Mode choice with attitudinal latent class: a Swiss case-study

Bilge Atasoy * Aurélie Glerum * Michel Bierlaire *

Second International Choice Modeling Conference, July 2011

*Transport and Mobility Laboratory (TRANSP-OR), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, {bilge.kucuk, aurelie.glerum, michel.bierlaire}@epfl.ch

Abstract

Attitudinal attributes play an important role in behavior of individuals in various contexts. In this study we focus on the travel behavior and our main objective is to come up with segments of individuals that have different mode choice preferences with the help of the attitudinal attributes. These segments are important for the design of more efficient public transport policies specific to the characteristics of different customers. In order to identify the segments of individuals, most importantly the potential users of public transport, factor analysis techniques are utilized with the attitudinal attributes and the socio-economic characteristics of individuals. The learnings from this exploratory analysis are exploited in the construction of the class-membership model in our hybrid choice model. Maximum likelihood estimation is done simultaneously for the latent class model, including the measurement equations for the psychometric indicators, and the class-specific choice models. The results for the presented model with two latent classes show that middle-aged individuals with high income who are active in their professional and social life have higher value of time and are less elastic to the changes in the transport offer compared to the rest of the population.

1 Introduction

Transport mode choice behavior is explained with various variables including the socio-economic characteristics of individuals, modal attributes and more recently psychometric indicators that express individuals' attitudes, perceptions, and lifestyle preferences. These psychometric indicators are integrated into choice models in order to measure the latent characteristics that drive the travel behavior. In this study our aim is to capture the latent segmentation of individuals in order to identify different mode choice preferences and design public transport policies more efficiently.

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 individuals in peri-urban and suburban 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.

Latent class choice models are studied in literature in various contexts. As a review on the methodology of generalized random utility models including the latent class models we refer to Walker and Ben-Akiva (2002) and Ben-Akiva et al. (2002). Latent segmentation is utilized in order to capture different taste parameters, choice sets, and decision protocols. Ben-Akiva and Boccara (1995) study mode choice behavior of commuters and allow different choice sets for different segments of the population. Gopinath (1995) presents latent class models for mode choice behavior and shows that segments of populations have different decision protocols for the choice process as well as different sensitivities for the time and cost. Furthermore Gopinath (1995) integrates attitudinal indicators in the framework such that latent class model includes the measurement model for these latent class indicators. Hosoda (1999) works on the mode choice models for shopping trips where continuous latent variables are included in the latent class membership model and the latent variables are measured with attitudinal indicators.

More recently Walker and Li (2007) study lifestyle preferences to form latent classes of individuals that have different residential location choice. They estimate the class membership and choice models simultaneously and they provide future directions in including psychometric indicators in the latent class framework as additional measurements for the class membership. Wen and Lai (2010) explore a latent market segmentation for international airline passengers' preferences using stated preference data and latent class model.

Latent class modeling with psychometric indicators is a fruitful area with potential improvements in the explanatory power of choice models. However the identification of latent classes needs to be paid attention in terms of many aspects including identification issues resulting from missing data, determining the number of classes to be used, the naming of the latent classes according to the characteristics of the belonging individuals, the homogeneity of the latent classes etc. All these issues are addressed by Collins and Lanza (2010) with examples.

In this paper we present a latent class choice model where class membership and class specific choice models are estimated simultaneously and psychometric indicators are included in the maximum likelihood estimation to strengthen the model. The psychometric indicators are related to attitudes of individuals regarding modes of transportation. For the preliminary analysis regarding the same project we refer to Hurtubia et al. (2010) and Atasoy et al. (2010) where latent variables are used to better explain the travel behavior with the attitudinal attributes.

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

2 Latent class choice model

The presented latent class choice model, which is strengthened by psychometric indicators, is based on the framework provided by Walker and Ben-Akiva (2002) who present the generalized random utility model with extensions of latent variables and latent classes. The framework we use consists of two components: a latent class model and class-specific discrete choice models, each having its own set of measurement and structural equations. The latent class model includes a class membership model and the measurement equations for psychometric indicators. In the framework, unobserved variables are represented by ellipses and observable variables by rectangles. Besides, dashed lines correspond to the measurement equations and straight lines represent the structural equations as in Figure 1.

Latent classes are not defined beforehand so we can not deterministically relate the individuals to the segments. However the estimation of the class membership probabilities is done jointly with the choice model with the help of the observed characteristics of the individuals. In this paper the class membership model is specified as a logit formulation. The probability of individual n belonging to class s given explanatory variables X_n is given by equation (1) where the parameters γ are to be estimated.

$$P(s|X_n;\gamma) \tag{1}$$

Since our assumption is that latent segments of the population show different mode choice behavior, the utility functions are defined separately for each class. Therefore utility of alternative i for individual n in latent class s is given as in equation (2) where

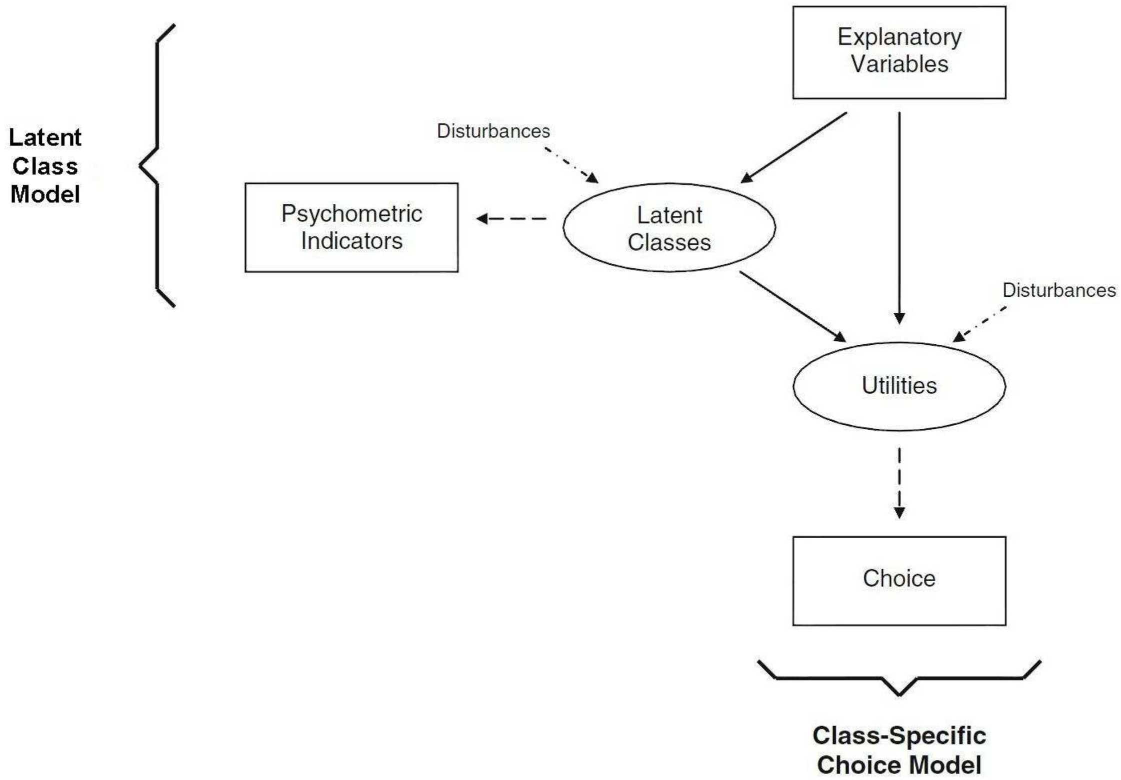

Figure 1: Framework for latent class choice model with indicators (Walker and Li, 2007)

it is explained by the modal attributes X_i and the individual characteristics X_n. β^s is a set of parameters that can be defined specific to each class s and ϵ_{in}^s is an error term.

$$U_{in}^s = V(X_n, X_i; \beta^s) + \varepsilon_{in}^s \tag{2}$$

Having defined the utility function, class-specific choice probability for alternative i is given by:

$$P(i|X_n, X_i, s; \beta^s, \theta_\varepsilon^s) = Prob[U_{in}^s \geq U_{jn}^s, \forall j \in C_s], \tag{3}$$

where θ_ε^s is the standard deviation of the error term in equation (2) and C_s is the choice set for class s. The choice model is also specified as a multinomial logit model.

As we mentioned previously, the latent class choice model includes the psychometric indicators such that the measurement model for indicators is simultaneously estimated with the class membership model and the choice models. The probability of an individual n giving a particular response I_n to an indicator conditional on being in latent class s is given by:

$$P(I_n|s). \tag{4}$$

This probability, which is called *item-response probability*, is defined as a parameter to be jointly estimated with the choice and class membership models. When we put all the models together, the joint probability of observing choice i and indicator I_n is given by:

$$P(i, I_n|X_n, X_i; \beta, \gamma, \theta_\varepsilon) = \sum_{s \in S} P(i|X_n, X_i, s; \beta^s, \theta_\varepsilon^s) P(I_n|s) P(s|X_n; \gamma). \tag{5}$$

Maximum likelihood estimation is used to estimate the unknown parameters. The likelihood function (L) can be written as in equation (7) with the definition of y_{in}^s for the measurement of choice as in equation (6). The choice set is class specific since the two classes have different available alternatives as explained in section 4.2.

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

$$L = \sum_{s \in S} \prod_{n} \prod_{i \in C_s} P(i, I_n | X_n, X_i; \beta, \gamma, \theta_\varepsilon)^{y_{in}^{s}} \tag{7}$$

3 Data Collection and Factor Analysis

The data collection campaign consists of three surveys in the area of study (non-urban areas served by Car Postal). First, a qualitative survey (informal interviews) is performed to have an insight about the behavior of individuals which is valuable for the construction of the subsequent surveys. The second survey is a revealed preferences (RP) regarding the 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 situations with proposed new alternatives, which were designed in the light of the gained insight in the first two surveys. This section focuses on the identification of latent classes by making use of the individual characteristics and the 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) 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 devise with them for seven days, recording all their movements. The geocoded results were shown afterwards to the respondents, where they identified the transport modes and the 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. For the details of the qualitative survey and the results we refer to Doyen (2010).

The main use of this survey considering this paper is to come up with the attitudes and perceptions of individuals that play an important role in the mode choice process. To reveal these attitudes and perceptions a list of psychometric indicators are selected which helps in the identification of latent classes. This list of psychometric indicators are determined based on the examples found in the literature: Kitamura et al. (1997), Redmond (2000), Ory and Mokhtarian (2005), Vredin Johansson et al. (2006). 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 the trade-off statements to be able come up with a reasonable set of data.

For example, the following indicators are revealing respondent's attitudes towards environmental issues:

- *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 the 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 the statements in a similar way. A list of 54 statements with a 5 level likert scale to indicate the level of agreement (from strong disagreement to strong agreement) is generated and included in the RP survey. The 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. The travel diary is the skeleton of this dataset where the respondents registered all the trips performed during a specified day. The collected information consists of the origin, destination, cost, travel time, chosen mode and the activity at the destination.

The model presented in this paper is based on the data from the RP survey, which resulted with a collection of 1763 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 2265 observations relating the sequences of trips, psychometric indicators and socioeconomic attributes.

The resulting sample has some socio-demographic categories that were oversampled such as individuals with a high education level, middle-aged and male respondents. In Table 1 we present the statistics for the corresponding socio-economic characteristics compared to the data of the Federal Census of 2000. Therefore the observations in this dataset are weighted according to the representation of the categories related to education level, age and gender. These weights for the individuals are used for having a weighted average value for the outputs of the model such as market shares, cost and time elasticities of demand in order to better represent the Swiss population.

Category		Sample	Population
Education			
	University	14.2%	6.2%
	Vocational university	16.2%	10.6%
	Certificate of Vocational Training and Education	61.0%	50.9%
	Compulsory school	7.6%	27.6%
	No school diploma	1.0%	4.7%
Age			
	16-19 years	2.3%	8.2%
	20-39 years	21.2%	33.4%
	40-64 years	55.9%	41.6%
	65-79 years	18.7%	12.7%
	80 years and above	1.8%	4.1%
Gender			
	Male	53.0%	49.0%
	Female	47.0%	51.0%

Table 1: Statistics on socio-demographic categories

3.2 Factor Analysis

Before the construction of the latent class choice model, an exploratory analysis is carried out to identify the potential segmentation of the population in terms of mode choice. With this purpose, factor analysis method is applied to a subset of variables which are found to be relevant. Although factor analysis does not provide distinct classes for us as clustering techniques, it helps to understand which characteristics or attitudes of individuals are strong in determining the groups. For the analogies between latent class modeling and factor analysis we refer to the work of Collins and Lanza (2010).

Factor analysis is a method that is used to explain the variability among the observed variables by assigning them to a lower number of unobservable variables. The relation between the observed variables (x_k) and the unobserved factors (F_j) is given by the following equation:

$$x_k = \overline{x_k} + \sum_j \rho_{kj} F_j + \varphi_k, \tag{8}$$

where $\overline{x_k}$ is the mean value of the answer for observed variable k and φ_k is an error term following a normal distribution. The factor loadings (ρ_{kj}) quantify the correlation between the observed variable and unobserved factors. If the factor loading is close to 1 in absolute value it means that there is a strong relation.

Our broader aim is to identify the potential users of public transport and it is found to be interesting to have 2 latent classes. Firstly, psychometric indicators are studied and the ones explaining a bigger variance are afterwards analyzed together with the individual characteristics and the mode choice. We have also included the choice variable in the analysis to be able to see the relation between the mode choice behavior of individuals and their socio-economic characteristics as well as their opinions that are revealed by the psychometric indicators.

There are 3 indicators in the presented factor analysis which were found to be most relevant ones improving the estimation results of the integrated model. These are listed as follows:

- **Ind1 - PT children:** *It's hard to take public transportation when I travel with my children.*
- **Ind2 - Flexibility car:** *With my car, I can go where I want when I want.*
- **Ind3 - Family oriented:** *I would like to spend more time with my family and friends.*

Preliminary analysis showed that the family attributes of individuals play an important role together with their income level and some other socio-economic characteristics. The factor loadings can be seen in Table 2 where the ones with an absolute value higher than 0.1 are presented. In the literature cut-off value is usually taken as 0.25 and therefore the loadings higher than 0.25 are presented as bold indicating the most explaining variables. Looking at the results, the following segments can be identified with their most outstanding characteristics:

- **Class 1 - Independent**: Middle-aged respondents that live with their family and children and who have high income.
- **Class 2 - Dependent**: Young individuals who are mostly students and old respondents. This class of individuals are single people or couples without children and they travel more with public transport compared to the first class.

Table 2: Results of factor analysis

	Factor 1	Factor 2
Choice PT		**0.250**
Socio-economics		
$N_{children}$	**0.517**	
Student/trainee	0.117	**0.770**
N_{cars}	0.203	
HighIncome	**0.252**	
Education		-0.123
Age ≥ 60	**-0.375**	
Family status		
Couple without children	**-0.606**	
Couple with children	**0.927**	**-0.368**
Living with parents	0.159	**0.956**
Single	**-0.371**	
Single parent	-0.170	
Roommate	-0.142	
Psychometric Indicators		
Ind1 - PT children		
Ind2 - Flexibility car		-0.130
Ind3 - Family oriented	0.135	

It is seen that the individuals in the first group want to spend more time with their family and friends compared to the second group of individuals. This is an expected result looking at their characteristics. Similarly, the first segment of individuals rely more on the car alternative compared to the second segment. We call the first segment as *independent* and the second as *dependent*. The idea behind this naming is that the second group of individuals are either very young and student which makes them economically dependent or they are old which limits their physical activity. Although we do not see a strong factor loading for the indicator *PT children*, our decision was to include this indicator since it is observed that it has a significant role in the segmentation as explained in section 4.3 (See Table 4).

4 Model Specification and Estimation Results

As described in section 2 the integrated model has two parts: a latent class model and class-specific discrete choice models. Latent classes are explained with the individual characteristics through a class membership model. The latent class model also includes the measurement model for the psychometric indicators. In the discrete choice part, the utility of the alternatives are explained by the modal attributes and the socioeconomic characteristics separately for each class.

4.1 Latent class model

Latent class model includes a class membership model and class-specific measurement models for psychometric indicators. The class membership model determines the probability of belonging to a class through the structural equations. The measurement equations for the indicators help to identify the latent classes by making use of the responses given to the indicators.

With the learnings from the exploratory analysis (section 3.2) the structural equations are built for the two latent classes and the probability of belonging to each class is defined with a binary logit formulation. The class membership model is given by:

$$V_{Independent} = ASC_{in} + \alpha_{family} \text{ Family} + \alpha_{income} \text{ HighIncome}, \tag{9}$$

$$V_{Dependent} = \alpha_{single} \text{ Single}, \tag{10}$$

where *Family*, *HighIncome*, and *Single* are all dummy variables. *Family* is equal to 1 if the individual is living with his/her children (couple with children and single parent), *High Income* is 1 if the household income is high, and *Single* is 1 if the person lives either alone or with parents. Although there are other characteristics defining the classes as seen in Table 2, the tested models with these variables were not successful.

As mentioned in section 3.2 we work with the indicators of *flexibility car*, *family oriented*, and *PT children*. The measurement equations for these indicators are written by defining the item-response probabilities as parameters. Therefore the probability of individual n in latent class s responding r to an indicator, $P(I_n = r|s)$ for $r = 1, ..., 5$, is estimated with the help of the observed responses.

4.2 Class-specific discrete choice model

Mode choice in the context of this study is considered as the main mode for a cycle of trips, since each observation corresponds to a cyclic sequence of trips. The alternative modes 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 1906 observations.

For the two latent classes, *independent* and *dependent*, a specific mode choice model is developed. The time and cost parameters are defined specifically and different socio-economic attributes are included in the two classes. These allow to capture the taste heterogeneity in the population. Furthermore for the class *independent* we have all the 3 alternatives available. However the individuals belonging to class *dependent* does not have the soft mode alternative. For the whole dataset, the percentage of respondents using the soft mode as the main mode for their cycle of trips is low (5%). Therefore the second class, which includes the old people as well, did not allow the inclusion of the soft mode. Actually, it makes sense to assume that individuals belonging to class *dependent* do not consider the soft mode.

The utility functions of the first latent class, *independent*, for the alternatives of private mode, public transport and soft mode are given by the equations (11), (12), and (13) respectively. Utilities for private mode and public transport include the travel time (TT_{PM}, TT_{PT}) and the cost variables (C_{PM}, C_{PT}) as modal attributes. Apart from the modal attributes, the utility for private mode includes other explanatory variables. N_{cars} is the number of cars in the household, $N_{children}$ represents the number of children in the household, *French* is a dummy variable which is 1 if the respondent is living in the French speaking part of Switzerland, and *WorkTrip* is a dummy variable which is 1 for the work related cycle of trips. For public transport there is a dummy variable *Urban* which is 1 if the respondent lives in a commune that is characterized as being urban. Finally the utility of soft mode is explained by the distance of the trip (D_{SM}) and the number of bikes in the household (N_{bikes}).

$$\begin{aligned} V^1_{PM} = ASC^1_{PM} + \beta^1_{cost} C_{PM} + \beta^1_{TT_{PM}} TT_{PM} + \beta_{N_{cars}} N_{cars} \\ + \beta_{N_{children}} N_{children} + \beta_{language} French + \beta_{work} WorkTrip \end{aligned} \tag{11}$$

$$V^1_{PT} = \beta^1_{cost} C_{PT} + \beta^1_{TT_{PT}} TT_{PT} + \beta_{urban} Urban \tag{12}$$

$$V^1_{SM} = ASC^1_{SM} + \beta_{distance} D_{SM} + \beta_{N_{bikes}} N_{bikes} \tag{13}$$

The latent class *dependent* has two modes in the choice set. For both private mode and public transport we have the time and cost variables as for the first latent class of individuals (see the equations (14) and (15)). In addition to that for public transport there is a dummy variable *Student* which is 1 if the respondent is a student or a trainee.

$$V^2_{PM} = ASC^2_{PM} + \beta^2_{cost} C_{PM} + \beta^2_{TT_{PM}} TT_{PM} \tag{14}$$

$$V^2_{PT} = \beta^2_{cost} C_{PT} + \beta^2_{TT_{PT}} TT_{PT} + \beta_{student} Student \tag{15}$$

The class-specific choice models are built assuming extreme value distribution for the error terms associated with the utility function of the alternatives. Therefore, for each class s, a multinomial logit model is obtained for the probabilities of choosing each alternative i as in equation (16).

$$P^s_i = \frac{\exp\left(V^s_i\right)}{\sum_{j \in C_s} \exp\left(V^s_j\right)} \tag{16}$$

4.3 Estimation results for the integrated latent class choice model

Having specified all the structural and measurement equations, the likelihood function in equation (7) is built and the parameters are estimated with maximum likelihood estimation. For the estimation, the extended version of the software package BIOGEME (Bierlaire (2003)) is used, extensions partly being explained in Bierlaire and Fetiarison (2009).

The estimation results can be seen in Table 3. First part in the table corresponds to the class membership model. The results support our discussion in section 3.2 such that for the first class *independent*, the parameter for *Family* is positive as well as the parameter for *HighIncome*. On the other hand for the second class we have a positive parameter for the dummy variable *Single*.

In the second part of the table we have the results for the class-specific choice models. All the parameters have the expected signs. The travel time and cost negatively affect the utilities of private mode and public transport for both classes and distance also has a negative effect on the utility of soft mode for the latent class *independent*.

Given that an individual is in class *independent*, the utility of private mode increases with the number of cars and the number of children in the household. The respondents in the French speaking part of Switzerland have more tendency to use private mode and the individuals living in urban communes use public transport more compared to others. When the trip is done for work purposes the utility of private mode decreases. Moreover, the number of bikes in the household is positively correlated with the utility of soft mode. When we look at the class *dependent* the utility of public transport is higher for the students and individuals in internship.

Table 3: Estimation results

	Affected utility					**Class**		**Estimated results**	
Parameter	V^1_{PM}	V^1_{PT}	V^1_{SM}	V^2_{PM}	V^2_{PT}	1- Indep.	2- Dep.	Value	t-test
ASC_{in}						x		-0.642	-3.470
α_{family}						x		4.010	4.570
α_{income}						x		0.479	2.110
α_{single}							x	0.861	3.880
ASC^1_{PM}	x							-1.100	-3.750
ASC^2_{PM}				x				0.274	1.210*
ASC^1_{SM}			x					0.529	1.320*
β^1_{cost}	x	x						-0.0229	-2.520
β^2_{cost}				x	x			-0.321	-4.720
$\beta^1_{TT_{PM}}$	x							-0.0159	-2.740
$\beta^2_{TT_{PM}}$				x				-0.108	-5.280
$\beta^1_{TT_{PT}}$		x						-0.00643	-2.190
$\beta^2_{TT_{PT}}$					x			-0.0433	-4.810
$\beta_{distance}$			x					-0.200	-3.680
$\beta_{N_{cars}}$	x							1.340	8.050
β_{urban}		x						0.366	2.200
β_{work}	x							-0.807	-4.970
$\beta_{language}$	x							1.320	6.500
$\beta_{N_{children}}$	x							0.395	4.800
$\beta_{N_{bikes}}$			x					0.206	3.460
$\beta_{student}$					x			3.820	5.100

(* Statistical significance $<$ 95%)

In Table 4 we provide the jointly estimated item-response probabilities for the three indicators that are described in section 3.2. To remind that when the response is 5 it means that the respondent strongly agrees with the indicator statement. The results for the first indicator related to children show that, since the individuals in the second group do not have children they are neutral (level 3) to the statement which plays an important role in distinguishing the classes of individuals. When we look at the second indicator, the probability for agreeing (levels 4 and 5) with the statement of car being flexible is higher for the first group of respondents. When we interpret the results for the third indicator, it is seen that the first class of individuals wants to spend more time with their family and friends having higher probabilities for the response levels 4 and 5. The difference between the classes is more obvious when we put together the responses of 1 and 2 under the name of *No*, 4 and 5 under the name of *Yes* and keep 3 as the neutral response (-) as seen in Figure 2.

As pointed out in section 3.1 the observations in the dataset are weighted according to the statistical information of gender, education and age. Therefore the generated results are obtained by having a weighted average of all the observations to increase representativity of our results. First of all the aggregate class membership probabilities,

Table 4: Item-response probabilities for indicators

	Class 1			Class 2		
Response	Ind1	Ind2	Ind3	Ind1	Ind2	Ind3
1	0.161	0.0314	0.0136	0.0128	0.0184	0.00430*
2	0.241	0.0322	0.0471	0.0192	0.0282	0.0394
3	0.318	0.120	0.255	0.929	0.170	0.411
4	0.174	0.370	0.488	0.0337	0.366	0.435
5	0.106	0.446	0.196	0.00530	0.417	0.111

(* Statistical significance < 90%)

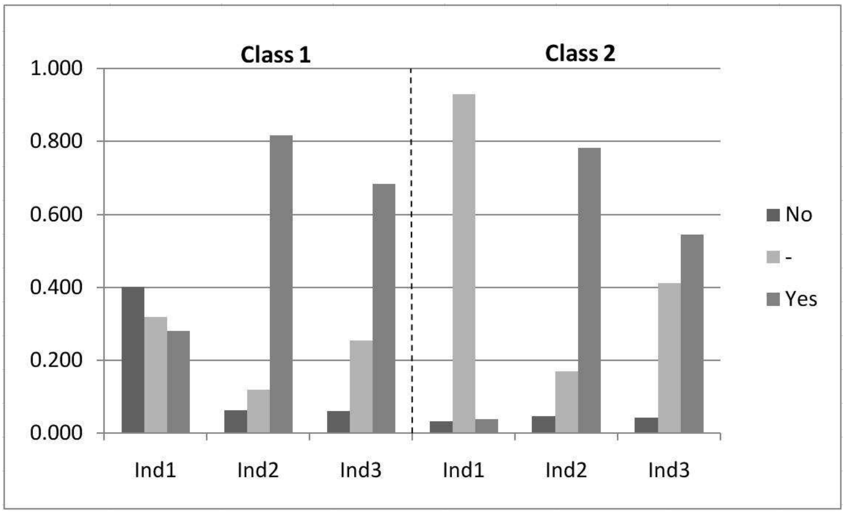

Figure 2: Estimated item-response probabilities

also called *prevalence*, are calculated as seen in Table 5. Individual class membership probabilities are used to obtain aggregate values for demand elasticities, market shares etc.

Table 5: Class membership probabilities

	Class membership
Class 1: Independent	53.72%
Class 2: Dependent	46.28%

The demand elasticities for time and cost are calculated for the two latent classes as well as at an aggregate level, by considering the weights for the observations as displayed in Table 6. The first thing we see is that the demand for public transport is more elastic in both aspects compared to private mode. Individuals belonging to the class *independent* are less elastic to time and cost for both modes which is expected due to their characteristics. We also observe that the demand is more elastic to the changes in time compared to cost. When we interpret the results we can make the following evaluations. When the time and cost is decreased by 1%:

- there is a potential increase of 0.17% and 0.03% in the demand for private mode respectively for class *independent*,
- there is a potential increase of 0.34% and 0.12% in the demand for public transport repectively for class *independent*,

- there is a potential increase of 0.41% and 0.15% in the demand for private mode respectively for class *dependent*,
- there is a potential increase of 0.90% and 0.48% in the demand for public transport respectively for class *dependent*.

Table 6: Demand elasticities for time and cost

		Time elasticity	Cost elasticity
Class 1: Independent	PM	-0.17	-0.03
	PT	-0.34	-0.12
Class 2: Dependent	PM	-0.41	-0.15
	PT	-0.90	-0.48
Aggregate	PM	-0.28	-0.09
	PT	-0.59	-0.28

With the developed model we can also make an analysis on the market shares of different modes. In Table 7 we present the market shares for private mode, public transport and soft mode. These values are also normalized with the weights that are calculated for each observation. We see that the market share for public transport is relatively higher for the class *dependent*. Overall market shares are estimated as 63%, 32%, and 5% for private mode, public transport, and soft mode respectively. The market shares for our data sample corrected with the weights are also presented in the table.

Table 7: Market shares

	PM	PT	SM
Class 1: Independent	61.23%	29.17%	9.60%
Class 2: Dependent	67.73%	32.27%	-
Aggregate	62.70%	32.35%	4.95%
Weighted sample	61.30%	33.95%	4.74%

Value of time (VOT) analysis is an important indicator of travel behavior models which gives the willingness to pay of the respondents in case of a reduction in the travel time by one hour. With the estimated time and cost parameters, we obtain meaningful value of time (VOT) values for the considered transport modes as seen in Table 8. VOT is higher for the first class of individuals as expected and it is higher for private mode compared to public transport for both classes. We can calculate the aggregate VOT with the help of class membership probabilities which results with 32 CHF/h and 13 CHF/h for private and public modes respectively.

Table 8: Value of time

	VOT_{PM} (CHF/h)	VOT_{PT} (CHF/h)
Class 1: Independent	41.66	16.85
Class 2: Dependent	20.19	8.09
Aggregate	31.72	12.80

Validation of our latent class choice model is done by re-estimating the model on the 80% of the data and predicting the remaining 20%. This split of the dataset is performed randomly. As a result, for the first latent class (*independent*) 69% of the esimated choice probabilities are above 0.5 and 19% are above 0.9. For the second class (*dependent*) the same values are 71% and 19% respectively which says that the prediction power of the presented model is good.

5 Conclusions and Further Research

While the attitudinal attributes are being studied in literature in the context of latent variable models, usage of this attributes in the construction of latent segments in the population is not studied widely. In this paper we present a latent class choice model where we integrate psychometric indicators in the framework. The latent segmentation we obtained supports our assumption such that the population is heterogeneous in terms of the mode choice behavior. The resulting segmentation consists of two segments, first of which is the class *independent* representing the individuals who are mostly middle-aged, living with family and children, who have high education and high income. The second class, *dependent*, corresponds to the young individuals who are mostly students and trainees as well as old people who are mostly retired.

The purpose of the latent class analysis is to design more appropriate strategies and policies for different segments of the population. For our case, it is clear that the actions to be taken to improve the public transport offer should be different for the two latent segments. The individuals in class *independent* are not elastic to cost and they need solutions for their family related characteristics, specifically their children. They find using car flexible and therefore to attract them public transport services should be designed in a more flexible way. On the other hand the second class of individuals are more elastic to cost which is already considered in most of the public transport services where there are discounts for young and old people.

The presented model is an evidence for the potential improvement in the travel behavior models with the latent segmentation that includes the psychometric indicators. We believe that the investigation of models with more than two latent classes is a promising future direction which allows more specific policies for different groups of individuals. Furthermore, inclusion of more psychometric indicators is believed to help in the identification of the latent segments and to increase the forecasting power of the models. The interpretation of the presented results give concrete ideas for the actions to be taken to increase the market share of public transport and the future steps are believed to strengthen the interpretability of models.

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 Ricardo Hurtubia for his work on the design of the survey and the preliminary analysis, Sonia Lavadinho for her help with the survey design, Antonin Danalet for his help with data collection and preparation, and Lidija Stankovikj

who initiated the work on clustering techniques and latent class models in the context of her semester project.

References

Atasoy, B., Hurtubia, R., Glerum, A. and Bierlaire, M. (2010). Demand for public transport services: Integrating qualitative and quantitative methods, *Swiss Transport Research Conference (STRC)*.

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

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A. and Munizaga, M. A. (2002). Hybrid choice models: Progress and challenges, *Marketing Letters* **13**(3): 163–175.

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

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

Collins, L. and Lanza, S. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*, first edn, Wiley Series in Probability and Statistics, Boston.

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.

Gopinath, D. A. (1995). *Modeling heterogeneity in discrete choice processes: Application to travel demand*, PhD thesis, Massachusetts Institute of Technology.

Hosoda, T. (1999). *Incorporating unobservable heterogeneity in discrete choice model: Mode choice model for shopping trips*, PhD thesis, Massachusetts Institute of Technology.

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

Kitamura, R., Mokhtarian, P. L. and Laidet, L. (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 Mokhtarian, P. (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's thesis, University of California, Davis, USA.

Vredin Johansson, M., Heldt, T. and Johansson, P. (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 Ben-Akiva, M. (2002). Generalized random utility model, *Mathematical Social Sciences* **43**(3): 303–343.

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

Wen, C.-H. and Lai, S.-C. (2010). Latent class models of international air carrier choice, *Transportation Research Part E* **46**: 211–221.