Heterogeneity - an overview

Key concepts
& study plan

Experimental
design

Data collection
& processing

**Model specification
& estimation**

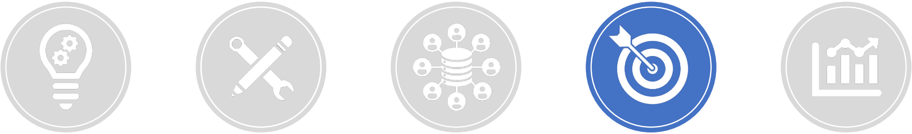
Interpretation
& application

Heterogeneity - an overview

What is heterogeneity?

- Heterogeneity is another word for variability
- In a choice modelling context, this relates to:
  - differences across people
  - differences across choices for same person
    - e.g. different settings, different points in time
- Main focus is in differences in sensitivities (e.g. cost sensitivity)
- Heterogeneity in preferences can lead to differences in choice outcomes

Why does heterogeneity matter?



Key concepts
& study plan

Experimental
design

Data collection
& processing

**Model specification
& estimation**

Interpretation
& application

Why does heterogeneity matter?

Mathematical fit

- Incorporating heterogeneity in our models increases flexibility
- Ability to better explain choices
- Models with heterogeneity invariably obtain better model fit

Why does heterogeneity matter?

Behavioural insights

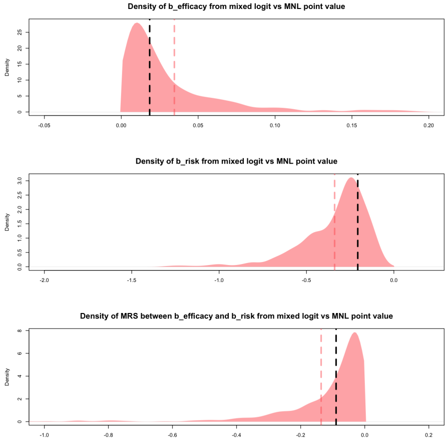
- Provides insights into why some people make specific choices
  - e.g. underlying health conditions reduce likelihood of certain treatment options
- Can ensure more robust appraisal of cost vs benefits
  - e.g. if we know how relative importance of waiting time and risk varies across patients
- Can help shape a more efficient provision of services/goods
  - e.g. make a variety of services available, covering widest range of preferences
- Can help shape strategies to nudge behaviour
  - e.g. identify drivers of vaccine hesitancy

Why does heterogeneity matter?

What if I'm only interested in the mean?

Not all point values are equal!

Aggregate preferences from a model allowing for heterogeneity may be different (and less biased) than those from a model assuming homogeneity



Why does heterogeneity matter?

Why do we get bias?

- Not accounting for heterogeneity increases amount of noise in model

$$
\begin{aligned}
U_{jnt} &= \sum_{k=1}^{K} \beta_{n,k}\, x_{jnt,k} + \varepsilon_{jnt} \\
&= \sum_{k=1}^{K} \beta_{k}\, x_{jnt,k} + \varepsilon_{jnt} + \sum_{k=1}^{K} \left(\beta_{n,k} - \beta_{k}\right) x_{jnt,k}
\end{aligned}
$$

- Unobserved part of utility now correlated with deterministic part
  $\longrightarrow$ breaks core assumption of additive in errors random utility
  $\longrightarrow$ potential bias
- Even if not *interested* in heterogeneity, will have less bias by accounting for it

Deterministic *vs* random heterogeneity

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Deterministic *vs* random heterogeneity

Deterministic heterogeneity: link to observed information

Option 1: discrete segmentations, with separate models

- e.g. male *vs* female, business *vs* leisure
- same as a joint model with segment-specific parameters
- assumes that differences exist in sensitivities to all attributes

Option 2: differences only for some attribute-covariate pairs

- interaction with categorical variables
  - e.g. interaction with gender, implying different sensitivities for men and women
- interactions with continuous covariates
  - e.g. continuous interaction with income, implying different sensitivity for each possible income

Deterministic *vs* random heterogeneity

Need for random taste heterogeneity

- Some taste heterogeneity cannot be explained deterministically
- Data limitations
  - we do not know everything about individuals in our data
  - constraints on behaviour, unobserved socio-demographics, etc
- Idiosyncratic reasons
  - two apparently *identical* individuals may have different sensitivities
- Solution is to allow for *random* heterogeneity
- Preferences not random, simply unobserved
  - use random heterogeneity to deal with this

Deterministic *vs* random heterogeneity

Mixture models

- Aim:
  - accommodate random taste heterogeneity
- Method:
  - allow choice probabilities to vary as function of (unobserved) distribution of sensitivities

Deterministic *vs* random heterogeneity

Key differences across specifications

- Model specifications without any heterogeneity
  - same probability for all individuals when faced with same choice scenario
- Model specifications with deterministic heterogeneity
  - probabilities vary across individuals
  - we **know** where on that distribution each person is located
- Model specifications with random heterogeneity
  - probabilities also vary across individuals
  - we **do not know** where on that distribution each person is located
- In models combining deterministic with random heterogeneity, we can be more certain about where on the distribution a person is

Deterministic *vs* random heterogeneity

Two broad categories of models

Finite mixtures

- Allow for a limited number of possible values for sensitivities
- Two different implementations:
  - Discrete mixtures: heterogeneity in individual parameters
  - Latent class: finite set of combinations of values for different parameters

Continuous mixtures

- Use continuous statistical distributions to capture heterogeneity
- Most flexible type of random utility model in theory
- Known as Mixed Logit, or Random Parameters Logit

Deterministic *vs* random heterogeneity

Basic idea

- Same underlying idea for finite and continuous mixtures
- Analyst specifies an underlying model, typically (but not necessarily) MNL
  - in technical terms often referred to as the *kernel* of the mixture model
- If sensitivities of an individual were known, would have a probability for the choices as in MNL (or other kernel)
- But sensitivities are not known

Deterministic *vs* random heterogeneity

A VERY simple example

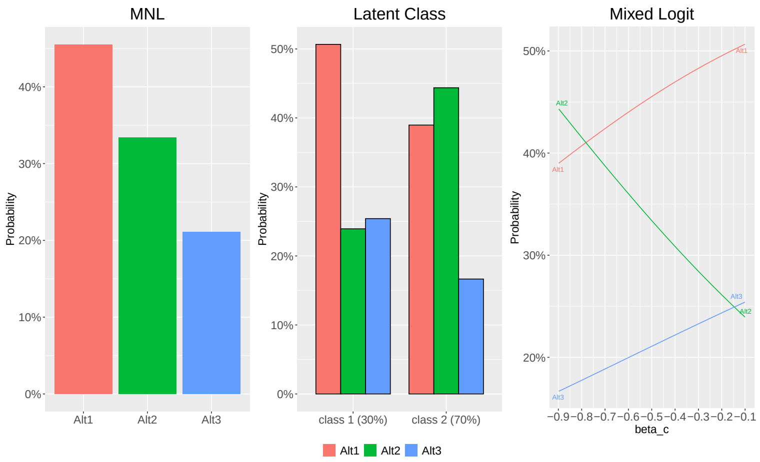
- Random cost (TC) coefficient random, fixed constants and sensitivities to free flow time (FFT), slowed down time (SDT) and tolls (TOLL)

	δ_1	δ_2	FFT	SDT	TC	TOLL
Alt 1	1	0	30	7	2.3	5
Alt 2	0	1	36	8	1.2	6
Alt 3	0	0	15	6	2.5	7
β	0.2	0.1	-0.03	-0.08	β_{TC}	-0.5

- Three cases:
  - MNL: $\beta_{TC} = -0.5$
  - Latent class: two classes, with weights of 30% and 70%, and values of β_{TC} of -0.1 and -0.9
  - Mixed Logit: β_{TC} is distributed uniformly on $[-0.9, -0.1]$

Deterministic *vs* random heterogeneity

Choice probabilities vary as a function of β_{TC}



Where is the heterogeneity?

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Where is the heterogeneity?

Misattributing deterministic heterogeneity

- Model estimation will use whatever flexibility we give it to try and improve fit
- Allowing heterogeneity in only some attributes risks lower fit and misattribution

Number of individuals	500		500		500		500	
Number of modelled outcomes	5000		5000		5000		5000	
Estimated parameters	2		3		3		4	
LL(final)	-2167.405		-2083.096		-2061.137		-2060.733	
Adj.Rho-square (0)	0.374		0.3981		0.4044		0.4042	
AIC	4338.81		4172.19		4128.27		4129.47	
BIC	4351.84		4191.74		4147.82		4155.54	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficacy	0.0199	11.78	0.0279	16.64	0.0202	14.1	0.0212	11.78
risk	-0.3651	-25.18	-0.3647	-26.37	-0.2958	-23.36	-0.3025	-20.12
efficacy interaction for men			-0.021	-12.79			-0.0029	-0.99
risk interaction for men					-0.2033	-13.33	-0.1816	-6.57

- Ignoring heterogeneity in β_k can lead to bias in β_l
  - Bad idea to keep the cost coefficient fixed!

Where is the heterogeneity?

Misattributing random heterogeneity

- As with deterministic heterogeneity, model will use whatever flexibility we give it
- Allowing for heterogeneity in only some attributes risks misattribution

Number of individuals	500		500		500		500	
Number of modelled outcomes	5000		5000		5000		5000	
Estimated parameters	3		3		4		5	
LL(final)	-2160.457		-2150.033		-2150.017		-2149.356	
Adj.Rho-square (0)	0.3758		0.3788		0.3785		0.3784	
AIC	4326.91		4306.07		4308.03		4308.71	
BIC	4346.47		4325.62		4334.1		4341.3	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficacy (mean for lognormal)	-3.9673	-40.54	-3.8255	-48.46	-3.8257	-48.52	-3.7959	-46.23
efficacy (sd for lognormal)	-0.3579	-6.19	0	NA	-0.0316	-0.83	-0.0928	-0.97
risk (mean for lognormal)	-0.9883	-23.84	-0.9401	-22.92	-0.94	-22.92	-0.9301	-21.87
risk (sd for lognormal)	0	NA	0.262	8.54	0.2614	8.55	0.3277	5.27
cholesky term	0	NA	0	NA	0	NA	0.158	1.29

Where is the heterogeneity?

Where is the heterogeneity?

taste heterogeneity

- differences in relative sensitivities to individual attributes
- differences in baseline preferences for different options

scale heterogeneity

- differences across individuals in amount of noise (from analyst's perspective)

process heterogeneity

- differences in how information is processed, and what decision rule is used

- In practice, difficult/impossible to disentangle, main focus is taste heterogeneity

Heterogeneity across people and across choices

Key concepts
& study plan

Experimental
design

Data collection
& processing

Model specification
& estimation

Interpretation
& application

Heterogeneity across people and across choices

Different possibilities

- Main scope for heterogeneity is across decision makers
- But also scope across choices for the same individual
  - choices at different points in time, or for different types of choices (e.g. long and short trips)
- With deterministic heterogeneity, easy to accommodate both at the same time
- With random heterogeneity, main interest is across people, but can also accommodate intra-individual heterogeneity

Key reference: *Hess, S. & Train, K.E. (2011), Recovery of inter- and intra-personal heterogeneity using mixed logit models, Transportation Research Part B, 45(7), pp. 973-990.*

Example with latent class: *Song, F., Hess, S. & Dekker, T. (2023), Uncovering the link between intra-individual heterogeneity and variety seeking: the case of new shared mobility, Transportation, 51, pp. 371-406.*

Implications for data

Key concepts
& study plan

Experimental
design

Data collection
& processing

Model specification
& estimation

Interpretation
& application

Implications for data

Data requirements

- Capturing heterogeneity places additional demands on data
- Need rich data, with trade-offs allowing us to observe different choice outcomes
- Heterogeneity in sensitivities may not lead to different outcomes
- Focus on categorical variables in some fields makes capturing heterogeneity harder
  - proliferation of parameters
  - identification issues