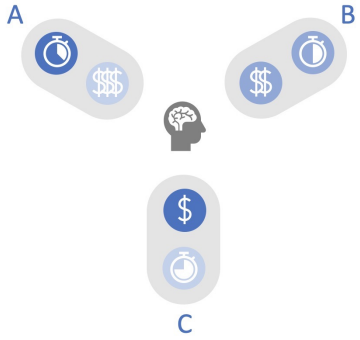
Non-compensatory and semi-non-compensatory models

- Some changes cannot be counteracted, or only partially
- Non-compensatory example: Elimination by aspects (EBA)
- Semi-non-compensatory example: Random Regret Minimisation (RRM)

Decision rules

Random Utility Maximisation (RUM)

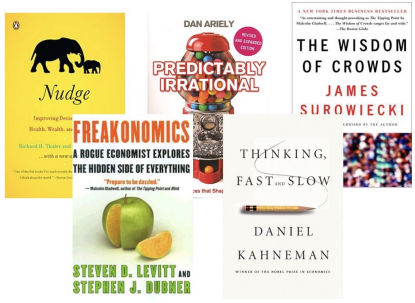
- Key paradigm since the early 1970s
- Grounded in micro-economic theory
- Inherent characteristics:
  - context independence
  - fully compensatory



Decision rules

RUM criticisms

- Many different RUM structures
- Criticised by behavioural economists and mathematical psychologists



Decision rules

Elimination by aspects (EBA): overview

- Gradual elimination of alternatives

Attribute	Drug A	Drug B	Drug C	Drug D	Rule
Cost	\$5 per day	\$10 per day	\$15 per day	\$20 per day	Cost $\leqslant$ \$15
Risk	1 in 500	1 in 1,000	1 in 2,000	1 in 5,000	Risks $\leqslant$1 in 1,000
Success	60%	70%	80%	90%	Sucess rate $\geqslant$ 75%

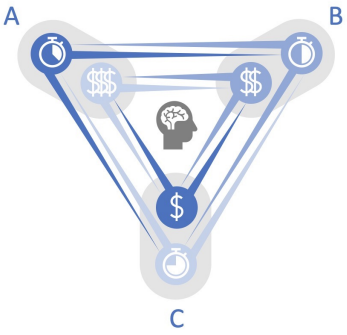
- Clearly not compensatory

Key reference: *Tversky, A. (1972), 'Elimination by aspects: A theory of choice', Psychological Review 79, 281-299.*

Decision rules

Random Regret Minimisation (RRM): overview

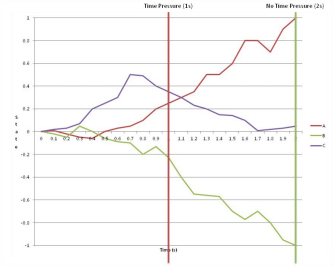
- Choice in non-risky situations
- Compare alternatives at level of attributes
- Minimise anticipated regret from inferior performance
- Inherent characteristics:
  - context dependence
  - semi non-compensatory
  - can capture compromise effects

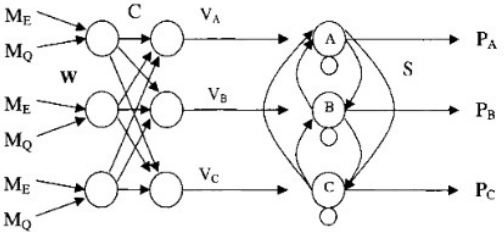


Key reference: *Chorus, C.G. (2012), 'Random Regret-based Discrete Choice Modeling, A Tutorial', Springer.*

Decision rules

Decision field theory (DFT): overview





Key reference: *Busemeyer, J. R. and Townsend, J. T. (1992). Fundamental derivations from decision field theory. Mathematical Social Sciences, 23(3):255-282.*

Implementation: *Hancock, T.O., Hess, S. & Choudhury, C.F. (2018), Decision field theory: improvements to current methodology and comparisons with standard choice modelling techniques, Transportation Research Part B, 107, pp. 18-40.*

Conclusions on decision rules

- Benefits:
  - behavioural realism
  - additional insights
- Losses:
  - no link to micro-economic theory
  - computational tractability
  - additional data requirements
- Question: do the benefits outweigh the losses?

Key reference: *Hess, S., Daly, A.J. & Batley, R. (2018), Revisiting consistency with random utility maximisation: theory and implications for practical work, Theory and Decision, 84(2), pp. 181-204.*

Appropriate uses of choice models

Appropriate uses of choice models

Much depends on the data

- Not every type of choice can or should be modelled
- Be careful also in model application/forecasting
  - is forecast scenario compatible estimation data?
  - how far outside of data range can we forecast?
- Recognise difference between *within* people and *across* people trends
  - can cross-sectional data tell us anything about intra-person differences?

Appropriate uses of choice models

The difference between why and how

- Models used to understand role of different attributes
- But causality itself is often not clear
  - does somebody not choose an expensive product due to cost, or are there other reasons?
  - reverse also happens, e.g. choosing expensive products due to perceived higher quality
- A choice model cannot tell us *how* a decision was made
  - we do not see into people's brains
  - any model is an approximation
  - all we can say is whether one type of decision rule works better than another one

Summary

Summary

Key points from this class

- Care required with type of data, with big differences between revealed and stated preference data
- Analyst needs to select an appropriate decision rule
- Major trade-offs between realism and applicability
- Choice models seek to explain why a specific alternative is chosen, not how that choice is made

Summary

Suggested reading

- Train, K.E. (2009), Discrete Choice Methods with Simulation, Cambridge University Press, free online access `https://eml.berkeley.edu/books/choice2.html`
  - Chapter 1, Section 1.2
  - Chapter 2, page 11-24
  - Chapter 7, Sections 7.2, intro to 7.3, 7.4 up until the equations, 7.5


Questions?
Apollo
www.ApolloChoiceModelling.com
The most flexible choice modelling software (up to a probability)

Additional material: RP *vs* SP

Additional material: RP *vs* SP

Not all surveys ask respondents to make choices!

- Ranking data
- Rating data
- Best-worst data
- Forced choice data
- Discrete continuous data

Additional material: RP *vs* SP

Ranking data

- Ask for full ranking
- Burdensome
- Middle rankings difficult
- Now often replaced by best-worst data

				
access time (minutes)	0	5	15	60
in vehicle time (minutes)	180	240	120	30
cost (£)	30	10	65	95
Preference ranking	3	4	2	1

Additional material: RP *vs* SP

Rating data

- Ask for rating for each alternative
- Gives further insights
- But may differ greatly across respondents
- Now used mainly for attitudes and perceptions

				
access time (minutes)	0	5	15	60
in vehicle time (minutes)	180	240	120	30
cost (£)	30	10	65	95
Rating (0-100)	50	20	85	90

Additional material: RP *vs* SP

Likert scale data for attitudes

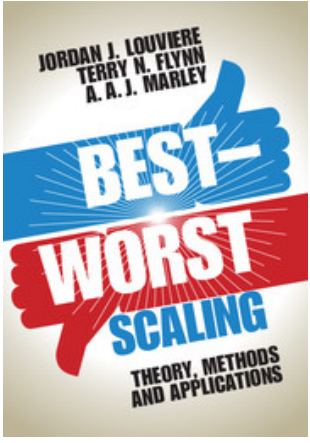
We would finally like you to indicate your level of agreement with the following statements. For each question, please provide your level of agreement on 5 point scale going from strongly disagree to strongly agree.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I am deeply concerned about COVID-19	○	○	○	○	○
I believe the measures put in place by the government to restrict transmission need to be strengthened	○	○	○	○	○
I believe the measures put in place by the government to restrict transmission should be relaxed	○	○	○	○	○
I believe that the risks of vaccination outweigh the benefits	○	○	○	○	○
There are significant risks in rapidly developing a vaccine for COVID-19	○	○	○	○	○
I am concerned about the impact of COVID-19 restrictions on my personal freedoms	○	○	○	○	○

Additional material: RP *vs* SP

Best-worst data

- Range of different formats
- Especially popular in marketing and health



Additional material: RP *vs* SP

Best-worst case 1: object case

Most important	Attribute	Least important
	Access time	✓
	In vehicle time	
✓	Travel cost	

- Popular for selecting attributes for SC
- No information on numeraire of the attributes or ranges of levels
- Link between perceived importance and impact might be weak

Most		Least
	Pesticides used on crops	
	Hormones given to livestock	
	Irradiation of foods	✓
	Excess salt, fat cholesterol	
✓	Antibiotics given to livestock	

Please consider the food safety issues in the table above and tick which concerns you most and which concerns you least.

Source: Flynn, 2014

Empirical comparison: *Song, F., Hess, S. & Dekker, T. (2021), A joint model for stated choice and best worst scaling data using latent attribute importance: application to high speed rail, Transportmetrica A, 17(4).*

Additional material: RP *vs* SP

Additional material

Best-worst case 2: profile case

- Ask for best and worst feature of alternative
- Can estimate utilities for each attribute level
- Simpler than SC, good for vulnerable respondents
- Can struggle when combining desirable with undesirable attributes

Best feature		Worst feature
	Access time: **15 mins**	✓
✓	In vehicle time: **120 mins**	
	Travel cost: **£65**	

Best		Worst
	Some problems walking about	
✓	No problems with self-care	
	Some problems with performing usual activities	
	No pain or discomfort	
	Moderately anxious or depressed	✓

Imagine you were living in the health state described above. Tick which aspect of this would be best to live with and which would be worst to live with.

Source: Flynn, 2014

Discussion paper: *V. Soekhai, B. Donkers, B. Levitan & E.W. de Bekker-Grob (2021), 'Case 2 best-worst scaling: For good or for bad but not for both', Journal of Choice Modelling, 41, 100325.*

Additional material: RP *vs* SP

Additional material

Best-worst case 3: multi-profile case

				
access time (minutes)	0	5	15	60
in vehicle time (minutes)	180	240	120	30
cost (£)	30	10	65	95
Most preferred				✓
Least preferred		✓		

	Phone 1	Phone 2	Phone 3	Phone 4
Phone Style	Clam or flip phone	 Candy Bar or straight phone	 Swivel flip	 PDA phone with touch screen input
Handset Brand	A	B	C	D
Price	$49.00	$199.00	$249.00	$129.00
Built-in Camera	No camera	5 megapixel camera	2 megapixel camera	3 megapixel camera
Wireless Connectivity	No Bluetooth or WiFi connectivity	Bluetooth and WiFi connectivity	WiFi connectivity	Bluetooth connectivity
Video Capability	No video recording	Video recording (up to 1 hour)	Video recording (more than 1 hour)	Video recording (up to 15 minutes)
Internet Capability	Internet Access	Internet Access	No Internet access	No Internet access
Music Capability	No music capability	MP3 Music Player only	FM Radio only	MP3 Music Player and FM Radio
Handset Memory	64 MB built-in memory	2 GB built-in memory	512 MB built-in memory	4 GB built-in memory

Source: Flynn, 2014

Key reference: *Marley, A.A.J., Louviere, J.J. & Flynn, T.N. (2015), 'Best-Worst Scaling', Cambr. Univ. Press*

Empirical comparison: *Giergiczny, M., Dekker, T., Hess, S. & Chintakayala, P. (2017), 'Testing the stability of utility parameters in repeated best, repeated best-worst and one-off best-worst studies', European Journal of Transport and Infrastructure Research, 17(4), pp. 457-476.*

Additional material: RP *vs* SP

Best-worst case 3 *vs* stated choice

How do we analyse BW?

- Option 1: sequence of two choices, with analyst-assumed order
- Option 2: choice between all possible best-worst outcomes

Advantages and disadvantages

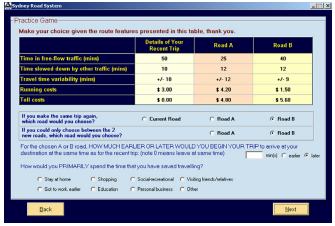
- Possibly smaller standard errors
- But is process to choose worst consistent with process to choose best?
- Is B-W useful when interest is in predicting first preferences?
- Barely used in some fields, potentially useful in others

Additional material: RP *vs* SP

Best-best and forced choice

access time (minutes)	0	5	15	60
in vehicle time (minutes)	180	240	120	30
cost (£)	30	10	65	95
Choice				✓

access time (minutes)	0	5	15
in vehicle time (minutes)	180	240	120
cost (£)	30	10	65
Choice			✓



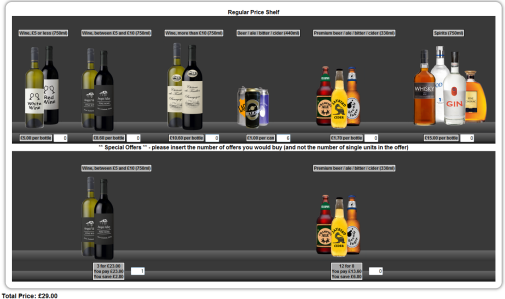
Source: Hess, S., Rose, J.M. & Hensher, D.A. (2008)

Empirical comparison: *Huls, S. P.I., Lancsar, E., Donkers, B. & Ride, J. (2022), 'Two for the price of one: If moving beyond traditional single-best discrete choice experiments, should we use best-worst, best-best or ranking for preference elicitation?' Health Economics, 31(12), 2630-2647*

Additional material: RP *vs* SP

Multiple discrete-continuous and discrete-discrete data

DAY	Sunday						
	Start		End		What were you doing (mode)	Where	
N°	Hour	Min	Hour	Min		Street 1	Street 2
1	10	0	11	20	Wake up, breakfast	Michimalongo 15	NA
2	11	20	14	0	Tidy up at home	Michimalongo 15	NA
3	14	0	14	5	Going to the shop (walk)	NA	NA
4	14	5	14	10	In the shop	Michimalongo central	NA
5	14	10	14	15	Going back home (walk)	NA	NA
6	14	15	15	30	Lunch	Michimalongo 15	NA
7	15	30	15	40	Going to a friend's home (walk)	NA	NA
8	15	40	15	50	At a friend's home	verbas buenas alto	NA
9	15	50	16	0	Going back home (walk)	NA	NA
10	16	0	19	0	Stay at home	Michimalongo 15	NA
11	19	0	19	30	Going to the doctor (walk)	NA	NA
12	19	30	21	30	Staying at the doctor	Salas	O'Higgins
13	21	30	22	0	Going back home (walk)	NA	NA
14	22	0	0	0	Stay at home, sleep	Michimalongo 15	NA

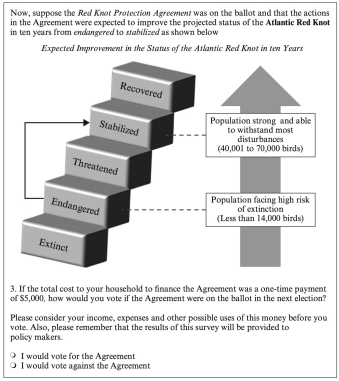


Additional material: RP *vs* SP

Additional material

Contingent valuation (CV), willingness to pay

- Open ended question or dichotomous choice, possibly double bounded
- Criticised as being too direct
- Framing possibly substantially influences answer (especially in terms of *bid levels*)
- CV largely discredited in some cases



Now, suppose the *Red Knot Protection Agreement* was on the ballot and that the actions in the Agreement were expected to improve the projected status of the **Atlantic Red Knot** in ten years from *endangered* to *stabilized* as shown below

Expected Improvement in the Status of the Atlantic Red Knot in ten Years

3. If the total cost to your household to finance the Agreement was a one-time payment of $5,000, how would you vote if the Agreement were on the ballot in the next election?

Please consider your income, expenses and other possible uses of this money before you vote. Also, please remember that the results of this survey will be provided to policy makers.

- ❍ I would vote for the Agreement
- ❍ I would vote against the Agreement

George Parsons and Kelley Myers, "Fat tails and truncated bids in contingent valuation: an application to an endangered shorebird species," Ecological Economics, Vol. 129, pp. 210–19, copyright 2016,

Additional material: RP *vs* SP

"Stated Choice" or "Discrete Choice Experiment"?

- In many fields, SC is now referred to as Discrete Choice Experiment (DCE)
- DCE is not necessarily a helpful term
  - gives idea that stated choices can only be discrete
- Worrying number of people think DCE=choice modelling
  - some authors say *"we conducted a DCE"* and think of both data and analysis
  - ignores that data and models are two different things
  - ignores existence of RP
  - criticism of SP as data source attached to choice modelling as analytic technique

Additional material: RP *vs* SP

Is SC "conjoint", or "choice based conjoint"?

- Terms used widely in consulting and marketing
- In conjoint, respondents evaluate products independently of each other
  - suitable if competition does not need to be considered
- In SC, respondents simultaneously consider multiple products

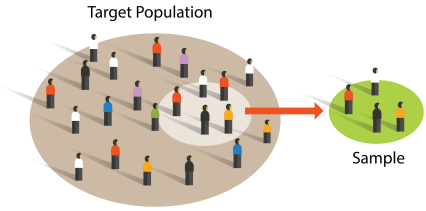
Louviere, J.J., Flynn, T.N. & Carson, R.T. (2010), Discrete Choice Experiments Are Not Conjoint Analysis, Journal of Choice Modelling, Volume 3, Issue 3, pp. 57-72.
Also: https://www.greenbook.org/marketing-research/conjoint-or-dcm-choice-models-overview-34754

Additional material: sampling

Additional material: sampling

Introduction to sampling

- Target population depends on aims
  - general population
  - subset of the population, for example if only people with specific health conditions are within scope
- Study uses subsample
- Sampling can impact results

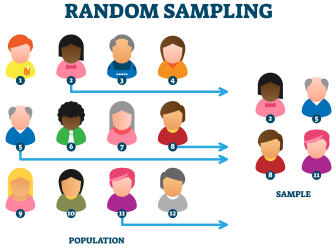


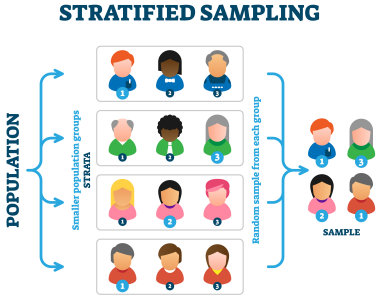
Key reference: *Ben-Akiva, M. & Lerman, S. (1985), 'Discrete Choice Analysis', chapter 8, MIT press.*

Additional material: sampling

Additional material

Random sampling *vs* stratified random

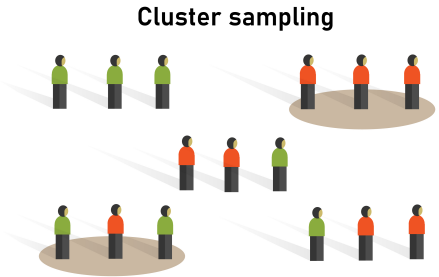


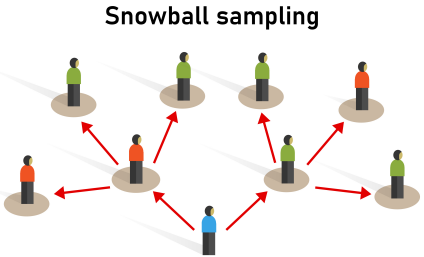


- Random sampling may have too few people from small subgroups
- Stratified sampling addresses this, but needs to be dealt with during modelling

Additional material: sampling

Cluster and snowball sampling





- Likely correlations in preferences

Additional material: sampling

Endogenous sampling

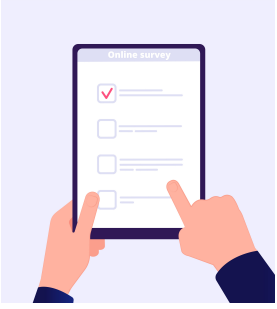
- Sampling related to preferences
  - fully choice based
  - intercept based
- Corrections needed, and approaches for this are difficult

Additional material: sampling

Additional material

Online panels and unrepresentative preferences

- Growing popularity of online panels
  - unrepresentative preferences
  - risk of bots



Additional material: sampling

Key take-aways on sampling

- Sampling approach matters
- Correction approaches exist
- Key issue is unobserved bias