Demand for public transport services: Integrating qualitative and quantitative methods

Bilge Atasoy

Aurélie Glerum

Ricardo Hurtubia

Michel Bierlaire

STRC 2010 **September 2010**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

STRC 2010

Demand for public transport services: Integrating qualitative and quantitative methods

Bilge Atasoy
TRANSP-OR
EPFL
Lausanne
phone: +41 21 693 93 29
fax: +41 21 693 80 60
bilge.kucuk@epfl.ch

Aurélie Glerum
TRANSP-OR
EPFL
Lausanne
phone: +41 21 693 24 35
fax: +41 21 693 80 60
aurelie.glerum@epfl.ch

Ricardo Hurtubia
TRANSP-OR
EPFL
Lausanne
phone: +41 21 693 93 29
fax: +41 21 693 80 60
ricardo.hurtubia@epfl.ch

Michel Bierlaire
TRANSP-OR
EPFL
Lausanne
phone: +41 21 693 25 37
fax: +41 21 693 80 60
michel.bierlaire@epfl.ch

September 2010

Abstract

This research is in the context of a mode choice study in Switzerland. This paper represents the discrete choice modeling part of this study. A comprehensive data collection campaign is carried out which also includes psychometric indicators for attitudes, perceptions and lifestyle preferences. With the help of these indicators an integrated choice and latent variable model is built including the latent attitudes of *attitude against public transport* and *environmental concern*.

Keywords

discrete choice, latent variable models, mode choice

1 Introduction

In transport mode choice modeling, qualitative aspects are becoming more important as a result of the desire for better demand estimates. Although socio-economic attributes cover some qualitative characteristics, there are other unobserved factors (latent variables) that are important in mode choice behavior such as lifestyle preferences, personal attitudes or perceptions. These variables enrich the choice models providing a better insight about the decision making process. The integration of the latent variables requires qualitative methods to be able to come up with an initial set of these factors. The purpose of this study is to combine the qualitative and quantitative methods to have a more powerful transport mode choice model at hand.

The research is carried out in the context of a collaborative work between EPFL's Transportation Center (TraCe) and CarPostal, the public transport branch of the Swiss Postal Service. The main purpose of this study is to analyze the travel behavior of people in low-density areas, where CarPostal typically serves, and afterwards propose new public transport alternatives according to the respondents' willingness to pay for these potential services in order to increase the market share of public transport.

This paper presents an integrated discrete choice and latent variable model including attitudes as latent variables. These variables are not observed but can be deduced from psychometric indicators which are collected in a survey. Indicators are a series of statements to which respondents give their level of agreement. With the help of these indicators, latent variables can be measured and structural equations can be built in order to integrate the latent variables into the mode choice context.

The rest of the paper is organized as follows: section 2 summarizes the theoretical formulation for the integrated choice and latent variable models. In section 3 the data collection and the factor analysis for the identification of latent variables are explained. Section 4 provides the model specification together with the estimations and finally we conclude and talk about future research in section 5.

2 Integrated choice and latent variable model

The integrated choice and latent variable model presented in this paper uses the extended framework for choice behavior provided by Ben-Akiva *et al.* (1999) and Walker and Ben-Akiva (2002). Framework consists of two components: a discrete choice model and a latent variable model, each having its own set of measurement and structural equations. Unobserved variables are represented by ellipses and observable variables by rectangles. Besides, dashed lines correspond to the measurement equations and straight lines represent structural equations as in

Figure 1.

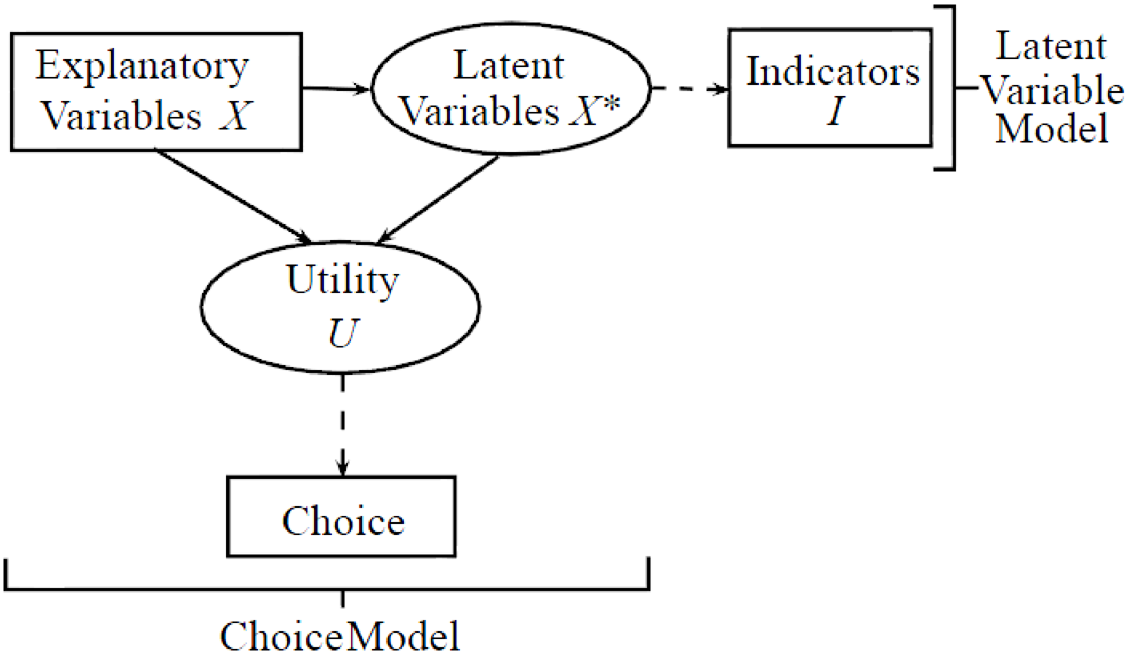

Figure 1: Integrated choice and latent variable model

Latent variables, that are represented by X_n^*, are unobserved variables related to attitudes, perception, preferences which can be measured with indicators such as psychometric indicators in surveys. Therefore measurement equations are built in the form of equation (1) to be able to measure latent variables with the psychometric indicators:

$$I_n = f(X_n^*; \alpha) + \upsilon_n, \tag{1}$$

where I_n is the indicator for individual n which is a function of the latent variables, a set of parameters (α) and an error term (υ_n). Density function of the indicators, $f(I_n|X^*; \alpha, \theta_\upsilon)$, can be obtained using the distribution of υ_n with a standard deviation of θ_υ.

The structural equations for the latent variables, shown in equation (2), are built in the same way as the classical utility function with explanatory variables (X_n) like socio-economic characteristics of individual n. In the equation, λ is a set of parameters and ω_n is an error term. The assumptions regarding the distribution of ω_n is employed in writing the density function of latent variables, $f(X^*|X_n; \lambda, \theta_\omega)$, θ_ω being the standard deviation of ω_n. The simultaneous estimation of these structural and measurement equations enables us to include unobserved constructs in choice models.

$$X_n^* = h(X_n; \lambda) + \omega_n. \tag{2}$$

Having defined the relations related to the latent variables, the utility for choosing alternative i can be expressed as a function of the individual characteristics (X_n), modal attributes (X_i) and latent variables with the following structural equation:

$$U_{in} = V(X_n, X_i, X_n^*; \beta) + \varepsilon_{in}, \tag{3}$$

where β is a set of parameters and ε_{in} is an error term . In a discrete choice context the probability of individual n choosing alternative i can be written as follows:

$$P(i|X_{n,}X_i, X_n^*; \beta, \theta_\varepsilon) = Prob[U_{in} \geq U_{jn}, \forall j \in C_n], \tag{4}$$

where θ_ε is the standard deviation of the error term in equation (3) and C_n is the choice set of individual n. With the integrated model there are two sets of measurement equations which result with a joint probability of observing choice i and indicator I_n expressed in equation (5). Since X_n^* is not observable, to be able to write this probability density functions of latent variables and indicators are incorporated.

$$P(i, I_n|X_{n,}X_i; \beta, \alpha, \lambda, \theta_\varepsilon, \theta_\upsilon, \theta_\omega) =$$

$$\int_{X^*} P(i|X_n, X_i, X^*; \beta, \theta_\varepsilon) f(I_n|X^*; \alpha, \theta_\upsilon) f(X^*|X_n; \lambda, \theta_\omega) \mathrm{d}X^*. \tag{5}$$

Maximum likelihood estimation is used to estimate the unknown parameters. The log-likelihood function (L) can be written as in equation (7) with the definition of y_{in} as in 6.

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

$$L = \sum_n \sum_{i \in C_n} y_{in} \log P(i, I_n, |X_{n,}X_i; \beta, \alpha, \lambda, \theta_\varepsilon, \theta_\upsilon, \theta_\omega). \tag{7}$$

As mentioned before, indicators are used to measure the latent variables and they do not directly influence the choice. Therefore once the parameters of the model are estimated, probability of choice in 8 can be used for application purposes. This formula can be applied to a data set containing the observable variables of modal attributes and individual characteristics.

$$P(i|X_n, X_i) = \int_{X^*} P(i|X_n, X_i, X^*) f(X^*|X_n) \mathrm{d}X^*. \tag{8}$$

There are many examples of the application of integrated choice and latent variable approach in different choice contexts. For example, Ben-Akiva and Boccara (1995); Espino *et al.* (2006) study mode choice behavior for suburban trips, Walker and Li (2007) work on residential location choice with lifestyle preferences, and Abou-Zeid *et al.* (2008) study travel behavior with measurement of travel well-being. These are just a few examples of the studies in the literature, which report an improvement in the quality of the estimates and the achievement of more realistic models when including unobserved factors through the latent variable approach.

3 Data Collection and Factor Analysis

The data collection campaign consists of three surveys in the area of study (non-urban areas served by CarPostal). First, a qualitative survey (informal interviews) is performed to have an insight about behavior of people which is valuable for the construction of the subsequent surveys. The second survey is a revealed preferences (RP) regarding travel behavior and a set of psychometric indicators to measure the latent variables. The third survey is the stated preferences (SP) survey where the respondents were presented hypothetical choice situation with proposed new alternatives, which were designed in the light of the gained insight in the first two surveys. This section focuses in the construction of a set of latent variables together with their psychometric indicators that are included in RP survey and the analysis of the results regarding these variables for modeling purposes.

3.1 Data Collection

The qualitative survey was conducted by Urban Sociology Laboratory (LASUR) of EPFL which consists of interviews to 20 individuals in the Swiss canton of Vaud, focusing on residential choice, mobility biography, and mobility habits. In addition to this, each of the 20 respondents were asked to carry a GPS with them for seven days, recording all their movements. The geocoded results were shown afterward to the respondents, where they identified the transport modes and purposes associated to each trip. During this part, additional (and informal) questions were made in order to complement the information already collected in the first part of the interviews.

The main use of this survey considering this paper is to identify the potential latent variables through a detailed analysis which is provided in Doyen (2010). Together with this analysis several papers in literature are used to enrich our set of candidates including Kitamura *et al.* (1997), Bagley and Mokhtarian (2002), Ory and Mokhtarian (2005) and Espino *et al.* (2006). This process resulted with a set of latent variables to be measured with RP survey. These potential latent variables include the attitude towards public transportation, car, environmental issues; lifestyle preferences related to residential choice, family, working hours etc.

To be able to measure these latent variables through RP survey a list of psychometric indicators are determined based on examples found in the literature Redmond (2000), Vredin Johansson *et al.* (2006), Kitamura *et al.* (1997), Ory and Mokhtarian (2005). These indicators are built as statements related with each of the potential latent variables, where the respondents are expected to give their level of agreement with the statements. The indicators designed in such a way that there are both negative and positive statements as well as trade-off statements to be able come up with a reasonable set of data.

For example, for the latent variable of *environmental concern*, some of the proposed statements are the following:

- *People and employment are more important than the environment*
- *I am concerned about global warming*
- *We should increase the price of gasoline to reduce congestion and air pollution*

These statements try to capture the positive or negative attitude related to environmental issues. It is worth to note that the third statement displays a trade-off between the increased price of gasoline and reduced pollution, which aims to identify the difference between the thoughts and actions. The rest of the latent variables are related with statements in a similar way. A list of 54 statements with a 5 level likert scale to indicate the level of agreement (from total disagreement to total agreement) is generated and included in the RP survey. RP survey also includes a travel diary and a set of questions related to the socioeconomic characteristics of the respondent together with his/her household. Travel diary is the skeleton of this data set where respondents registered all the trips performed during a specified day. The collected information consists of origin, destination, cost, travel time, chosen mode and activity at the destination.

The model presented in this paper is based on the data from RP survey, which resulted with a collection of 1124 completed surveys. For each respondent, cyclic sequences of trips (starting and ending at the same location) are detected and their main transport mode is identified. The data is used to generate the estimation database, with 1339 observations relating sequences of trips, psychometric indicators and socioeconomic attributes.

It is important to note that the observations in this data set are weighted according to the statistical data of Switzerland considering 6 dimensions: presence of driving license, gender, education, number of cars in the household, age, and number of people in the household. These weights for individuals are used to correct the values of elasticities and similar statistical information according to the proportions of each socioeconomic segment in the Swiss population of the towns selected in the survey.

3.2 Factor Analysis

As mentioned in section 3.1 an initial set of latent variables are selected to construct the set of indicators. However it is better to make use of the data to identify the latent variables to be used in the modeling process. Therefore factor analysis methodology is used to identify the relations between the indicators with collected answers for the level of agreement on the 54 statements.

First of all, an exploratory factor analysis is performed with the whole list of indicators in order to identify the correlations between indicators, grouping them in factors that represent the latent variables. The relation between the indicators (I_k) and the unobserved factors (F_j) is given by the following equation:

$$I_k = \overline{I_k} + \sum_j \rho_{kj} F_j + \varphi_k, \tag{9}$$

where $\overline{I_k}$ is the mean value of the answer for indicator k and φ_k is an error term following a normal distribution. The factor loadings (ρ_{kj}) quantify the correlation between the indicator k and factor j.

The factor analysis gives with 17 relevant factors with an eigenvalue greater than 1, that together explain 57% of the total variance among indicators. For modeling purposes first 6 factors (those that explain most of the variability) are found to be interesting and with these 6 factors a confirmatory factor analysis is performed the results of which can be seen in Table 1. This table shows the factor loadings and their related indicators. The factor loadings which have an absolute value greater than 0.2 are displayed which are all significant at the 95% level. The list of statements for the given indicators are provided in Table 6 in the Appendix.

When we analyze the factor loadings with their signs we can group them and come up latent variables. For example, for Factor 1 the three biggest factor loadings in absolute value are those of indicators 17, 22, and 23. These indicators are the following statements in the survey:

Table 1: Factor loadings

Indicator	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	-0.540	0.592	-	-	-	-
2	-0.252	0.452	-	-	-	-
3	0.357	-0.357	-	-	-	-
4	-	-0.435	-	-	-	-
5	-	0.627	-	-	-	-
6	-	0.581	-	-	-	-
9	-	-	0.271	-	-	-
11	-	-	-	-	-	0.258
14	0.311	-	-	0.290	-	-
15	- 0.224	-	-	-	-	-
16	0.521	-	-	-	-	-
17	0.613	-	-	-	-	-
18	-	-	-	0.622	-	-
20	0.385	-	-0.344	-	-	-
22	0.564	-	-	-	-	-
23	0.562	-	-	-	-	-
24	0.378	-	-0.302	-	-	-
25	-0.269	-	0.665	-	-	-
26	-	-	0.758	-	-	-
28	-	-	-0.323	-	-	-
29	-	-	0.295	-	-	-
30	-	-	0.846	-	-	-
32	-	-	0.430	-	-	0.615
33	-	-	-	-	-	0.750
34	-	-	-	-	0.425	-
35	-	-	-	-	0.274	0.231
38	-	-	-	-	0.684	-
39	-	-	-0.254	-	0.421	-
40	-	-	-	-	-0.639	-
46	-	-	-	0.315	-	-
47	-	-	-	0.551	-	-
48	-	-	-	-0.509	-	-
50	0.209	-	-	-	-	-

- I_{17} : *It's hard to take public transport when I have bags or luggage.*
- I_{22} : *I don't like to change transport modes when I travel.*
- I_{23} : *If I use public transport, I have to cancel some activities that I do, so I take the car.*

If a person displays a high agreement with the above statements, he has a negative attitude for public transport stressing the difficulties and inconvenience in using public transport. When other indicators are analyzed in a similar way, it is seen that this factor represent a negative attitude for public transport which can be named as *attitude against public transport*.

Extending the analysis to other factors, it is possible to conclude that factors from 2 to 6 represent *environmental concern*, *public transport awareness, status seeking*, *pro-high-density*, and *personalized service* respectively. *Environmental concern* represent the behavior of people who are concerned with environmental issues and who are ready to take action for that. *Public transport awareness* explains the behavior of people who have used public transport throughout their life and who are aware of the possibilities and conditions in using public transport. Latent variable of *status seeking* shows the behavior of giving importance to having valuable belongings and showing it to other people. *Pro-high-density* is the behavior regarding the desire to live in strategical centers of the city with its all social facilities and finally *Personalized service* capture the behavior of people for whom it is an important issue to have a contact with the drivers to have a better service.

4 Model Specification and Estimation Results

As described in section 2 the integrated model has two parts: latent variable model and discrete choice model. Building the latent variable is related to constructing the measurement equations for latent variables with psychometric indicators and defining the latent variable with explanatory variables through structural equations. In the discrete choice part, utility of alternatives are explained by modal attributes, socioeconomic characteristics and latent variables. Since integrated models are complicated models, to save time initial analysis was done separately for the two parts which resulted with a good set of explanatory variables for each. Afterwards the integrated model was estimated in the light of the initial analysis. A model with the preliminary analysis of this study can be found in Hurtubia *et al.* (2010).

4.1 Latent variable model

The resulting factors from the factor analysis in section 3.2 are candidates for the latent variables to be used in the model estimation. From these candidates we worked with *attitude against public transport* and *environmental concern* for the integrated model. Although we have analyzed others these were better in terms of results. In this section we provide the specification for these 2 latent variables. To remind that, the statements for the indicators can be found in Table 6 in the Appendix.

In the latent variable model, structural equations and measurement equations for the attitudes (represented by Att) were built in the form of the equation (10) and (11). The details for the specification is provided for the considered attitudes in the rest of this section.

$$Att = \overline{Att} + \sum_{e} \lambda_e X_e + \omega, \tag{10}$$

where $\overline{Att}$ is estimated mean value of the latent attitude and X_e is a set of explanatory variables. ω is the error term which is assumed to have a Normal distribution with a mean of 0 and a standard deviation of θ_ω.

Measurement equations were built with the most relevant indicators of the latent attitudes in the form of a regression which has a similar structure as the factor analysis in equation (9):

$$I_k = a_k + \alpha_k \, Att + \upsilon_k \qquad \forall k, \tag{11}$$

where a_k and α_k are parameters to be estimated and Att is the latent variable defined by equation (10). The error term is normally distributed with mean 0 and standard deviation θ_{υ_k}.

According to Table 1 it is seen that indicators 16, 17, 22 have high factor loadings for *attitude against public transport* which can be used for measurement equations. For the structural equation, several variables were regressed and number of cars in the household, dummy variables indicating the region in Switzerland and dummy variable for high education level found to be significant. The specification can be represented by the framework in Figure 2. Estimation results say that people with high education have a positive attitude towards public transport, people living in the regions which are the German speaking part of Switzerland has more tendency to use public transport and number of cars in the household increase the negative attitude towards public transport.

Latent variable *environmental concern* has the 4 most explaining indicators of 1, 2, 5, and 6. The explanatory variables are the number of bikes in the household, age which is piecewise linear starting from age of 45, and a dummy variable for the people with high education level (university degree) as seen in Figure 3. It can be concluded that high number of bikes in the household and high education level are signs of a high level of environmental concern. Besides when people get older they start to think more about environmental issues.

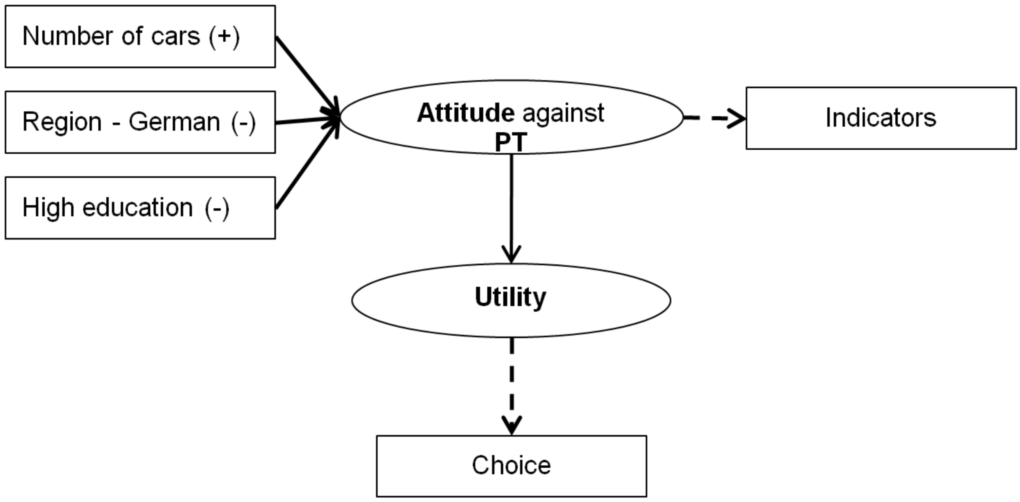

Figure 2: Attitude against public transport

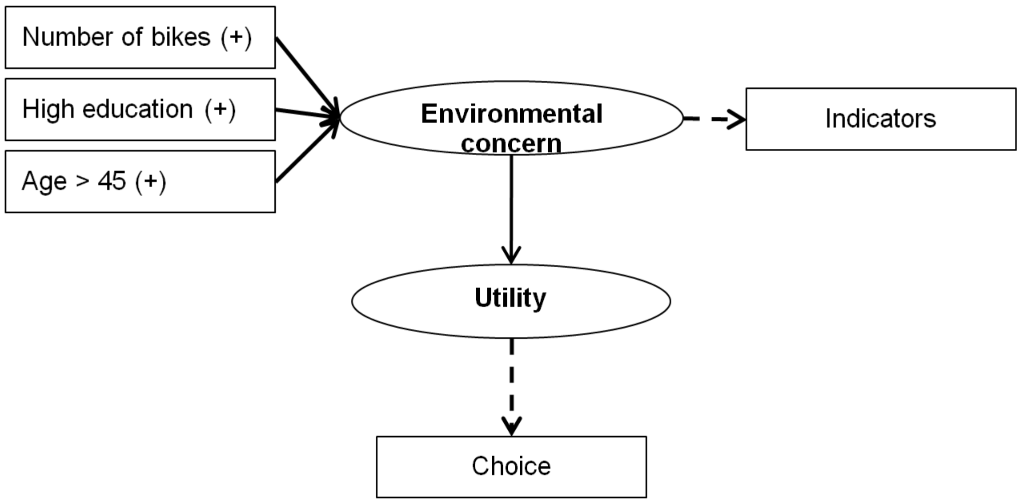

Figure 3: Environmental Concern

4.2 Discrete choice model

A logit model was estimated, choice being the main mode for a cycle of trips, which start and end at the same location. The alternatives for the choice are private mode (PM) representing car, taxi, motorbike and car-sharing, public transport (PT) and soft mode (SM) including walking and bike. Observations with different choices than these three alternatives are excluded being out of scope which results with 1096 observations. Informations for cost and time are inputted according to the information provided by the respondent in the travel diary. Utility function for each alternative was defined as in equations (12)-(14).

$$V_{PM} = ASC_{PM} + \beta_{cost} C_{PM} + \beta_{TT_{PM}} TT_{PM} + \sum_s \beta_s X_s \tag{12}$$

$$V_{PT} = ASC_{PT} + \beta_{cost}C_{PT} + \beta_{TT_{PT}}TT_{PT} + \beta_{freq}F_{PT} + \beta_{attAPt}attAPt + \beta_{attEnv}attEnv \tag{13}$$

$$V_{SM} = ASC_{SM} + \beta_{distance}D_{SM} \tag{14}$$

C_{PM}, C_{PT} are the cost (in CHF), TT_{PM}, TT_{PT} are the travel time in minutes for private mode and public transport respectively and D_{SM} is the distance (in km) for soft mode. In the utility function for car X_s is a vector of explanatory variables including dummy variables for the trip cycles which include work related activities, presence of children and for the location of household (German or French speaking part of Switzerland).

In the utility function for public transport, F_{PT} is the average frequency (number of pairs of trips per day) of the public transport services. *attApt* and *attEnv* are the latent variables of *attitude against public transport* and *environmental concern* respectively which were described in section 3.2 and specified in section 4.1.

The choice model was built assuming extreme value distribution for the error terms associated with the utility functions of the alternatives. Therefore a multinomial logit model is obtained with the following probabilities for choosing each alternative.

$$P_i = \frac{exp(V_i)}{exp(V_{PM}) + exp(V_{PT}) + exp(V_{SM})} \quad i = PM, PT, SM. \tag{15}$$

4.3 Integrated choice and latent variable model

After having defined all the relations likelihood function is obtained by replacing equation (15) and structural and measurement equations for latent variables in equation (5). The estimation is done by using an extended version of the software package BIOGEME Bierlaire (2003) extensions partly being explained in Bierlaire and Fetiarison (2009). As there are two latent variables simulation is used for estimation. Estimation results, including reference results for a multinomial logit model are provided in Table 2.

As seen from the results all the parameters have the expected sign. Time and cost negatively affect private mode and public transport and distance has also a negative effect on the utility of soft mode. Presence of children in the household favors the usage of private mode as expected.

Table 2: Estimation results

	Affected Utility			Integrated Model		Multinomial Logit	
Parameter	V_{PM}	V_{PT}	V_{SM}	Value	t-test	Value	t-test
ASC_{PM}	x			0.157	0.15*	0.81	3.35
ASC_{PT}		x		0**	-	0**	-
ASC_{SM}			x	-0.409	-0.38*	0.218	0.56*
$\beta_{children}$	x			0.492	3.09	0.412	2.62
β_{cost}	x	x		-0.0493	-4.63	-0.0508	-3.91
$\beta_{distance}$			x	-0.221	-4.47	-0.222	-4.44
β_{freq}		x		0.649	3.22	0.701	3.51
β_{w}	x			-0.61	-3.97	-0.622	-4.1
β_{french}	x			1.05	6.22	1.09	6.5
$\beta_{TT_{PM}}$	x			-0.0211	-4.33	-0.0215	-3.83
$\beta_{TT_{PT}}$		x		-0.00847	-3.1	-0.00846	-2.79
β_{attAPt}		x		-0.63	-2.89	-	-
β_{attEnv}		x		0.326	1.89	-	-
$\overline{attAPt}$		x		3.45	54.33	-	-
$\overline{attEnv}$		x		3.04	34.67	-	-
$\lambda_{age>45}$		x		0.00609	2.59	-	-
λ_{bikes}		x		0.0605	4.17	-	-
λ_{cars}		x		0.129	3.52	-	-
$\lambda_{high-educ}$		x		0.262	5.9	-	-
$\lambda_{region3}$		x		-0.307	-3.66	-	-
$\lambda_{region4}$		x		-0.234	-2.02	-	-
$\lambda_{region5}$		x		-0.315	-3.01	-	-
$\lambda_{region6}$		x		-0.193	-2.12	-	-
$\lambda_{region7}$		x		-0.467	-3.01	-	-
a_{16}				0**	-	-	-
a_{17}				0.847	2.28	-	-
a_{22}				1.24	4.4	-	-
a_{1}				-1.77	-2.81	-	-
a_{2}				0.0318	0.07*	-	-
a_{5}				0**	-	-	-
a_{6}				1.06	5	-	-
α_{16}				1**	-	-	-
α_{17}				0.974	7.84	-	-
α_{22}				0.727	7.57	-	-
α_{1}				1.17	6.86	-	-
α_{2}				0.904	7.12	-	-
α_{5}				1**	-	-	-
α_{6}				0.87	15.63	-	-
θ_{attAPt}				-0.469	-6.33	-	-
θ_{attEnv}				-0.492	-5.44	-	-
θ_{16}				-0.255	-5.03	-	-
θ_{17}				-0.126	-3.21	-	-
θ_{22}				0.0171	0.71*	-	-
θ_{1}				0.0873	2.6	-	-
θ_{2}				-0.00741	-0.26*	-	-
θ_{5}				-0.174	-3.94	-	-
θ_{6}				-0.582	-12.87	-	-

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

Trip cycles including work related trips have a negative effect on the utility of private mode which suggests that frequent trips like going to work favor the public transport usage. French speaking part of Switzerland has more tendency to use private mode compared to the German speaking part. Furthermore frequency of the public transport service increases the utility of public transport.

When we look at the sign of the latent variables it is seen that *attitude against public transport* has negative sign and *environmental concern* has positive sign affecting the utility of public transport which is expected. The explanatory variables for the latent variables $\lambda_{age>45}$, λ_{bikes}, λ_{cars}, $\lambda_{high-educ}$, and $\lambda_{region3} - \lambda_{region7}$ have the expected signs as explained in section 4.1. Regions 3-7 (all being in the German speaking part of Switzerland) have better attitude against public transport compared to regions 1 and 2 (in French speaking part), number of cars in the household increase the biased attitude against public transport and high educated people have better image of public transport. For the *environmental concern* number of bikes in the household and high education level have positive effect. Besides, people are more concerned about environmental concepts as they get older.

High education level exists in the structural equations of both latent variables. It is same in absolute value but it enters the structural equation of *attitude against public transport* with a negative sign. Therefore high education level has a negative effect on *attitude against public transport* and a positive effect on *environmental* both of which at the end increase the utility of public transport. These kinds of structures enable us to capture the attitudes of people together with their socioeconomic characteristics.

To be able to see the relation between a_k, α_k and latent attitudes in measurement equations, the term $a_k + \alpha_k\, Att$ (see equation (11)) for each indicator is simulated through the error term of corresponding latent attitude. Furthermore the structural equation (10) of both latent variables are also simulated to find the estimated value for the latent variables (See Table 3). Note that, a_k and α_k are fixed as 0 and 1 respectively for I_{16} and I_5. Since the indicators of *attitude against public transport* are all stating the inconvenience of using public transport in the same direction, the term ($a_k + \alpha_k\, Att$) has a higher value than the value of the latent variable for all indicators. However, for *environmental concern* I_1 and I_2 correspond to the statements which include trade-off between cost (price of gasoline and taxes) and environmental issues as provided in Appendix in Table 6. This trade-off makes people state a lower degree of agreement compared to the statements of I_5 and I_6.

When we compare the results with the multinomial logit model there are not big differences in the parameter values in general since we add the latent variables as constants. However there are some differences like the parameter for the presence of children, $\beta_{children}$, becomes more significant in the integrated model since we consider the attitudes regarding the difficulties of using public transport with children. Furthermore we see that the responsiveness to cost

Table 3: Estimates for latent attitudes and indicators

AttAPt	**2.96**
$a_{16} + \alpha_{16}\, AttAPt$	2.96
$a_{17} + \alpha_{17}\, AttAPt$	3.73
$a_{22} + \alpha_{22}\, AttAPt$	3.40

AttEnv	**3.74**
$a_1 + \alpha_1\, AttEnv$	2.60
$a_2 + \alpha_2\, AttEnv$	3.41
$a_5 + \alpha_5\, AttEnv$	3.74
$a_6 + \alpha_6\, AttEnv$	4.31

slightly decreases with the introduction of latent attitudes in determining the choice. Since both time coefficients remain the same the change in the cost coefficient results with a higher value of time (CHF/h) as seen in Table 4. We also see that constant coefficient for the private mode becomes insignificant in the integrated model which means that we are able to explain the utilities better.

Table 4: Value of time

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

As mentioned at the end of section 3.1 the observations in the data set were weighted according to several statistical information. Time and cost elasticities for private mode and public transport (utility of soft mode does not include the information of time and cost) were calculated as seen in Table 5 considering these weights for each observation. We see that demand for public transport is more elastic in both aspects compared to private mode and people are more elastic for the changes in time than in cost. In particular, demand for private mode is inelastic for cost and for time we can talk about an increase of 0.20% in market share when the time for private mode is decreased by 1%. For public transport there is a potential increase of 0.34% and 0.17% in market share when the time and cost are decreased by 1% respectively. All in all these elasticity values show that demand is not highly elastic for time and cost which supports the idea of introducing latent variables of attitudes in mode choice models.

Table 5: Demand elasticities for time and cost

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

Validation was done by estimating the model on the 80% of the data and predicting the remaining 20%. As a result 66% of the esimated choice probabilities are above 0.5 and 19% are above 0.9 which says that the prediction power of the model is good.

5 Conclusions and Further Research

When latent attitudes are introduced to the mode choice models, it is seen that we are able to have a better understanding of the behavior of users. It is true that these integrated models are far more complicated models than logit models being more tricky in theory. However the methodology provided in this paper enables us to capture unobserved heterogeneity for the choice process with a better forecasting power.

In this paper we have included the latent variables of *attitude against public transport* and *environmental concern* which address the underlying decision making process of people which can not be directly observed. This model gives promising results about the importance of these attitudinal factors in mode choice and motivates us for further improvements. Other potential latent variables are also presented being the candidates for building future models which we believe to have better estimation power. In the presented specification the indicators are treated as continuous variables. However, the scaling for the level of agreement is not necessarily uniform between the levels 1-5. Therefore discrete specification of the indicators is an important further step. Furthermore, the number of indicators for the latent variables is another issue to be investigated.

From a broader point of view, these integrated models enable us to analyze the effect of different latent attitudes and perceptions on the choice process. The structural equations for the latent variables give the possibility to understand the role of different socio-economic characteristics for people with different attitudinal behavior. Therefore the methodology has the potential to provide different segments of population with different attitudes and perceptions combined with explanatory variables.

6 Acknowledgments

The authors would like to thank CarPostal for funding this study and specially Gregor Ochsenbein receives our appreciation by supporting us with his help and feedback. The qualitative survey mentioned was a joint work with the Urban Sociology Laboratory (LASUR) and the Urban and Regional Planning Laboratory (CEAT) at EPFL; special thanks go to Vincent Kaufmann, Martin Schuler and Etienne Doyen. Anne Curchod deserves a special acknowledgment managing many issues regarding the project. The authors would also like to thank Sonia Lavadinho for her help with the survey design and Antonin Danalet for his help with data collection and preparation.

References

Abou-Zeid, M., M. Ben-Akiva and M. Bierlaire (2008) Happines and travel behaviour modification, paper presented at *European Transport Conference (ETC)*, October 2008.

Bagley, M. N. and P. L. Mokhtarian (2002) The impact of residential neighborhood type on travel behavior: A structural equations modeling approach, *The Annals of Regional Science*, **36** (2) 279–297.

Ben-Akiva, M., D. McFadden, T. Garling, D. Gopinath, J. Walker, D. Bolduc, A. Boersch-Supan, P. Delquie, O. Larichev, T. Morikawa, A. Polydoropoulou and V. Rao (1999) Extended framework for modeling choice behavior, *Marketing Letters*, **10** (3) 187–203.

Ben-Akiva, M. E. and B. Boccara (1995) Discrete choice models with latent choice sets, *International Journal of Research in Marketing*, **12**, 9–24.

Bierlaire, M. (2003) Biogeme: a free package for the estimation of discrete choice models, paper presented at *Swiss Transport Research Conference (STRC)*, March 2003.

Bierlaire, M. and M. Fetiarison (2009) Estimation of discrete choice models: extending biogeme., paper presented at *Swiss Transport Research Conference (STRC)*, September 2009.

Doyen, E. (2010) Résultats d'une enquete qualitative avec suivi GPS sur 20 personnes, *Technical Report*, Urban Sociology Laboratory (LASUR), Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Espino, R., C. Roman and O. J.D. (2006) Analyzing demand for suburban trips: A mixed rp/sp model with latent variables and interaction effects, *Transportation*, **33** (3) 241–261.

Hurtubia, R., B. Atasoy, A. Glerum, A. Curchod and M. Bierlaire (2010) Considering latent attitudes in mode choice: The case of switzerland, paper presented at *World Conference on Transport Research (WCTR)*, July 2010.

Kitamura, R., P. L. Mokhtarian and L. Laidet (1997) A micro-analysis of land use and travel in five neighborhoods in the san francisco bay area, *Transportation*, **24**, 125–158.

Ory, D. and P. Mokhtarian (2005) Don't work, work at home or commute? discrete choice model of the decision for san francisco bay area residents, *Technical Report*, Institute of Transportation Studies, University of California, Davis.

Redmond, L. (2000) Identifying and analyzing travel-related attitudinal, personality, and lifestyle clusters in the san francisco bay area, Master Thesis, University of California, Davis, USA, August 2000.

Vredin Johansson, M., T. Heldt and P. Johansson (2006) The effects of attitudes and personality traits on mode choice, *Transportation Research Part A: Policy and Practice*, **40** (6) 507–525.

Walker, J. and M. Ben-Akiva (2002) Generalized random utility model, *Mathematical Social Sciences*, **43** (3) 303–343.

Walker, J. L. and J. Li (2007) Latent lifestyle preferences and household location decisions, *Journal of Geographical Systems*, **9** (1) 77–101.

Appendix:

Table 6: Psychometric Indicators

Ind	Statement	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	We should increase the price of gasoline to reduce congestion and air pollution	-0.540	0.592	-	-	-	-
2	We need more public transport, even if it means higher taxes	-0.252	0.452	-	-	-	-
3	Environmentalism harms the small businesses	0.357	-0.357	-	-	-	-
4	People and employment are more important than the environment	-	-0.435	-	-	-	-
5	I am concerned about global warming	-	0.627	-	-	-	-
6	We must act and make decisions to reduce emissions of greenhouse gases	-	0.581	-	-	-	-
9	I use my travel time productively	-	-	0.271	-	-	-
11	I often reconsider my choice of transport mode	-	-	-	-	-	0.258
14	I am not comfortable when I travel with people I do not know well	0.311	-	-	0.290	-	-
15	Taking the bus helps to make the city more comfortable and welcoming	- 0.224	-	-	-	-	-
16	It's hard to take public transportation when I travel with my children	0.521	-	-	-	-	-
17	It's hard to take public transportation when I travel with bags or luggage	0.613	-	-	-	-	-
18	It is very important to have a nice car	-	-	-	0.622	-	-
20	When I take the car, I know I'll arrive on time	0.385	-	-0.344	-	-	-
22	I don't like to change transport modes when I travel	0.564	-	-	-	-	-
23	If I use public transport instead of my car, I have to cancel some activities	0.562	-	-	-	-	-
24	The bus schedule is sometimes hard to understand	0.378	-	-0.302	-	-	-
25	I know well which bus or train I must take, regardless of where I'm going	-0.269	-	0.665	-	-	-
26	I know the bus schedule by heart	-	-	0.758	-	-	-
28	When I'm in a strange city, I feel very disoriented	-	-	-0.323	-	-	-
29	I use the Internet for schedules and departure times of buses or trains	-	-	0.295	-	-	-
30	I have used public transport all my life	-	-	0.846	-	-	-
32	I know some of the drivers of the buses I take	-	-	0.430	-	-	0.615
33	I find it important to talk with the drivers of public transport	-	-	-	-	-	0.750
34	I like living in a neighborhood where there are a lot of things to do	-	-	-	-	0.425	-
35	The terms of accessibility and mobility are important in choosing a home	-	-	-	-	0.274	0.231
38	I would like to live in the center of a big city	-	-	-	-	0.684	-
39	I would like to live in a commune on the outskirts of a city	-	-	-0.254	-	0.421	-
40	I would like to live in the countryside	-	-	-	-	-0.639	-
46	I can identify the status of people by looking at their car	-	-	-	0.315	-	-
47	The pleasure of having something beautiful is to show	-	-	-	0.551	-	-
48	For me, the car is just a convenient way to move	-	-	-	-0.509	-	-
50	I do not like staying at the same place for a long time	0.209	-	-	-	-	-