

# Commuting mode choice with latent preference heterogeneity: a case study for employees of the EU Institutions in Luxembourg

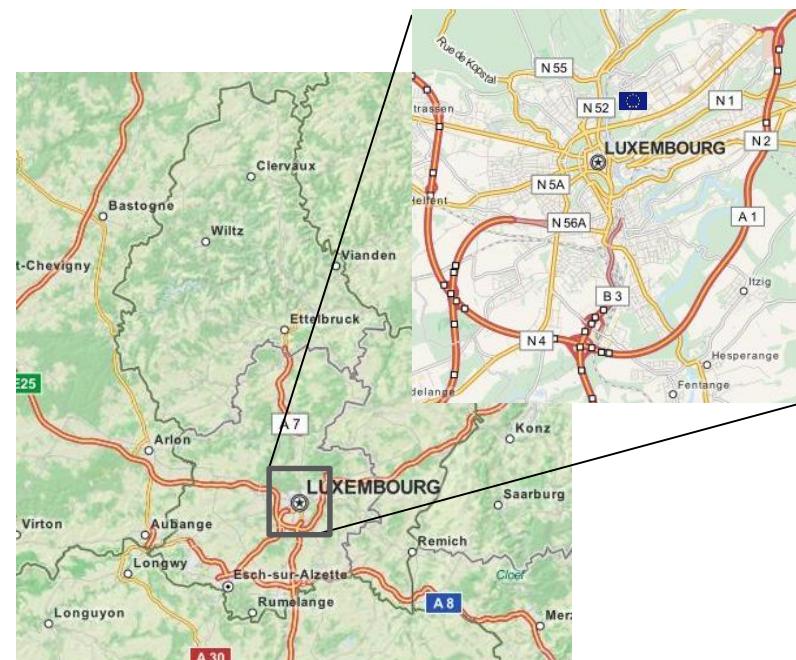
ThéoQuant, Besançon, May 20, 2015

T.-Y. Ma, P. Gerber, S. Carpentier, **S. Klein**  
LISER (CEPS/INSTEAD)  
Université Aix-Marseille (France)



## Research question

- ❑ Analyse the determinants of mode choice behaviours in terms of psychological, spatial, journey and socio-demographic attributes
- ❑ Case study : Mobility of the employees of E.U. Institutions in Luxembourg



## Commuting mode choice modeling

- ❑ The most widely used model: **logit model**
  1. Logit model: individual-specific variables
  2. Conditional logit model: alternative-specific variables
  3. **Mixed conditional logit model: individual-specific + alternative-specific variables**
- ❑ **Issues of MNL models**: how to incorporate the response heterogeneity into the MNL models?
  - **Mixed logit model**
  - **Endogenous market segmentation approach (latent class model)**: Accommodate systematic heterogeneity in a practical manner and jointly determine the number of segments and the segment-specific choice model parameters

## Why latent class model ?

### □ Latent class model

- Does not require the analyst to make specific assumptions about the distributions of parameters across individuals
- Allow to endogenously identify different preference homogenous groups
- Empirical comparisons show that the latent class model outperforms the mixed logit model in terms of goodness of fit (Greene and Hensher, 2003; Shen, 2009)

## Latent class mode choice model

- The probability that **transport mode  $j$**  is chosen by **individual  $i$** , conditional on the individual belonging to **segment  $r$** , follows the MNL form as:

$$P_{ij|r} = \frac{\exp(\tilde{\mathbf{X}}_i' \boldsymbol{\gamma}_j)}{\sum_{j'=1}^J \exp(\tilde{\mathbf{X}}_i' \boldsymbol{\gamma}_{j'})}$$

- The likelihood function of individual  $i$  can be written as :

$$L_i = \sum_{r=1}^R P_{ir} L_{i|r}$$

$L_{i|r}$  is the choice likelihood of individual  $i$ , conditional on latent class  $r$

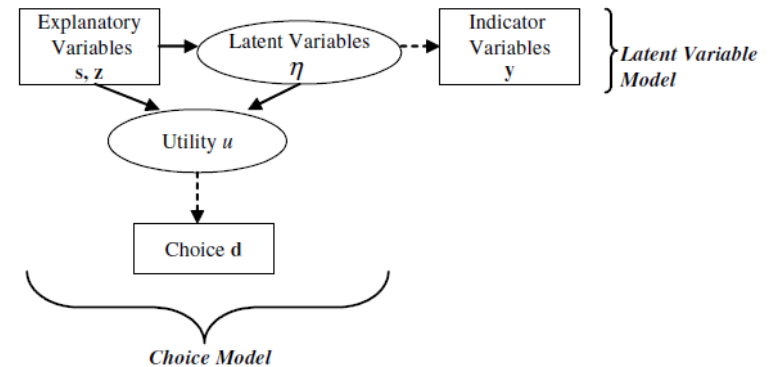


Fig. 1. Integrated choice and latent variable model (Ben-Akiva et al., 1999, p. 195).

## Latent class mode choice model : parameters estimation

The total log-likelihood function of the sample with  $N$  individuals can be obtained as:

$$LL = \sum_{i=1}^N \ln L_i = \sum_{i=1}^N \ln \left[ \sum_{r=1}^R P_{ir} L_{i|r} \right] = \sum_{i=1}^N \ln \left[ \sum_{r=1}^R P_{ir} \prod_{j \in J} P_{ij|r}^{\delta_{ij}} \right]$$

### □ Model estimation

- The parameters to be estimated in the latent class model are the parameter vectors  $\gamma$  and  $\beta$
- Estimation methods : **expectation-maximum likelihood (EM) algorithm**
- Implementation: STATA Icllogit package

## Data collection

- ❑ Mobility survey for employees working in EU Institutions in Luxembourg and Strasbourg (October-November 2012)
  - **European Investment Bank (EIB)** : 131 individuals (~6.2%)
  - **Court of Justice of the European Union (CURIA)**: 239 individuals (~11.2%)
- ❑ Web-based survey about their perceptions of the city of working place, mobility practice of trips, attitudes for transport mode, and socio-demographic characteristics
- ❑ Choice set: car and public transport (bus and/or train)
  - Only 5% by walk or bicycle
  - Use of bus and train are correlated (Pearson Correlation Coefficients 0.36864). Regroup them as public transport
- ❑ After data cleaning, **286 individuals** remained for use in our empirical study

## Model specification

- ❑ **Alternative-specific variables** : travel time, travel cost, season ticket subscription, free parking at working place
- ❑ **Individual-specific variables**: gender, couple, number of children, professional status (manager or not), presence of working spouse, number of cars, residence in Luxembourg, flexible working time or not, attitudinal variables,
- ❑ **Choice set**: car and public transport
- ❑ **Model specification**: conditional logit model, mixed conditional logit model and latent class model
- ❑ Model selection
- ❑ Model validation



**Table 3. Mode choice share and mode-specific covariates**

Mode	Choice	Average travel cost <sup>2</sup> (euro)	Average travel time <sup>1</sup> (min)	Season_Ticket <sup>3</sup>	Free_Parking <sup>4</sup>
Car	62.2%	3.6	22.3	0	86.4%
Public transport (bus and train)	37.8%	2.1	44.4	62.9%	0

- Remark: 1. Travel time is estimated from a respondent's home location to his workplace. In the case of public transport, it is the minimum travel time of overall journey from home to stations or bus stops located in a reasonable waking distance range (1km for railroad station and 0.5 km for bus stops)
2. For car, travel cost is calculated as the daily commute distance multiplied by the average monetary cost of fuel consumption. For public transport, it is calculated as the daily average price for season ticket users. The average fuel consumption for a car is 7.0 liters/100 km. For public transport, its cost is calculated using the actual rate available on <http://www.cfl.lu>.
3. Season ticket is a binary variable representing an individual's subscription to a season ticket for public transport
4. Free parking is a binary variable indicating whether the parking facility is free at or near a respondent's workplace

## Attitudinal indicators

- Three Indicators
  - Exploratory factor analysis for all attitudes -> three factors are identified
  - Compute factor scores for each individuals which represents a individual's relative standing on each of the factors
  - Factor score of Pro\_train = summation (over all variables) of loading coefficient\* value of variable (standardized)
  - Att\_pro\_mode = 1 if its factor score is maximum among three factors; 0, otherwise.

**Table 2. Factors in travel mode preference**

Factor	Variable	Loading*
<i>Pro-train</i>	Train is rapid	0.464
	Train is flexible	0.446
	Train is ecological	0.395
	Train is punctual	0.369
	Bus is flexible	0.355
<i>Pro-bus</i>	Bus is rapid	0.552
	Bus is flexible	0.476
	Bus is punctual	0.405
<i>Pro-car</i>	Car is flexible	0.524
	Car is rapid	0.489
	Train is ecological	0.365

Remark: only the correlations between the covariates and the factors with values greater than 0.3 are reported.

## Comparison of model-fit statistics for different numbers of latent classes

Number of classes	Log-likelihood at convergence	Number of parameters	CAIC	BIC
2	-131.88	21	403.55	382.55
3	-112.05	37	470.37	433.37
4	-115.86	53	584.49	531.49

Remark: Criteria of Bayesian Information Criteria (BIC)  
Constrained Akaike Information Criterion (CAIC)

Covariate	CMNL		MCMNL		Latent class model			
	Coef.	Z-value	Coef.	Z-value	Class 1		Class 2	
					Coef.	Z-value	Coef.	Z-value
<i>Class-specific mode choice model</i>								
Constant	-0.108	-0.22	-2.053**	-2.14	-0.465	-0.44	-0.367	-0.36
Travel_time	-0.051***	-3.52	-0.051***	-3.36	-0.103***	-3.04	-0.067**	-2.45
Travel_cost	-0.219***	-3.12	-0.188**	-1.98	-0.269*	-1.87	-0.018	-0.15
Free_parking	0.998**	2.56	0.899**	2.03	2.150***	3.14	-1.590	-1.19
Season_ticket	-1.380***	-4.4	-1.271***	-3.71	-1.820**	-2.01	-2.887**	-2.15
<i>Effects on latent membership probability</i>								
Constant					-5.449*	-1.76		
Male			-0.239	-0.77	-1.122	-1.36		
Couple			-0.381	-0.75	-1.313	-1.11		
N_children			0.437***	2.82	0.822**	2.13		
Working_spouse			-0.195	-0.49	0.196	0.23	Reference class	
N_car			1.005***	3.32	2.379**	2.51		
Res_Lux			0.990*	1.93	4.235**	2.28		
Flex_time			-0.105	-0.25	0.915	0.78		
Att_pro_car			0.532	1.17	3.508*	1.83		
Att_pro_bus			-0.842*	-1.91	-1.618*	-1.69		
Att_pro_train			-0.510	-1.09	-1.695*	-1.64		
Class share	---		---			68.90%		31.10%
Log-Likelihood value $LL(\beta)(LL(0))$	-159.239	(-189.586)	-138.129	(-189.586)	-131.888	(-189.586)		
McFadden's R <sup>2</sup> (adjusted R <sup>2</sup> )	0.160	(0.134)	0.271	(0.192)	0.304	(0.194)		
Percent concordant	69.93%		73.78%		85.66%			
Number of observations (individuals)	572	(286)	572	(286)	572	(286)		

1. Reference mode is public transport. 2. \*\*\* p-value ≤ 0.01, \*\* 0.01 < p-value ≤ 0.05, \* 0.05 < p-value ≤ 0.1. 3 The McFadden's adjusted R<sup>2</sup> is computed as  $1 - [LL(\beta) - k] / LL(0)$ , where k is the number of parameters of the model. 4 All the models are statistically significant at 0.0001 level compared to the null model with only constant

- ❑ **Goodness of fit** : Latent class model > MCMNL > CMNL (log-likelihood ratio test, statistically significant at least 0.05 level)
- ❑ **Percent concordant (% of corrected prediction): latent class model (85.66%) > MCMNL (73.78%) > CMNL model (69.93%)**
- ❑ **Cross validation**: 80% randomly selected sample for model estimation. 20% for test. Average corrected prediction for the 20% tested sample: **latent class model (75.88%) > MCMNL (67.11%)**
- ❑ **Value of time (VOT)** (ratio of the estimated travel time and travel cost parameters) implied by the model: For class 1 (car preferred users) is estimated as **23.0 (0.103/0.269\*60)** euros/hr. For class 2, it cannot be estimated due to travel cost is not statistically significant (H0 (the coefficient of travel cost is zero) cannot be rejected)

	Latent class	MCMNL model	Minimum wage of a qualified employee
VOT(euros/hr)	23.0	16.3	13.9

Variable	Class 1 ( <i>Car preferred users</i> )	Class 2 ( <i>Public transport preferred users</i> )
<u>Mode choice is car</u>	82.7%	16.9%
Male	40.1%	56.2%
Couple	82.7%	77.5%
<u>N children</u>	1.05	0.79
<u>Manager</u>	8.12%	2.25%
Working_spouse	65.5%	52.8%
<u>N car</u>	1.77	1.38
<u>Res Lux</u>	87.8%	50.6%
Flex_time	87.8%	77.5%
<u>Att pro car</u>	40.6%	4.5%
Att_pro_bus	18.3%	44.9%
Att_pro_train	13.7%	42.7%
Distance (km) to workplace (average (median))	13.30 (8.6)	23.71 (15.10)
Average (median) travel time by car (min)	20.1 (17.0)	27.3 (24.5)
Average (median) travel time by public transport (min)	40.5 (35.9)	53.0 (46.1)
Average reported travel time (S.D.) (min)	30.6 (18.1)	48.7 (24.4)
Class share	68.9%	31.1%

## Sensitivity analysis and policy implications

- ❑ The sensitivity analysis is based on the choice elasticities of explanatory variables, which is of particular interest for policy implications
- ❑ **Elasticity for continuous variable**: % change of choice probability with respect to 1% change in an explanatory variable
- ❑ **Elasticity for dummy variable** : % change in choice probability when a dummy variable changes from 1 to 0.
- ❑ Direct elasticity v.s. cross elasticity
- ❑ If the elasticity value is greater than **1.0**, it is considered as elastic in response to changes in an explanatory variable

## Market share prediction with respect to the variations of travel time

	Class 1					Class 2					All				
<i>Increments of travel time by car</i>															
%	0	20	40	60	80	0	20	40	60	80	0	20	40	60	80
<i>Market share (%)</i>															
car	90.9	86.8	78.2	69.5	61.9	13.5	10.1	7.9	7.9	4.5	66.8	62.9	56.3	50.4	44.1
PT	9.1	13.2	21.8	30.5	38.1	86.5	89.9	92.1	92.1	95.5	33.2	37.1	43.7	49.7	55.9
<i>Decrements of travel time by public transport</i>															
%	0	-10	-20	-30	-40	0	-10	-20	-30	-40	0	-10	-20	-30	-40
<i>Market share (%)</i>															
car	90.9	89.3	83.3	73.1	64.0	13.5	9.0	7.9	5.6	2.3	66.8	64.3	59.8	52.1	44.8
PT	9.1	10.7	16.8	26.9	36.0	86.5	91.0	92.1	94.4	97.8	33.2	35.7	40.2	47.9	55.2

- ❑ Average travel time by PT -20% from 44.4 min. to 35.52 min. -> % of PT +7%
- ❑ Average travel time by PT -30% from 44.4 min. to 31.1 min -> % of PT + 14.7%



## Conclusion

- ❑ The empirical result suggests a **two-class segmentation** of mode choice behaviour
- ❑ The first class is qualified as a car-preferred class in which most employees living in the country of Luxembourg have shorter travel distance. Travel time and the availability of free parking play important roles in their choice of the car
- ❑ The second class is qualified as a public-transport-preferred class in which travel time and distance are much longer than those for employees in class 1
- ❑ **Individual's attitudes to transport modes have consistent significant influence on their mode choice preference**
- ❑ **Reducing travel time by public transport and reducing free parking availability could effectively reduce car use in Luxembourg city**
- ❑ Extension:
  - Mode choice analysis for the cross border workers based on EMF (Enquête Mobilité des Frontaliers, 2011) survey data set

# Thank you

Ma, T.-Y., Gerber, P., Carpentier, S., & Klein, S. (2015). Mode choice with latent preference heterogeneity: a case study for employees of the EU institutions in Luxembourg.

*Transportmetrica A: Transport Science*, 1-23.

<http://doi.org/10.1080/23249935.2015.1007175>

## Summary statistics of samples (N=286)

Variable	Definition	Means
<i>Socio-demographic and spatial characteristics</i>		
Male	1 if male, 0 female (% of 1)	45.0
Age34 <sup>1</sup>	1 if the age of the individual is between [20, 35) years, else 0 (% of 1)	21.7
Age35_44	1 if the age of the individual is between [35, 45) years, else 0 (% of 1)	38.5
Age45_54	1 if the age of the individual is between [45, 55) years, else 0 (% of 1)	26.9
Age 55	1 if the age of the individual is equal or greater than 55 years, else 0 (% of 1)	12.9
Single <sup>2</sup>	1 if single, else 0 (% of 1)	23.3
Couple_no children	1 if individual lives as a couple with no child, else 0 (% of 1)	30.1
Couple_children	1 if individual lives as a couple with children, else 0 (% of 1)	46.6
N_children	Number of young children less than 15 years of age in the household	0.9
Manager	1 if individual is a manager, else 0 (% of 1)	6.1
Working_spouse	1 if individual's spouse/husband has a job, else 0 (% of 1)	58.3
N_car	Number of cars in the household	1.5
Flex_time	1 if individual has flexible working hours, else 0 (% of 1)	83.8
Edu_high	1 if education level is superior or equal to bachelor degree (% of 1)	69.9
Res_Lux	1 if the country of individual's residence is Luxembourg, else 0 (% of 1)	78.0
Distance <sup>3</sup> (km)	Distance from home to workplace on average (median)	15.6

- Employees' attitude toward transport mode (286 individuals)
- Coding: totally agree (5), rather agree (4), neither one nor the other (3), rather disagree (2), totally disagree (1)

Covariate <sup>↵</sup>	All <sup>↵</sup>	Car users <sup>↵</sup>	PT users <sup>↵</sup>	↵
Bus is rapid <sup>↵</sup>	3.08 <sup>↵</sup>	3.04 <sup>↵</sup>	3.15 <sup>↵</sup>	↵
Bus is expensive <sup>↵</sup>	1.96 <sup>↵</sup>	2.07 <sup>↵</sup>	1.78 <sup>↵</sup>	↵
Bus is ecological <sup>↵</sup>	3.60* <sup>↵</sup>	3.63* <sup>↵</sup>	3.56* <sup>↵</sup>	↵
Bus is dangerous <sup>↵</sup>	2.33 <sup>↵</sup>	2.41 <sup>↵</sup>	2.19 <sup>↵</sup>	↵
Bus is flexible <sup>↵</sup>	3.04 <sup>↵</sup>	2.87 <sup>↵</sup>	3.32 <sup>↵</sup>	↵
Bus is tired <sup>↵</sup>	2.74 <sup>↵</sup>	2.76 <sup>↵</sup>	2.70 <sup>↵</sup>	↵
Bus is punctual <sup>↵</sup>	3.27 <sup>↵</sup>	3.34 <sup>↵</sup>	3.15 <sup>↵</sup>	↵
Train is rapid <sup>↵</sup>	3.41 <sup>↵</sup>	3.35 <sup>↵</sup>	3.49 <sup>↵</sup>	↵
Train is expensive <sup>↵</sup>	3.13 <sup>↵</sup>	3.22 <sup>↵</sup>	2.97 <sup>↵</sup>	↵
Train is ecological <sup>↵</sup>	3.89* <sup>↵</sup>	3.85* <sup>↵</sup>	3.94* <sup>↵</sup>	↵
Train is dangerous <sup>↵</sup>	2.16 <sup>↵</sup>	2.19 <sup>↵</sup>	2.11 <sup>↵</sup>	↵
Train is flexible <sup>↵</sup>	3.16 <sup>↵</sup>	3.01 <sup>↵</sup>	3.41 <sup>↵</sup>	↵
Train is tired <sup>↵</sup>	2.42 <sup>↵</sup>	2.47 <sup>↵</sup>	2.33 <sup>↵</sup>	↵
Train is punctual <sup>↵</sup>	3.15 <sup>↵</sup>	3.19 <sup>↵</sup>	3.10 <sup>↵</sup>	↵
Car is rapid <sup>↵</sup>	3.91* <sup>↵</sup>	4.03** <sup>↵</sup>	3.71* <sup>↵</sup>	↵
Car is expensive <sup>↵</sup>	3.96* <sup>↵</sup>	3.93* <sup>↵</sup>	4.01** <sup>↵</sup>	↵
Car is ecological <sup>↵</sup>	2.00 <sup>↵</sup>	2.03 <sup>↵</sup>	1.94 <sup>↵</sup>	↵
Car is dangerous <sup>↵</sup>	3.06 <sup>↵</sup>	3.00 <sup>↵</sup>	3.17 <sup>↵</sup>	↵
Car is flexible <sup>↵</sup>	4.40** <sup>↵</sup>	4.54** <sup>↵</sup>	4.18** <sup>↵</sup>	↵
Car is tired <sup>↵</sup>	3.32 <sup>↵</sup>	3.23 <sup>↵</sup>	3.46 <sup>↵</sup>	↵

Remark: \* :  $3.5 \leq \text{value} < 4$ ; \*\* :  $4 \leq \text{value}$  ↵