MODELS CAPTURING LATENT CONSTRUCTS

Aurélie Glerum

Michel Bierlaire

Seventh Workshop on Discrete Choice Models
Lausanne, 25th August 2011

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

Introduction

- Two case studies
- Context of research
- New data types
- Research objectives

Data to models

- Specification & analysis
- Validation & forecasting

Models to data

Conclusion

Introduction

Two case studies

Recently work on two case studies:

Case study 1: mode choice study in low-density areas

Case study 2: vehicle choice including electric vehicles

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

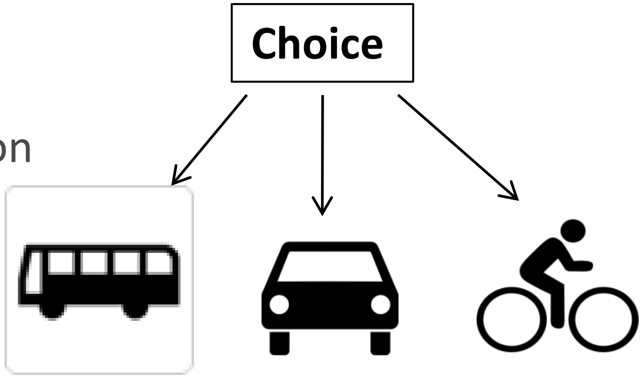
Introduction

Two case studies

Case study 1: mode choice study in low-density areas

Available data:

- Individuals trips (RP)
- Socio-economic information
- Mobility habits
- Opinions
- Perceptions



Work so far:

- Development of hybrid choice models (HCM)
- Focus on analysis of latent variables & integration into choice model (rather than forecasting)

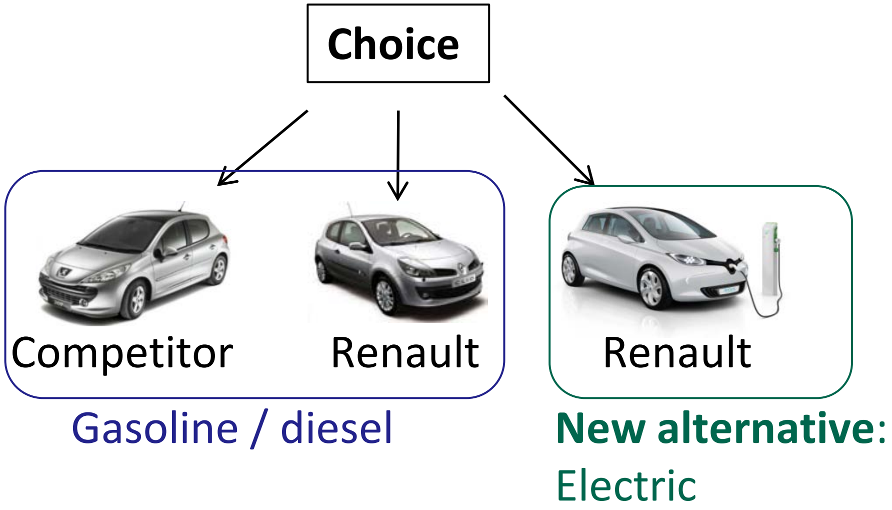
Introduction

Two case studies

Case study 2: vehicle choice including electric vehicles

Available data:

- Vehicle description (RP)
- Vehicle preferences (SP)
- Socio-economic information
- Mobility habits
- Opinions
- Perceptions



Work so far:

- Development of a logit model with multiple alternatives
- Focus on validation & forecasting (no integration of latent attitudes yet)

Introduction

Context of research

Context of research: recent progresses in DCM

- Focus on attitudes and perceptions
- Taken into account to model choice behavior

Motivation:

- Choice cannot only be explained by economic indicators (time, price, etc.)
- Important role of attitudes and perceptions in choice behavior

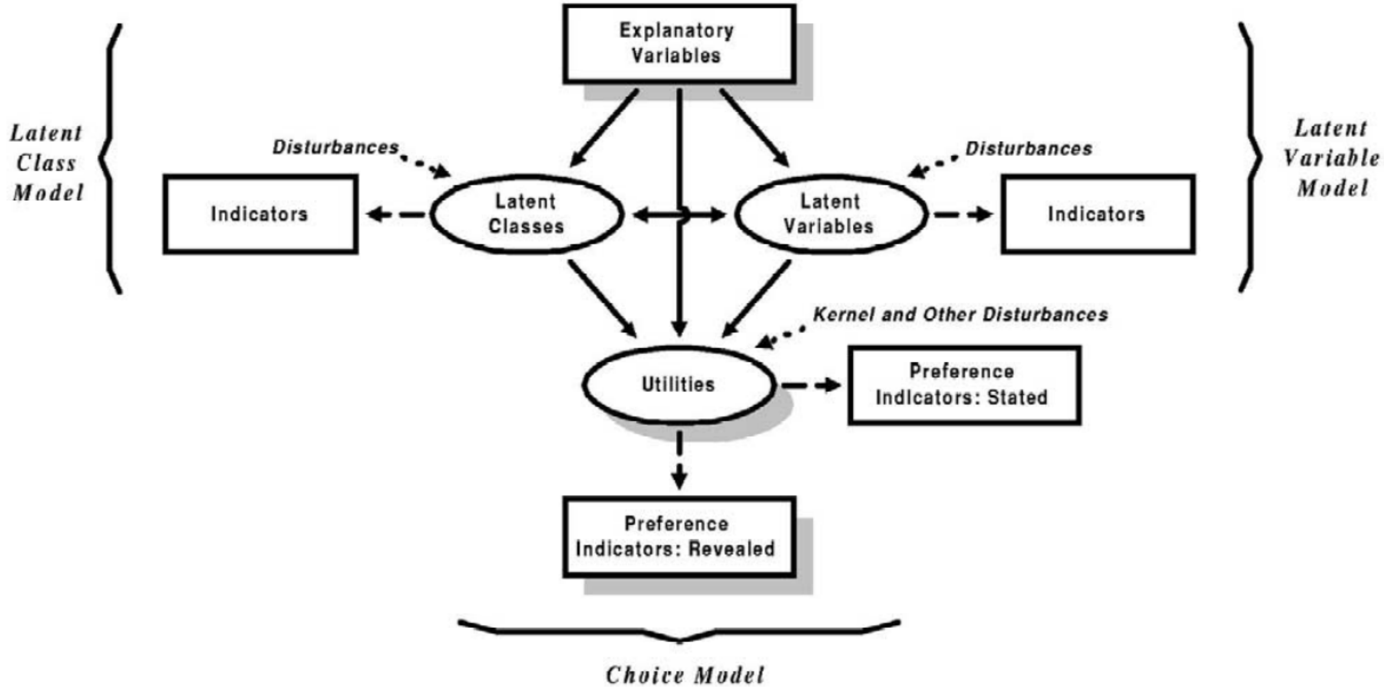
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Introduction

Context of research

DCM with **latent constructs** capturing **attitudes** and **perceptions**:
Hybrid choice model (HCM) (Walker, 2001; Ben-Akiva et al., 2002)



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Introduction

New data types

Issue:
- Latent aspects must be measured from real data

Recently: new types of data
- Data from survey with advanced designs developed by social scientists
- Data from new devices: smartphones, image analysis, eye-tracking devices, fMRI, etc.

Current drawback:
- Data not necessarily designed for choice models

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Objectives of this research:

Investigate potential issues:

Objective 1:

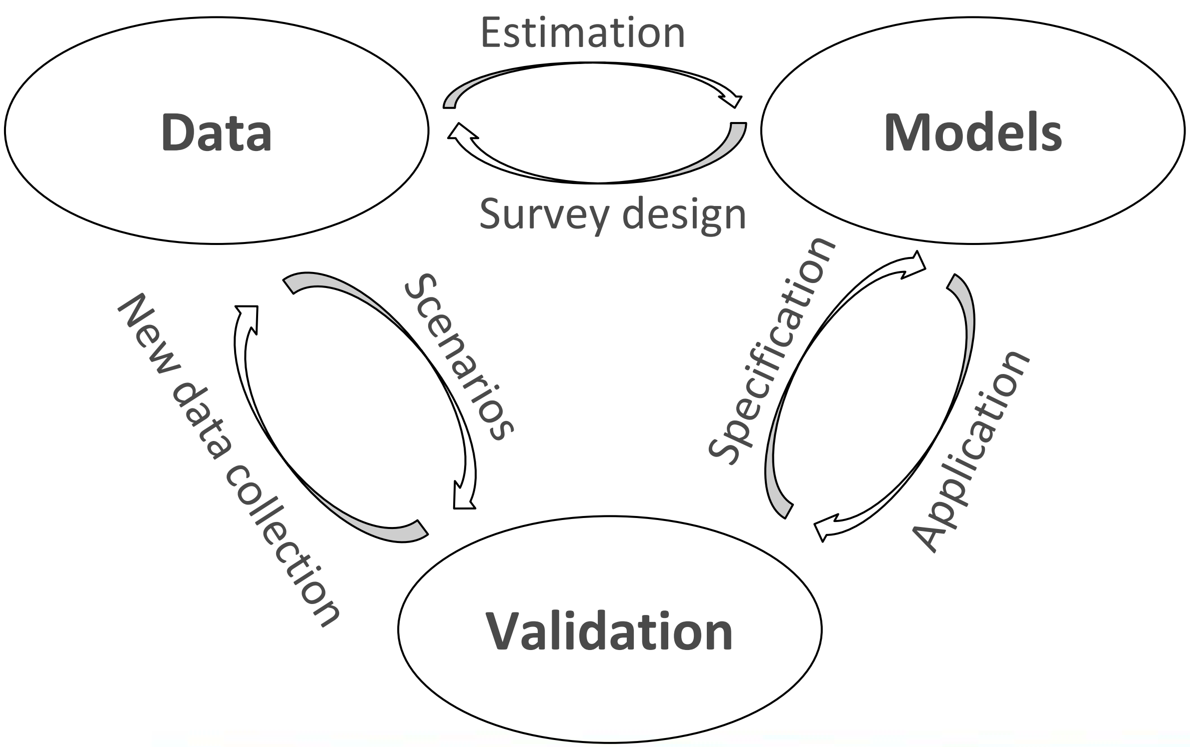
- Developments from **data to models**:
  1. Specification & analysis of HCM
  2. Validation & forecasting of HCM

Objective 2:

- Developments from **models to data**:
  1. Improvement of survey design

Introduction

Research objectives



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

Objective 1: data to models

1. Specification & analysis
2. Validation & forecasting

of HCM

Data to models

1. Specification & analysis

Issues in specification and estimation of HCM:

1. Relation between latent variable and its indicators

   ⟹ Work on measurement equations:

   1. Non-linear specifications of indicators
   2. Use of indicators from word questions

2. Estimation

   ⟹ Local optima, Bayesian inference, etc.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

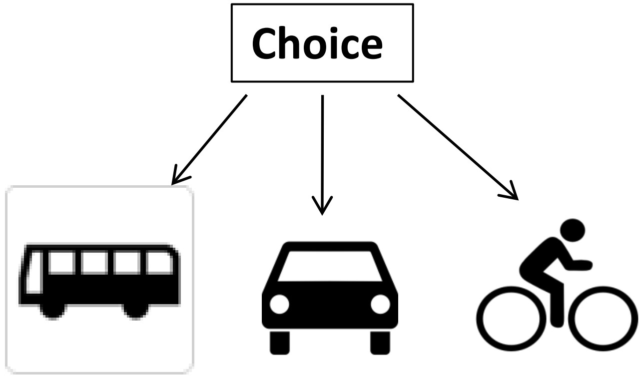
Data to models

1. Specification & analysis

1. Relation between latent variable and its indicators

Non-linear specifications of indicators in **measurement equations**

Example 1 from mode choice case study



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

1. Specification & analysis

Four themes in statements of opinion:

- Environment — *The price of gasoline should be increased in order to reduce traffic congestion and air pollution.*
- Mobility — *Taking the bus helps making a town more comfortable and welcoming.*
- Residential choice — *Accessibility and mobility conditions are important in the choice of an accommodation.*
- Lifestyle — *I always plan my activities a long time in advance.*

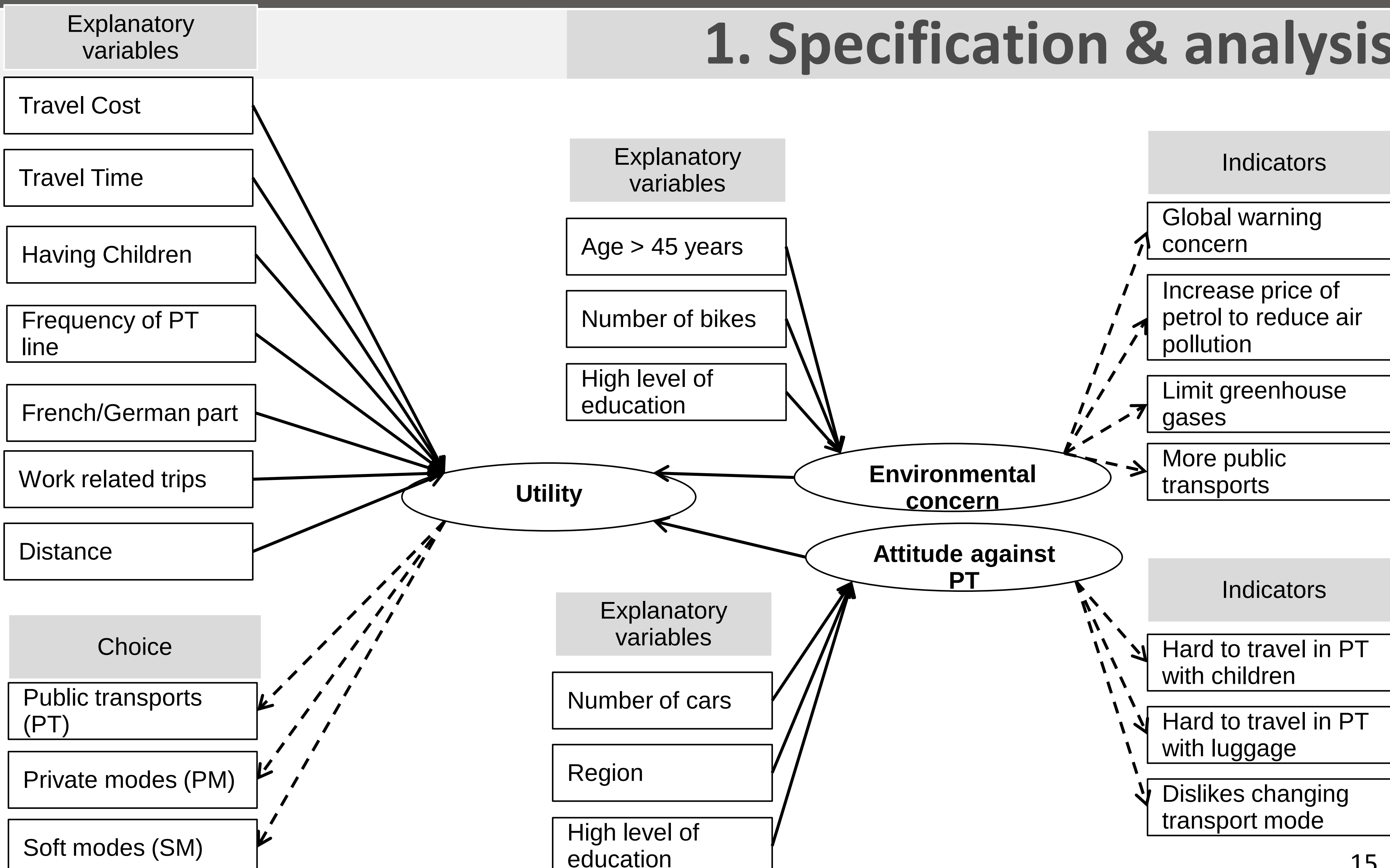
Respondents rate agreement on 5-point Likert scale:
Total disagreement (1) $\Longrightarrow$ Total agreement (5)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

1. Specification & analysis



Data to models

1. Specification & analysis

Measurement equation of indicators:

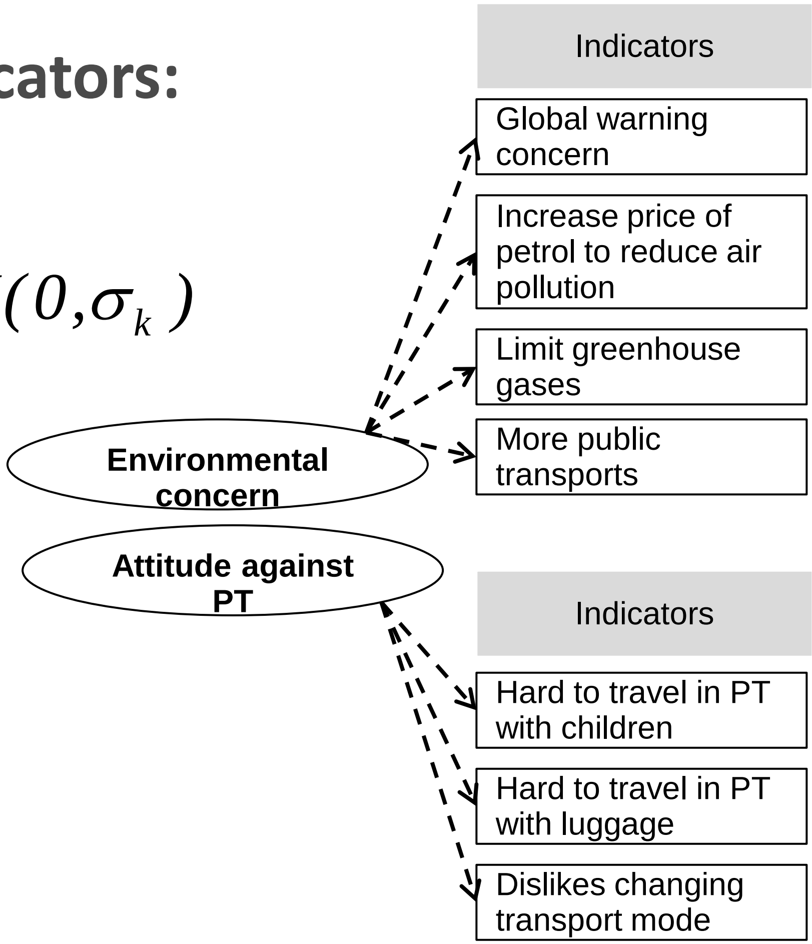
- Currently: linear

$$I_k = \alpha_k + \lambda_k A + \nu_k \text{, with } \nu_k \sim N(0, \sigma_k)$$

- Assumption might not always hold

$\Longrightarrow$ Non-linear specification

- Examples: agreeing-response bias (Schuman and Presser, 1981)



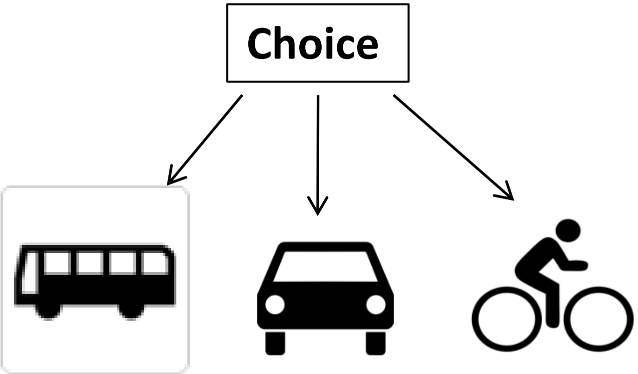
Data to models

1. Specification & analysis

1. Relation between latent variable and its indicators

Use of indicators from **word questions** in **measurement equations**

Example 2 from mode choice case study



Data to models

1. Specification & analysis

Adjective data for perception of transport modes

For each of the following transport modes, give three adjectives that describe them best according to you.

		Adjective 1	Adjective 2	Adjective 3
1	The car is:	**convenient**	**comfortable**	**expensive**
2	The train is:	**relaxing**	**punctual**	**restful**
3	The bus, the metro and the tram are:	**fast**	**frequent**	**cheap**
4	The post bus is:	**punctual**	**comfortable**	**cheap**
5	The bicycle is:	**stimulating**	**convenient**	**cheap**
6	The walk is:	**healthy**	**relaxing**	**independent**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

1. Specification & analysis

Data processing:

- Classification into themes:
    - Perception of cost
    - Perception of time
    - Difficulty of access
    - Flexibility
    - Comfort, etc.
- Attribution of scale from -2 to +2

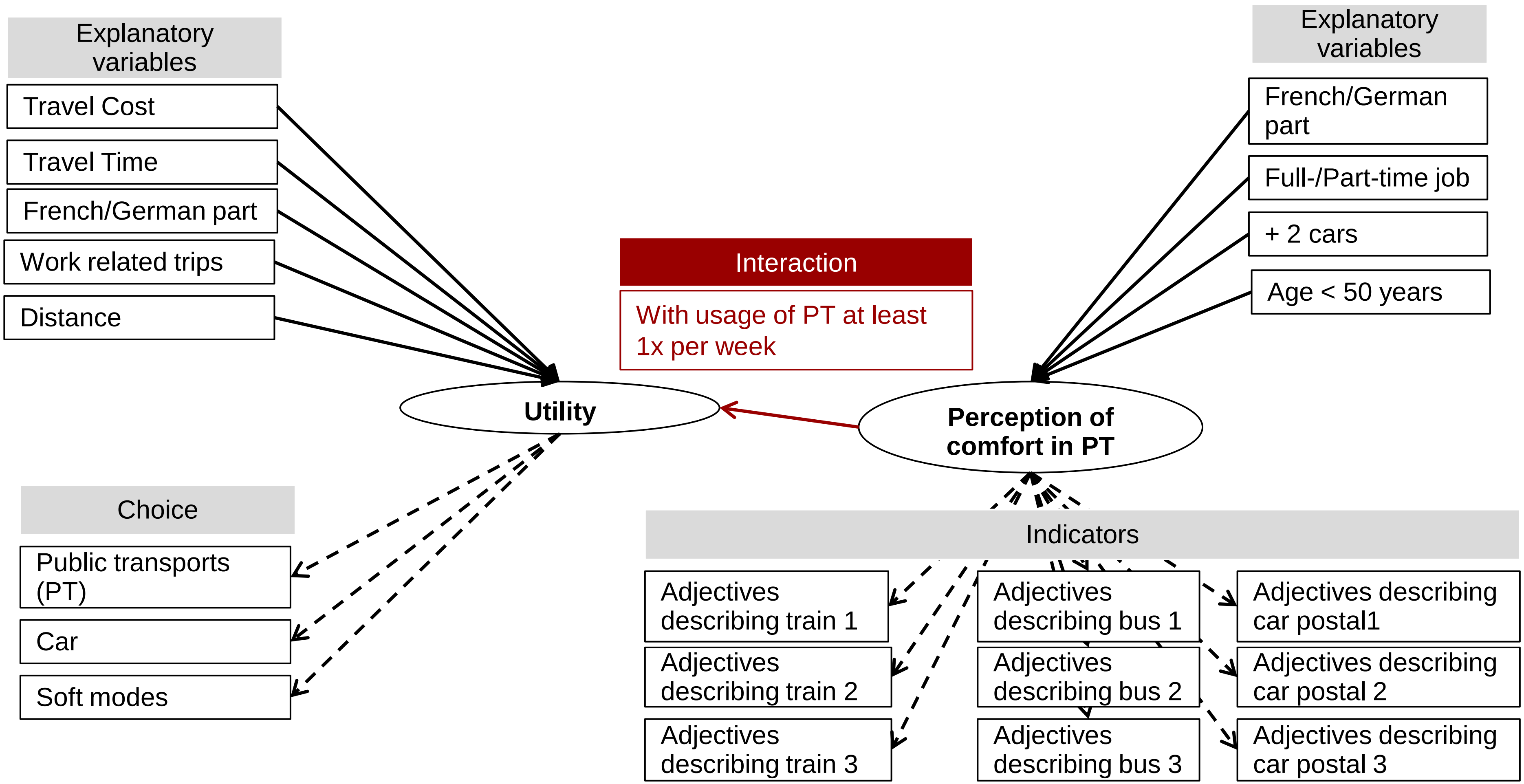
Comfort	Scale
hardly full	1
packed	-1
bumpy	-2
comfortable	1
hard	-1
irritating	-2
tiring	-1
unsuitable with bags	-1
uncomfortable	-1
bad air	-2

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

1. Specification & analysis



Data to models

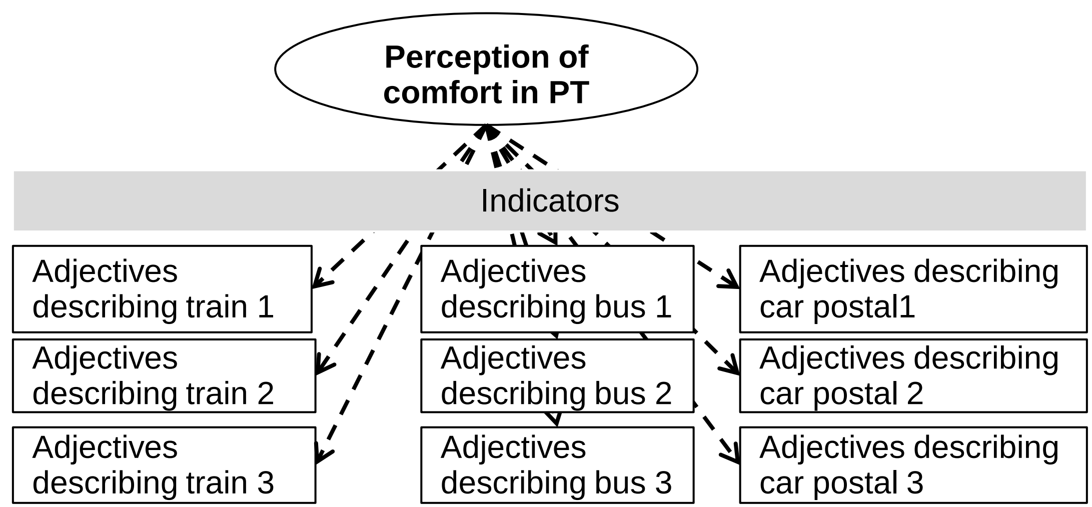
1. Specification & analysis

Measurement equations:

$$I_n = m(X_n^*;\alpha) + \upsilon_n$$

with $\upsilon_n \sim Logistic(0,1)$

$$I_n = \begin{cases} -2 \text{ if } -\infty < X_n^* \le \tau_1 \\ -1 \text{ if } \tau_1 < X_n^* \le \tau_2 \\ 0 \text{ if } \tau_2 < X_n^* \le \tau_3 \\ 1 \text{ if } \tau_3 < X_n^* \le \tau_4 \\ 2 \text{ if } \tau_4 < X_n^* \le +\infty \end{cases}$$



Data to models

1. Specification & analysis

Issue:

- Scale from -2 to +2 subjective to the modeler

Improvements:

- Order adjectives & obtain distribution with respect to each theme:

  $\Longrightarrow$ Text mining

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

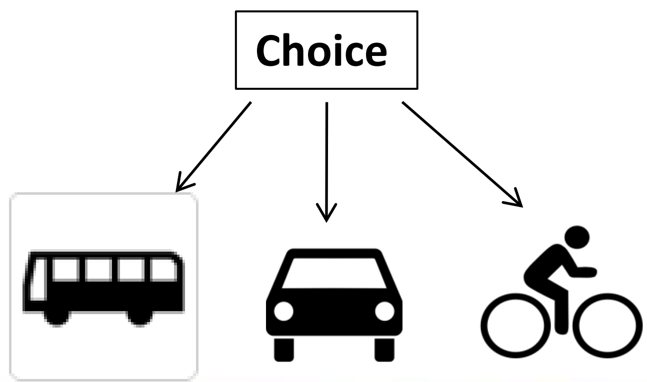
1. Specification & analysis

2. Issues related to estimation

Example:

- Difficulty of estimating models including several latent variables:
    - Bayesian inference using Gibbs sampling (Alvarez-Daziano and Bolduc, 2009)
    - Mode choice case study: only 2 latent variables

Example 3 from mode choice case study

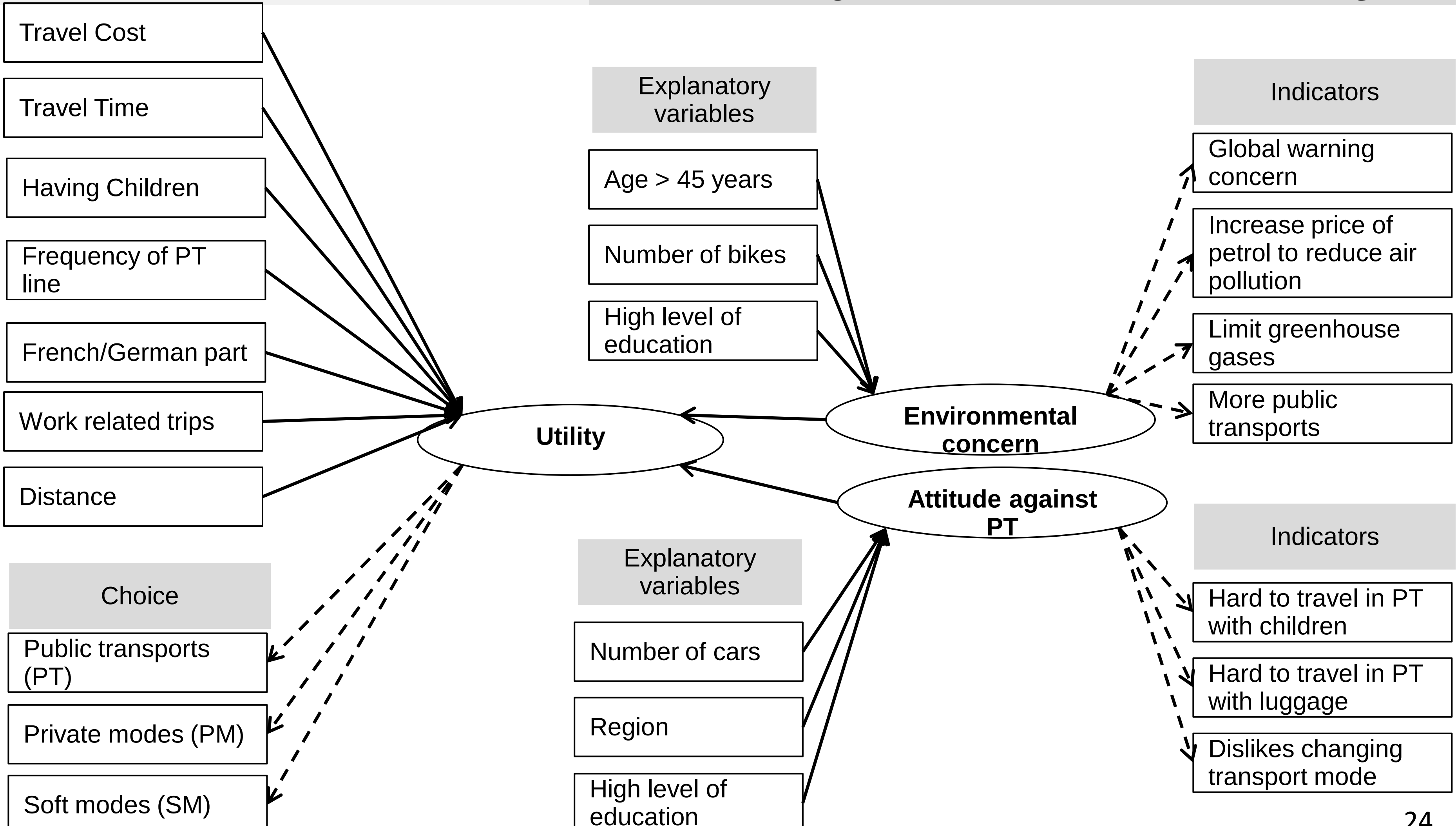


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

1. Specification & analysis



Data to models

Issues in validation & forecasting of HCM:

1. Analysis of demand indicators built on latent variables
2. Inclusion of aggregate market data for forecasting
3. Absence of market data for new alternatives (SP)

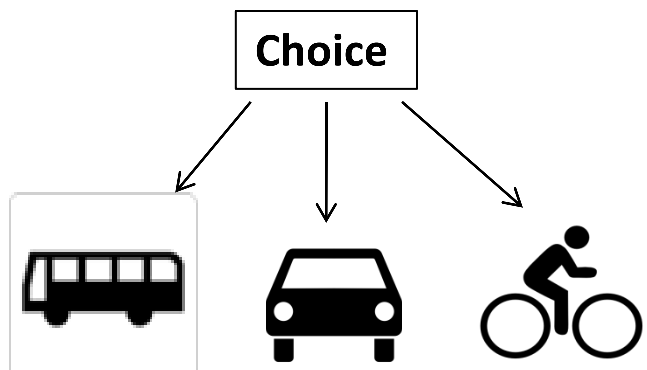
Data to models

1. Analysis of demand indicators built on latent variables

Computation of demand indicators depending on value of latent variable:

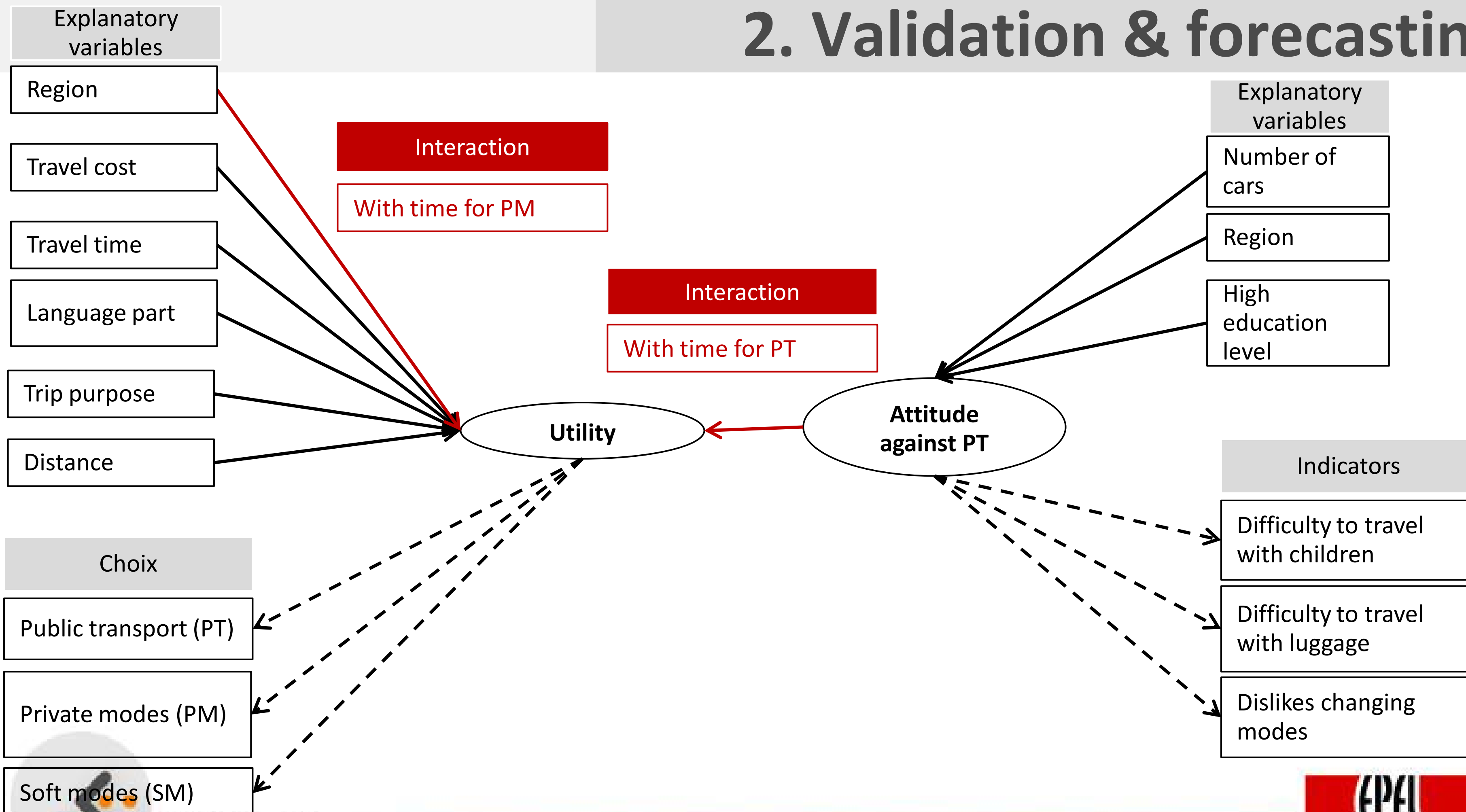
- Capture heterogeneity of value of time (VOT) in population (Abou-Zeid et al., 2010)

Example 4 from mode choice case study



Data to models

2. Validation & forecasting



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

2. Validation & forecasting

Value of time PT:

$$VOT_{PT,n} = \frac{\beta_{timePT} - \beta_{attPT} \cdot attPT_n}{\beta_{cost}}$$

Result:

- Individuals with more negative attitude against PT

  $\Longrightarrow$ Increase in TT will **decrease** probability to choose PT

  Individuals with more a positive attitude towards PT.

  $\Longrightarrow$ Increase in TT will **increase** probability to choose PT
- Impacts on VOT

Purpose:

- Investigate more indicators & impacts on latent variables

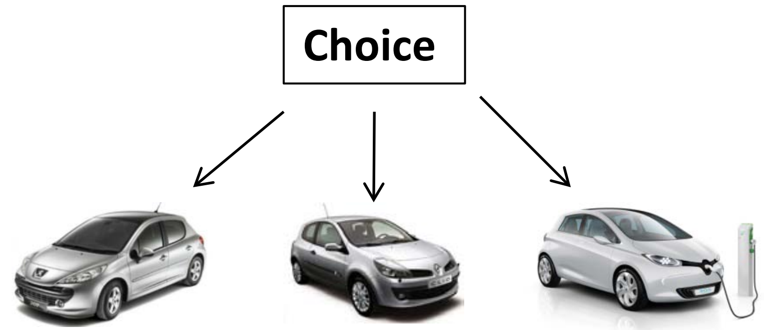
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

2. Inclusion of aggregate market data for forecasting

Inclusion of aggregate alternatives in SP survey to deal with missing information

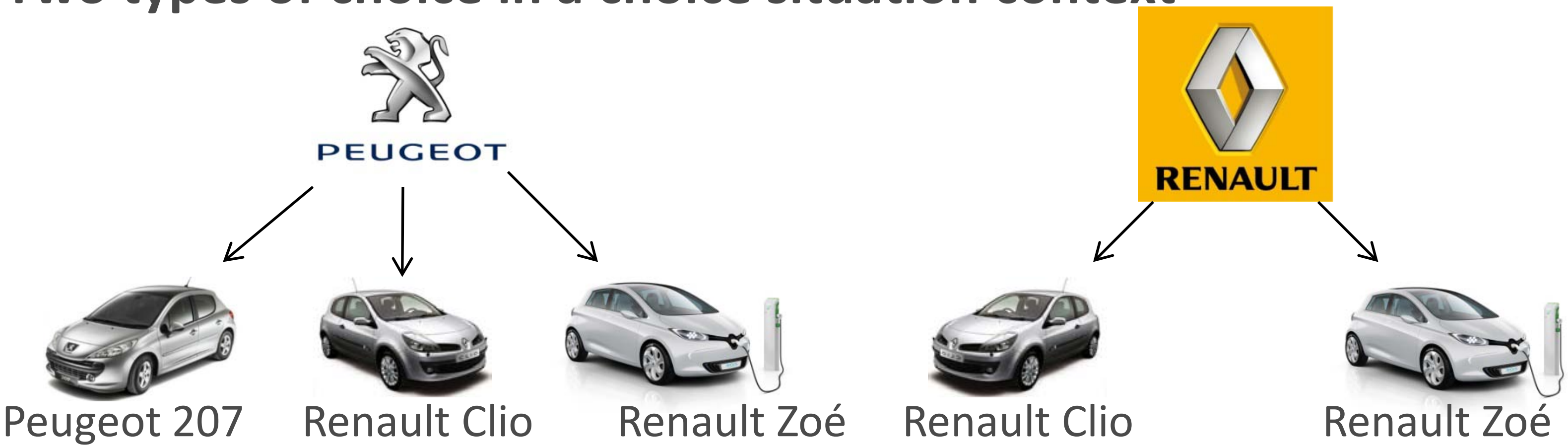
Example 5 from vehicle choice case study



Data to models

2. Validation & forecasting

Two types of choice in a choice situation context



Issue:

- Choice is supposed to represent all possible alternatives for decision maker
- Not the case for owners of Renault cars

Solution:

- Impute aggregate alternative of gasoline – competitors for these individuals

Data to models

2. Validation & forecasting

Aggregate alternative imputed for Competitors – Gasoline (CG):

$$V_{\mathrm{CG}} = \log \sum_{l \in L} \exp V_{ln}$$

$$V_{ln} = C_{\mathrm{CG}} + \sum_{s \in S_n} \beta_s \cdot x_s + \beta_{\mathrm{price}_{\mathrm{CG}}} \cdot \mathrm{price}_l + \beta_{\mathrm{UseCostGasoline}} \cdot \min(\mathrm{Cost100}_l, 12)$$

Create **aggregate alternative** from **prices** & **operating costs** of new cars on market
(matching segment of 2 other alternatives in choice situation)

TRANSP-OR

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data to models

2. Validation & forecasting

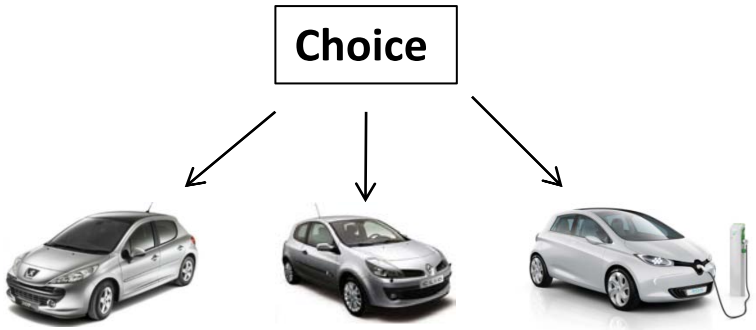
3. Absence of market data for new alternatives (SP)

Use:

- Market data of current alternatives
- SP survey data

To estimate possible share for new alternative

Example 6 from vehicle choice case study



Data to models

Evaluation of potential market share (MS) for EV

Acceptance rate EV in the questionnaire for CG owners (weighted)

Acceptance rate EV in the questionnaire for RG owners (weighted)

$$MS(RE) = \boxed{\%(\text{Choice RE} \mid \text{Owns CG})} \cdot \boxed{93.74\%} + \boxed{\%(\text{Choice RE} \mid \text{Owns RG})} \cdot \boxed{6.26\%}$$

Market share of competitors

Market share of Renault

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Software demo

Illustration: simulation viewer of new version of Biogeme (Bierlaire and Fetiarison, 2009)

Market share scenario with:

- Imputation of aggregate alternative
- Integration of market data to compute MS of EV

Variation of price of EV

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Models to data

Objective 2: models to data

Design of new surveys from learnings of models

Models to data

Improving design of questionnaires:

- Semi-open questions:
  - Lot of information from free report
  - Include more of such questions
- Further modeling developements:
  Semi-open → open questions

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusion

Presented **research agenda** based on expertise from two case studies

⟹ **2 main objectives:**

- **Data to models:**
  - Identify main issues in estimation of HCM
  - Improve specification (measurement equations)
  - Analysis of demand indicators related to latent variables
  - Include aggregate alternatives
  - Include market data

- **Models to data:**
  - Design new surveys for HCM based on learning from models

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Thanks

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE