Demand for public transport services: Integrating qualitative and quantitative methods

Bilge Atasoy
Aurélie Glerum
Ricardo Hurtubia
Michel Bierlaire

Transport and mobility laboratory
EPFL

STRC
September 1, 2010

Outline

Outline

Introduction

- The research is carried out with CarPostal: bus service in rural and low density areas of Switzerland.
- The aim of the project is to ...
  - Understand the travel behavior in the area of interest
  - Come up with latent attitudes and perceptions affecting travel behavior
  - Integrate these latent attitudes into mode choice context
  - Improve the market share of public transport

Data collection

Data collection campaign consists of the execution of 3 surveys:

- Qualitative survey
- Revealed preferences (RP) survey
- Stated preferences (SP) survey

Data collection

Data collection campaign consists of the execution of 3 surveys:

- Qualitative survey
  - In-depth interviews with 20 individuals
  - Identification of potential latent attitudes
  - Ideas for the design of quantitative surveys
- Revealed preferences (RP) survey
- Stated preferences (SP) survey

Data collection

Data collection campaign consists of the execution of 3 surveys:

- Qualitative survey
  - In-depth interviews with 20 individuals
  - Identification of potential latent attitudes
  - Ideas for the design of quantitative surveys
- Revealed preferences (RP) survey
  - Travel diary for a predefined day
  - Socio-economic characteristics
  - Habits
  - Psychometric indicators
- Stated preferences (SP) survey

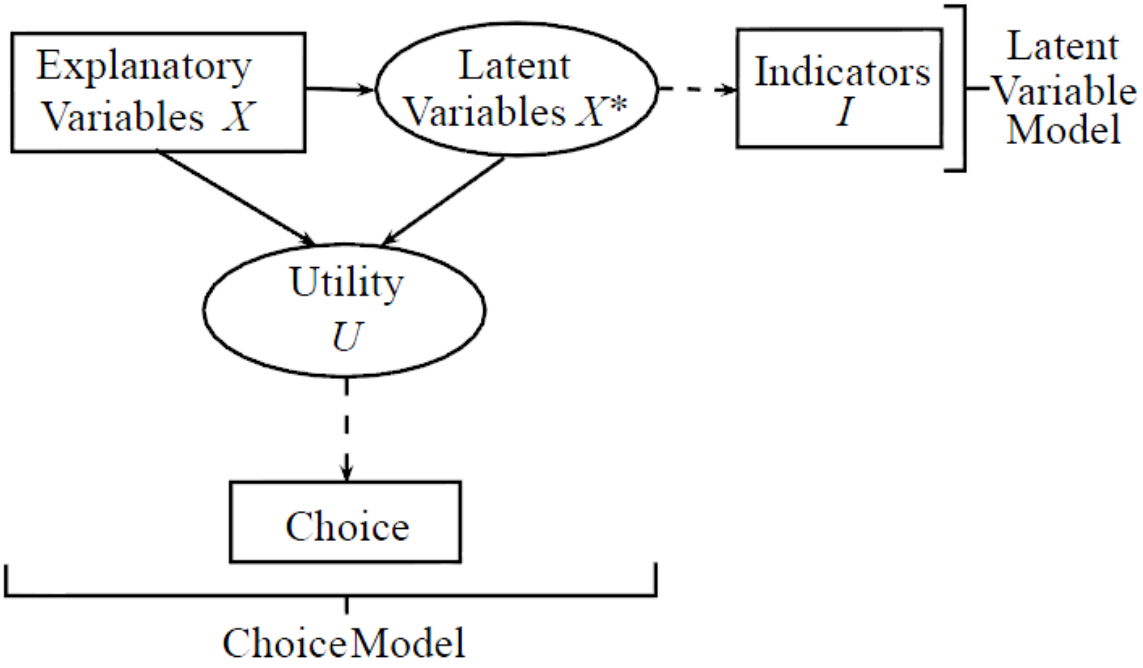
Data collection

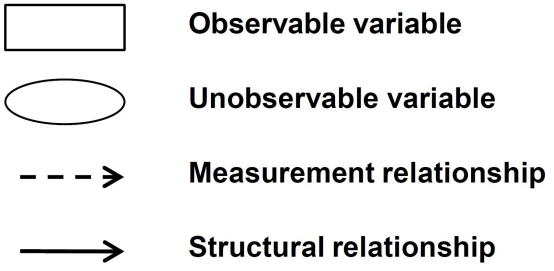
Data collection campaign consists of the execution of 3 surveys:

- Qualitative survey
  - In-depth interviews with 20 individuals
  - Identification of potential latent attitudes
  - Ideas for the design of quantitative surveys
- Revealed preferences (RP) survey
  - Travel diary for a predefined day
  - Socio-economic characteristics
  - Habits
  - Psychometric indicators
- Stated preferences (SP) survey
  - Hypothetical choice situations with improved alternatives:
    - information services, electric bikes, neighborhood service...

Outline

Integrated model framework





Integrated model equations

Structural equations:

- Latent var - explanatory var
  $X_n^* = h(X_n; \lambda) + \omega_n$
- Utilities
  $U_n = V(X_n, X_n^*; \beta) + \varepsilon_n$

Measurement equations:

- Latent var - indicators
  $I_n^* = f(X_n^*; \alpha) + \upsilon$
- Choice
  $P(y_{in} = 1) = P(U_{in} \geq U_{jn}, \forall j)$

Likelihood:

$$L_n(y_n, I_n | X_n; \alpha, \beta, \theta_\varepsilon, \theta_\upsilon, \theta_\omega) =$$

$$\int_{X^*} P(y_n | X_n, X_i, X^*; \beta, \theta_\varepsilon) f(I_n | X_n, X^*; \alpha, \theta_\upsilon) f(X^* | X_n; \lambda, \theta_\omega) \mathrm{d}X^*.$$

Outline

Latent variables

- With the psychometric indicators in RP survey we are able to estimate models with unobserved variables like *attitudes*, *perceptions*, *lifestyle preferences* etc.
- Factor analysis is performed to come up with the most powerful latent factors together with their most explaining indicators. First 6 of them are:
  - Attitude against public transport
  - Environmental concern
  - Public transport awareness
  - Status seeking
  - Pro high density
  - Personalized service

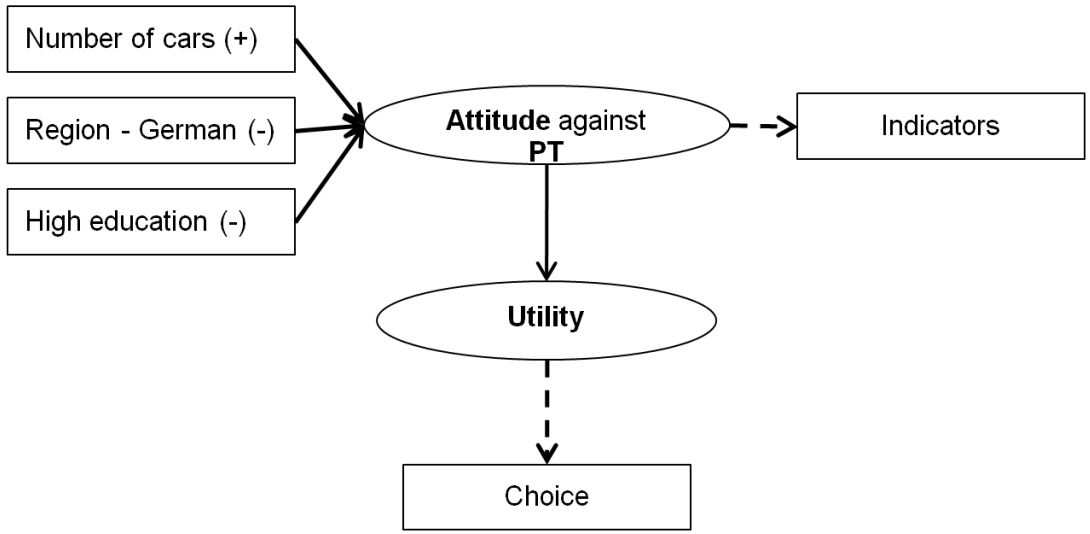
Latent variables

- With the psychometric indicators in RP survey we are able to estimate models with unobserved variables like *attitudes*, *perceptions*, *lifestyle preferences* etc.
- Factor analysis is performed to come up with the most powerful latent factors together with their most explaining indicators. First 6 of them are:
  - Attitude against public transport
  - Environmental concern
  - Public transport awareness
  - Status seeking
  - Pro high density
  - Personalized service

Attitude against PT

Psychometric indicators (level of agreement: likert scale 1-5)

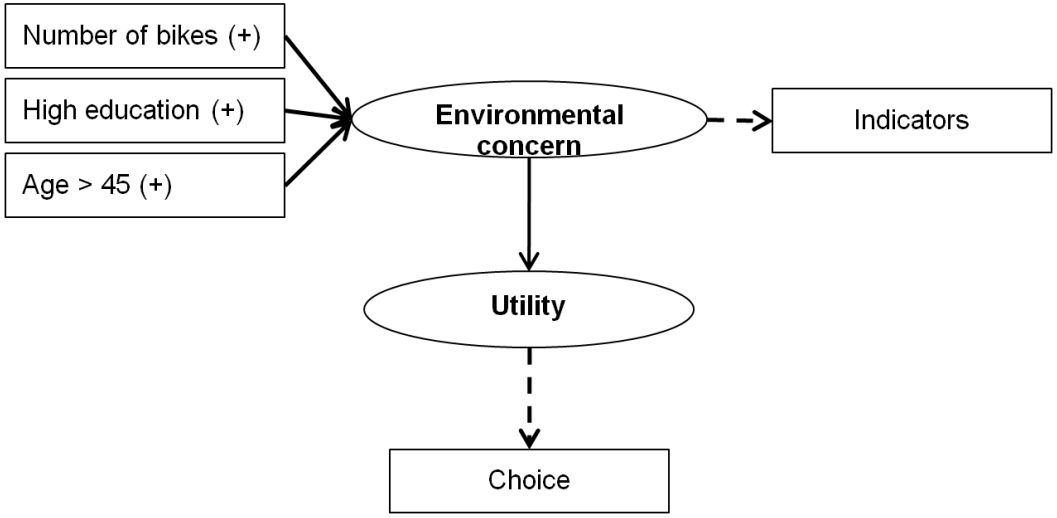
- It is hard to take PT when I travel with my children.
- I do not like to change means of transportation when I travel.
- It is hard to take PT when I have bags or luggage.



Environmental concern

Psychometric indicators (level of agreement: likert scale 1-5)

- I am concerned about global warming.
- We should increase the price of gasoline to reduce congestion and air pollution.
- We must act and make decisions to reduce emissions of greenhouse gases.
- We need more public transport, even if it means higher taxes.



Outline

Integrated latent variable and discrete choice model

- Data from RP survey is used (1124 received questionnaires).
- Concerned choice is the mode choice for each loop of trips in a day (1339 loops).
- There are 3 alternatives for choice:
  - Private mode: car, taxi, car-sharing, moto etc.
  - Public transportation
  - Soft mode: bike and walking

Model specification

$$
\begin{aligned}
V_{PM} &= ASC_{PM} + \beta_{cost}\ C_{PM} + \beta_{TT_{PM}}\, TT_{PM} \\
&\quad + \beta_{children}\ I_{children} + \beta_{w}\ I_{w} + \beta_{french}\ I_{french} \\
V_{PT} &= ASC_{PT} + \beta_{cost}\ C_{PT} + \beta_{TT_{PT}}\, TT_{PT} \\
&\quad + \beta_{freq}\ F_{PT} + \beta_{attAPt}\ attAPt + \beta_{attEnv}\ attEnv \\
V_{SM} &= ASC_{SM} + \beta_{distance}\ D_{SM}
\end{aligned}
$$

$$
P_i = \frac{exp(V_i)}{exp(V_{PM}) + exp(V_{PT}) + exp(V_{SM})} \quad i = PM, PT, SM.
$$

▸ Equations for latent variables

Outline

Estimation results - Choice model

		Integrated Model		Multinomial Logit	
Parameter	Utility	Value	t-test	Value	t-test
ASC_{PM}	PM	0.157	0.15*	0.81	3.35
ASC_{PT}	PT	0**	-	0**	-
ASC_{SM}	SM	-0.409	-0.38*	0.218	0.56*
$\beta_{TT_{PM}}$	PM	-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$	PT	-0.00847	-3.1	-0.00846	-2.79
β_{cost}	PM & PT	-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$	SM	-0.221	-4.47	-0.222	-4.44
$\beta_{children}$	PM	0.492	3.09	0.412	2.62
β_{w}	PM	-0.61	-3.97	-0.622	-4.1
β_{french}	PM	1.05	6.22	1.09	6.5
β_{attAPt}	PT	-0.63	-2.89	-	-
β_{attEnv}	PT	0.326	1.89	-	-
β_{freq}	PT	0.649	3.22	0.701	3.51

(* Statistical significance $<$ 90%, ** Fixed parameter)

Estimation results - Choice model

		Integrated Model		Multinomial Logit	
Parameter	Utility	Value	t-test	Value	t-test
ASC_{PM}	PM	0.157	0.15*	0.81	3.35
ASC_{PT}	PT	0**	-	0**	-
ASC_{SM}	SM	-0.409	-0.38*	0.218	0.56*
$\beta_{TT_{PM}}$	PM	-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$	PT	-0.00847	-3.1	-0.00846	-2.79
β_{cost}	PM & PT	-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$	SM	-0.221	-4.47	-0.222	-4.44
$\beta_{children}$	PM	0.492	3.09	0.412	2.62
β_{w}	PM	-0.61	-3.97	-0.622	-4.1
β_{french}	PM	1.05	6.22	1.09	6.5
β_{attAPt}	PT	-0.63	-2.89	-	-
β_{attEnv}	PT	0.326	1.89	-	-
β_{freq}	PT	0.649	3.22	0.701	3.51

(* Statistical significance $<$ 90%, ** Fixed parameter)

Estimation results - Choice model

		Integrated Model		Multinomial Logit	
Parameter	Utility	Value	t-test	Value	t-test
ASC_{PM}	PM	0.157	0.15*	0.81	3.35
ASC_{PT}	PT	0**	-	0**	-
ASC_{SM}	SM	-0.409	-0.38*	0.218	0.56*
$\beta_{TT_{PM}}$	PM	-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$	PT	-0.00847	-3.1	-0.00846	-2.79
β_{cost}	PM & PT	-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$	SM	-0.221	-4.47	-0.222	-4.44
$\beta_{children}$	PM	0.492	3.09	0.412	2.62
β_{w}	PM	-0.61	-3.97	-0.622	-4.1
β_{french}	PM	1.05	6.22	1.09	6.5
β_{attAPt}	PT	-0.63	-2.89	-	-
β_{attEnv}	PT	0.326	1.89	-	-
β_{freq}	PT	0.649	3.22	0.701	3.51

(* Statistical significance $<$ 90%, ** Fixed parameter)

Estimation results - Choice model

		Integrated Model		Multinomial Logit	
Parameter	Utility	Value	t-test	Value	t-test
ASC_{PM}	PM	0.157	0.15*	0.81	3.35
ASC_{PT}	PT	0**	-	0**	-
ASC_{SM}	SM	-0.409	-0.38*	0.218	0.56*
$\beta_{TT_{PM}}$	PM	-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$	PT	-0.00847	-3.1	-0.00846	-2.79
β_{cost}	PM & PT	-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$	SM	-0.221	-4.47	-0.222	-4.44
$\beta_{children}$	PM	0.492	3.09	0.412	2.62
β_{w}	PM	-0.61	-3.97	-0.622	-4.1
β_{french}	PM	1.05	6.22	1.09	6.5
β_{attAPt}	PT	-0.63	-2.89	-	-
β_{attEnv}	PT	0.326	1.89	-	-
β_{freq}	PT	0.649	3.22	0.701	3.51

(* Statistical significance $<$ 90%, ** Fixed parameter)

Estimation results - Choice model

		Integrated Model		Multinomial Logit	
Parameter	Utility	Value	t-test	Value	t-test
ASC_{PM}	PM	0.157	0.15*	0.81	3.35
ASC_{PT}	PT	0**	-	0**	-
ASC_{SM}	SM	-0.409	-0.38*	0.218	0.56*
$\beta_{TT_{PM}}$	PM	-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$	PT	-0.00847	-3.1	-0.00846	-2.79
β_{cost}	PM & PT	-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$	SM	-0.221	-4.47	-0.222	-4.44
$\beta_{children}$	PM	0.492	3.09	0.412	2.62
β_{w}	PM	-0.61	-3.97	-0.622	-4.1
β_{french}	PM	1.05	6.22	1.09	6.5
β_{attAPt}	PT	-0.63	-2.89	-	-
β_{attEnv}	PT	0.326	1.89	-	-
β_{freq}	PT	0.649	3.22	0.701	3.51

(* Statistical significance $<$ 90%, ** Fixed parameter)

Estimation results - Attitude against PT

Parameter	Value	t-test
β_{attAPt}	-0.63	-2.89
$\overline{attAPt}$	3.45	54.33
λ_{cars}	0.129	3.52
$\lambda_{high-educ}$	-0.262	5.9
$\lambda_{region-BL}$	-0.307	-3.66
$\lambda_{region-GR}$	-0.234	-2.02
$\lambda_{region-SG}$	-0.315	-3.01
$\lambda_{region-VS}$	-0.193	-2.12
$\lambda_{region-BE}$	-0.467	-3.01

Parameter	Value	t-test
a_{16}	0**	-
a_{17}	0.847	2.28
a_{22}	1.24	4.4
α_{16}	1**	-
α_{17}	0.974	7.84
α_{22}	0.727	7.57
θ_{attAPt}	-0.469	-6.33
θ_{16}	-0.255	-5.03
θ_{17}	-0.126	-3.21
θ_{22}	0.0171	0.71*

Estimation results - Attitude against PT

Parameter	Value	t-test
β_{attAPt}	-0.63	-2.89
$attAPt$	3.45	54.33
λ_{cars}	0.129	3.52
$\lambda_{high-educ}$	-0.262	5.9
$\lambda_{region-BL}$	-0.307	-3.66
$\lambda_{region-GR}$	-0.234	-2.02
$\lambda_{region-SG}$	-0.315	-3.01
$\lambda_{region-VS}$	-0.193	-2.12
$\lambda_{region-BE}$	-0.467	-3.01

Parameter	Value	t-test
a_{16}	0**	-
a_{17}	0.847	2.28
a_{22}	1.24	4.4
α_{16}	1**	-
α_{17}	0.974	7.84
α_{22}	0.727	7.57
θ_{attAPt}	-0.469	-6.33
θ_{16}	-0.255	-5.03
θ_{17}	-0.126	-3.21
θ_{22}	0.0171	0.71*

Estimation results - Environmental concern

Parameter	Value	t-test
β_{attEnv}	0.326	1.89
$attEnv$	3.04	34.67
$\lambda_{age>45}$	0.00609	2.59
λ_{bikes}	0.0605	4.17
$\lambda_{high-educ}$	0.262	5.9

Parameter	Value	t-test
a_1	-1.77	-2.81
a_2	0.0318	0.07*
a_5	0**	-
a_6	1.06	5
α_1	1.17	6.86
α_2	0.904	7.12
α_5	1**	-
α_6	0.87	15.63
θ_{attEnv}	-0.492	-5.44
θ_1	0.0873	2.6
θ_2	-0.00741	-0.26*
θ_5	-0.174	-3.94
θ_6	-0.582	-12.87

Estimation results - Environmental concern

Parameter	Value	t-test
β_{attEnv}	0.326	1.89
$attEnv$	3.04	34.67
$\lambda_{age>45}$	0.00609	2.59
λ_{bikes}	0.0605	4.17
$\lambda_{high-educ}$	0.262	5.9

Parameter	Value	t-test
a_1	-1.77	-2.81
a_2	0.0318	0.07*
a_5	0**	-
a_6	1.06	5
α_1	1.17	6.86
α_2	0.904	7.12
α_5	1**	-
α_6	0.87	15.63
θ_{attEnv}	-0.492	-5.44
θ_1	0.0873	2.6
θ_2	-0.00741	-0.26*
θ_5	-0.174	-3.94
θ_6	-0.582	-12.87

Results

- **Value of time**

	VOT_{PM} (CHF/h)	VOT_{PT} (CHF/h)
Integrated model	25.7	10.3
Multinomial logit	25.4	10.0

- **Validation** is done by estimating the model on 80% of the data and predicting the remaining 20%. 66% of the estimated choice probabilities are above 0.5 and 19% are above 0.9.
- **Demand elasticities**

	Time elasticity	Cost elasticity
Private mode	-0.20	-0.06
Public transport	-0.34	-0.17

Outline

Conclusions

- With latent variables we are able to have a better understanding of travel behavior.
- When latent attitudes are introduced effect of cost in the utilities is decreased compared to MNL.
- Utility of private mode is explained better compared to MNL since constant becomes insignificant.
- Existence of high education in structural equations of both latent attitudes

Future research

- Inclusion of more than 2 latent variables
- Discrete specification of indicators
- Modeling perceptions as well as attitudes
- Latent classes

- Analysis of SP data
- Model with SP data
- Proposal of improved alternatives

Thank you for your attention !

Equations for latent variables

Back

Structural equations for latent variables:

$$Att = \overline{Att} + \sum_e \lambda_e X_e + \omega,$$

Measurement equations for latent variables:

$$I_k = a_k + \alpha_k Att + \upsilon_k \qquad \forall k,$$