Incorporating latent variables into discrete choice model, the case of Swiss data set

Author:

Aliasghar Mehdizadeh Dastjerdi

Assignment:

Final project for discrete choice model course

1 Introduction

Discrete choice analysis has become a standard method to analyze travel choice behaviors, such as travel mode, residential and activity locations, car ownership etc. Different alternative discrete choice models have been developed to represent a behaviorally more realistic choice process. Nevertheless, a complete understanding and explanation of a decision-maker's choice behavior is not straightforward since there might be a substantial number of difficult to measure determinants influence the choice process. Decision-makers' attitudes, perceptions, beliefs and values are some examples of such determinants. It is not easy to identify these unobservable factors and the causal relationships among them. In this context, the hybrid choice model (HCM) is used to identify and incorporate them into a discrete choice analysis. HCM is an extended discrete choice modeling framework, which integrates different types of models into a single structure that is estimated concurrently. More specifically, HCMs include a latent variable model into a discrete choice model as to improve the explanatory power of the choice model by considering the effects of decision makers' latent attitudes.

2 Literature review

To date, a large body of studies has employed HCMs to examine the effects of people's latent attitudes on different travel choice behaviors. Table 1 shows a summary of the previous research. The findings and results of the research are not presented in Table1 since they are not of interest for the objectives of this study.

Table 1. Overview of previous studies applying the HCMs to investigate travel choice behaviors

Author	**Objective**	**Application Subject**	**Types of survey Choice set**	**Latent Variables**
Temme et al. (2008)	To investigate the influence of abstract motivations and attitudes on travel mode choice	Travel mode	RP (1) PT, (2) car +PT, (3) PT only	Flexibility, convenience and comfort, safety, power, hedonism, security
Popuri et al. (2011)	To study the influence of individual attitudes and perceptions about transportation on transport choices	Travel mode	RP (1) PT , (2) Auto	Need for reliable & stress free commute, need for privacy & comfort, dynamic work schedule & trip complexity, tolerance to walking & waiting, importance of PT, perceived safety of travel ambience
Kim et al. (2012)	To understand the role of travelers uncertainty for mode choice behavior	Travel mode	SP (1) auto, (2) train, (3) water transit	Environmental preferences Dislike of driving a car Water transit preference
Prato et al. (2012)	To identify of determinants of route choice behavior by integrating latent variables into the model	Travel route	RP 2 ~ 19 alternative routes per observation (generated by the branch and bound algorithm)	Mnemonic ability, habit within the choice environment, familiarity with the choice environment, spatial ability, time saving skill
Daly et al. (2012)	To investigate the effect of security on rail travel behavior.	Rail security system	SP 4 options including "not to choose"	Increased concern Reduced distrust
Atasoy et al. (2013)	To study the influence of attitudes towards mode choice behavior	Travel mode	RP (1) private motorized modes, (2) PT, (3) soft modes	Pro-car Environmental concern
Daziano & Bolduc (2013)	To investigate the effect of environmental concern on fuel vehicle type purchase decisions	Vehicle fuel type	SP (1) standard gas vehicle, (2) alternative fuel vehicle (3) hybrid vehicle, (4) hydrogen fuel cell vehicle	Environmental consciousness (in the context of Transport policy support & Transport problem evaluation)
Glerum et al. (2013)	To forecast the demand for electric vehicles considering attitudes and perceptions	Vehicle fuel type	SP (1) Competitors-gasoline, (2) Renault-gasoline, (3) Renault electric	Pro-leasing attitude Pro-convenience attitude (e.g., spaciousness, technology, etc.)
Mabit et al. (2014)	To examines how reference-dependent preferences and attitudes together explain individual choices for purchasing alternative-fuel vehicles	Vehicle fuel type	SP (1) conventional vehicle, (2) hybrid non-plugin vehicle, (3) bio-diesel vehicle, (4) electric vehicle	Appreciation of car features (joy of driving, comfort, design, etc.)

Author	Objective	Application Subject	Types of survey Choice set	Latent Variables
Soto et al. (2014)	To predict the consumer behavior in new vehicle purchases when alternative fuel technology vehicles are available.	Vehicle fuel type	SP (1) standard gasoline vehicle, (2) natural gas vehicle, (3) hybrid electric vehicle, (4) electric vehicle or diesel vehicle	Support transport policies Environmental concern, Attitudes towards technology A attitudes towards car
Paulssen et al. (2014)	To examine the influence of personal values and individual attitudes towards different alternative attributes on travel mode choice	Travel mode	RP (1) Transit, (2) auto	Values : Security, Hedonism , Power Attitudes: Comfort and convenience, Ownership, Flexibility
Kamargianni et al. (2014)	To identify the social interaction effects between teenagers and their parents regarding walking-loving behavior and then the effect of this on mode to school choice behavior	Travel mode	RP (1) walk, (2) bus, (3) car	Social environment : (Parents-Walking lovers) Decision makers: (teenagers-Walking lovers)
Kamargianni et al. (2015)	To investigated the subjective and objective factors influencing teenagers' school travel mode choice	Travel mode	SP (1) car, (2) Powered two wheelers, (3) bus, (4) walk (5) bike	Safety Consciousness, Green Lifestyle Physical Propensity
Scagnolari et al. (2015)	To analyze the mobility preferences of young people at nighttime by incorporating alcohol-related psychological as the latent variable	Travel mode	SP (1) car driver, (2) motorcycle, (3) car passenger,(4) bus/train, (5) mini-bus, (6) shared taxi (7) no-choice	Attitude toward alcohol consumption
Sottile et al. (2015)	To study the influence of pollution information and individual stress on modal shift from private car to P&R	Travel mode	SP (1) car, (2) park and Ride	Perception of traffic stress, Perception of information about stress, Personal norm
Fleischer et al. (2015)	To explore the influence of releasing safety information on consumer choice to flight with different airlines	Airlinetype	SP (1) Lufthansa, (2) easyJet (3) Aeroflot	Safety perception of airlines
Belgiawan et al. (2015)	To study the influence of expectations of others to buy a car on car purchase intentions of students	Car purchase	SP 7 scale ordered intention to buy a car	Symbolic-affective attitudes Independence Expectation of others

Author	Objective	Application Subject	Types of survey Choice set	Latent Variables
La Paix Puello et al. (2015)	To analyze observed and unobserved factors influencing cycling to railway stations	Travel mode	SP (1) Cycling, (2) others	Perception of connectivity Attitudes towards station environment Perceived quality of bicycle facilities
Cantillo et al. (2015)	To investigate the objective and subjective factors influencing the pedestrians' choice on how to cross an urban road	Pedestrian choice to cross an urban road	SP (1) crossing directly, (2) crossing by using a pedestrian bridge (3) crossing by using a crosswalk at a signalized intersection.	Perception of attractiveness Perception of safety/security
Thorhauge et al. (2016)	To investigate the objective and subjective factors that influence departure time choice	Departure time choice	SP 3 alternative departure times (1) the current departure time, (2) an earlier, (3) a later departure time	Psychological factors for departure time choice from the Theory of planned behavior: Attitude to being late Subjective norm Perceived behavioral control Intention to arrive on-time Attitude towards short travel time Perceived mobility necessities
Yazdanpanah & Hosseinlou (2016)	To determine the influence of personality traits on individual preference towards using PT	Travel mode	RP (1) parked at airport, (2) dropped-off, (3) van, (4) taxi, (5) metro, (6) bus	Personality traits (NEO Five-Factor Inventory test)

3 Theoretical framework for hybrid choice model

Under the assumption of rational decision making, individuals q will choose the alternative i (among a set of available alternatives) that maximizes their perceived utility. The utility is a random variable, formed by two components; V_{iq} called systematic utility and ε_{iq} error term:

$$U_{iq} = V_{iq} + \varepsilon_{iq}$$

The systematic utility represents the part of the utility that can be measured and is usually characterized through concrete and measurable properties of the alternatives and the individuals. The error term is the part of the utility that represents all unknown or abstract elements affecting the decision.

When considering a HCM modelling framework (Ben-Akiva et al., 2002), latent variables (η_{liq}) which are abstract and cannot be directly observed from revealed choices, are also included into the systematic utility. These variables are meant to represent attitudes and/or perceptions of the individuals. The latent variables are explained by a set of characteristics of the individuals and the alternatives (S_{iqr}), through so called structural equations, while explaining simultaneously, a set of attitudinal and/or perceptual indicators (y_{ziq}), previously gathered from the individuals, through so called measurement equations. Therefore, structural and measurement equations together form a latent variable model. The model identifies latent constructs as a function of the indicators, and captures the causal relationships between exogenous explanatory variables and the latent variables. This framework can be expressed by the following equations:

$$\eta_{liq} = \sum_{r} \alpha_{lri} \, S_{riq} + \vartheta_{liq}$$

$$y_{ziq} = \sum_{l} \gamma_{lzi} \, \eta_{liq} + \Sigma_{ziq}$$

Where;

- i: alternatives
- q: individuals
- r: exogenous variables
- l: latent variables
- z: indicators/items
- ϑ_{liq} and Σ_{ziq}: error terms, usually assumed normally distributed with mean zero and a certain covariance matrix
- α_{lri} and γ_{lzi}: parameters to be jointly estimated.

Under the assumption of linear parameters in V_{iq}, the utility function of HCM can be shown as;

$$U_{iq} = \sum_k \theta_{ki} X_{kiq} + \sum_l \beta_{li} \eta_{liq} + \varepsilon_{iq}$$

Where;

- k: number of explanatory variables
- X: explanatory variables
- θ_{ki}: estimated parameters related to the explanatory variables
- β_{li}: estimated parameters related to the latent variables

If the attributes are also assumed to be linear in the systematic utility, the estimated parameters θ_{ki} and β_{li} can be directly interpreted as marginal utilities.

Under the assumption that the error terms ε_{iq} are independent and identically distributed (IID) Extreme Value Type 1 (EV1) with the same variance, the differences between the utilities associated with the alternatives follow a Logistic distribution with mean zero and scale factor λ (usually being normalized to one, Walker (2002)), leading to the Multinomial Logit (MNL) model (Domencich & McFadden, 1975). Here, the probability of choosing alternative i is given by:

$$P_{iq} = \frac{e^{\lambda V_{iq}}}{\sum_j e^{\lambda V_{jq}}}$$

There are two methods to estimate the parameters of HCM namely, sequential and simultaneous estimations. The former is a two-step process in which, first the latent variables are estimated based on the structural equation model (SEM) and afterwards included in the discrete choice model as explanatory variables, whereas in the latter method the estimation of both parts of the model are performed simultaneously. The sequential estimation cannot guarantee consistent and unbiased estimators and can underestimate the standard deviation of the parameters (Train et al., 1987; Ben-Akiva et al., 2002). However, empirical evidence conducted by Raveau et al., (2010) and Bahamonde-Birke et al., (2010) using real and synthetic data revealed that both estimation methods are unbiased and the difference does not affect the mode estimates considerably.

In this study, the sequential method was used since simultaneous estimation cannot be implemented in R (the statistical computing tool used for this project).

4 Aim and scope

4.1 Background

There are a number of studies showing that attitudes towards environmental issues influence the travel choice behavior. Anable (2005) showed that increasing the awareness of people about the environmental issues of using car provide additional beliefs that can be targeted in order to change travel behavior. The results of the study conducted by Schüssler & Axhausen (2011) show that awareness of environmental problems and denial of environmental issues influence the travel mode choices. Atasoy et al. (2013) found that the environmental concern increases the utility of public transport so that the individuals who are sensitive to environmental issues use public transport more. Brazil & Caulfield (2013) found that providing information about CO2 emissions associated with public transport, in comparison to driving, increase the perceived utility of this mode.

There are also some studies indicating a relationship between having pro car attitudes and propensity to car use. Steg (2005) showed that people with a positive car attitude value symbolic, affective and instrumental motives of car use more favorably compared to those with a neutral or negative car attitude. De Vos et al. (2012) found that people with a pro car attitude are more likely to have an affinity toward car use. Likewise, Atasoy et al. (2013) showed that the people with a positive car attitude are more likely to use car instead of public transport.

4.2 Hypotheses

This study aimed to investigate the determinants of mode choice behavior by integrating two attitudinal latent variables namely, environmental concern and pro-car into discrete choice model. The main goal was to examine to what extent travel mode choice is affected by psychological factors versus microeconomic evaluation of the characteristic of the alternatives. More specifically, it is expected that the pro-car and environmental concern attitudes have a significant direct effect on travel mode choice. The following specific hypotheses were tested;

1) incorporating the latent variables improves the model estimation in comparison with a basic DCM without latent variable
2) people who are pro-car are more likely to choice car rather than public transport
3) people who are environmentally concern are more likely to choice public transport rather than car

This study, frankly speaking, does not add anything new and valuable to the scientific world. However, the capability of R to implement DCM and HCM, as shown by this study, can be a good motivation for the instructor of this course to consider this nice and user friendly software for the future semester.

5 Data set description

The dataset consists of survey data collected on the trains between St. Gallen and Geneva, Switzerland, during March 1998. The respondents provided information in order to analyze the impact of the modal innovation in transportation, represented by the Swissmetro. For more information about the data set and the process of data collection please see Antonini et al. (2007).

5.1 Variables for choice model

A number of changes were implemented in the original data set in order to prepare them for this study:

- Purpose 5,6,7,8 and 9 were aggregated and named PURPOSE 5
- AGE 6 was excluded from the data set
- INCOME 0 and 1 were aggregated and named INCOME 1
- The attributes of travel time and cost for all three alternative modes were divided by 100
- If the person has a GA (season ticket), her/his incremental cost became zero
- Choice indicator of zero was removed from the data set

The variables of the modified dataset are described in Tables 2.

Table 2. Description of the variables of the dataset

Variables	Description
ID	Respondent identifier
PURPOSE	Travel purpose: 1:Commuter , 2: Shopping, 3: Business, 4: Leisure, 5: Other
AGE	Age class of individuals: 1: age<=24, 2: 24<age<=39, 3: 39<age<=54, 4: 54<age<=65, 5: 65 <age
GENDER	Travelers gender: 0: Female, 1: Male
INCOME	Traveler's income per year [thousand CHF]: 1: under 50, 2: between 50 and 100, 3: over 100, 4: unknown
GA	It captures the effect of the Swiss annual season ticket for the rail system and most local public transport. 1: if the individual owns a GA, 0: otherwise.
AV.TRAIN	Train availability dummy
AV.CAR	Car availability dummy
AV.SM	SM availability dummy
TT.TRAIN	Train travel time [minutes]. Travel times are door- to-door making assumptions about car-based distances (1.25*crow-flight distance)
CO.TRAIN	Train cost [CHF]. If the traveler has a GA, this cost equals the cost of the annual ticket.
TT. SM	SM travel time [minutes] considering the future Swiss-metro speed of 500 km/h
CO.SM	SM cost [CHF] calculated at the current rail fare, without considering GA, multiplied by a fixed factor (1.2) to reflect the higher speed.
TT.CAR	Car travel time [minutes]
CO.CAR	Car cost [CHF] considering a fixed average cost per kilometer (1.20 CHF/km)
CHOICE	Choice indicator: 1: Train, 2: SM, 3: Car

The descriptive statistics of the attributes related to travel time and cost are summarized in Table 3. Table 4 describes the sample characteristics of the remaining variables.

Table 3. Descriptive statistics of the attributes of travel time and cost

Variable	**Min**	**Max**	**Mean**	**St. Dev**
TT.TRAIN	0.31	10.49	1.67	0.77
TT.SM	0.08	7.96	0.88	0.54
TT.CAR	0	15.6	1.24	0.89
CO.TRAIN	0	5.76	0.83	0.68
CO.SM	0	7.68	1.01	0.84
CO.CAR	0	5.2	0.79	0.55

Table 4. Sample characteristics

Variable	**Categories (%)**				
PURPOSE	Commuter 14.7	Shopping 11.8	Business 48.4	Leisure 21.5	Other 3.6
AGE	age<=24 6.6	24<age<=39 31.2	39<age<=54 35.7	54<age<=65 18.9	65 <age 7.6
GENDER	Female 25	Male 75			
INCOME	Under 50 18.9	Btw 50 & 100 34.9	Over 100 37.7	Unknown 8.5	
CHOICE	Train 13.2	SM 58	Car 28.8		
Number of unique ID: 1190					
Total number of observations: 10710					

5.2 Variables for latent model

There are two latent variable included in to the model and each of them was measured by three items. Tables 4 shows the items and statements used to describe the two latent variables.

Table 5. Latent variables items

Latent variables	**Items**	**Statements**
Pro-car	PC_1	Driving car gives me a feeling of freedom
	PC_2	I do not like to change mode when I travel
	PC_3	Driving a car is a cool way to travel
	PC_4	I do not feel more independent when I drive a car
Environmental concern	EC_1	I am concerned about global warming
	EC_2	We should increase the price of gasoline to reduce air pollutions
	EC_3	We must act and take decisions to limit CO2 emissions
	EC_4	The environmental impacts of using car are not very important

A balanced questionnaire design including both positively and negatively phrased was chosen because of a well-established and widely-used copying strategy with acquiescent responding, namely the tendency to agree with questionnaire items (Ferrando and Lorenzo-Seva, 2010). The respondents were asked to rate their level of agreement to each statement on a five-point Likert scale ranging from a total disagreement (response of 1) to a total agreement (response of 5).

Very important:
The original dataset does not contain any latent variable. Therefore, the latent variables and items are totally fake and generated by the author. The corresponding algorithm for the data generation is described in Appendix I.

Table 6 presents the means, standard deviations and the internal consistencies of the psychological constructs. Cronbach's alpha reliability test is used to measure the reliability of two or more construct indicators. Some professionals insist on a reliability score of 0.70 or higher in order to use a psychometric instrument. The data does not show good internal consistency with the low Cronbach's alphas. The reason is mainly related to the way the data generated and the unbalanced dataset regarding travel choices. Since this is only a course assignment I ignore this issue.

Table 6. Mean, Standard deviation and Cronbach's α for the items

Latent variables	Items	Mean	SD	Cronbach's α
Pro-car	PC_1	2.62	1.26	0.203
	PC_2	2.36	1.47	
	PC_3	2.58	1.21	
	PC_4	3.43	1.19	
Environmental concern	EC_1	3.44	1.21	0.162
	EC_2	3.66	1.44	
	EC_3	3.45	1.22	
	EC_4	2.57	1.21	

6 Theoretical framework

The theoretical framework hypothesized in the current study to explain mode choice behavior is shown in Figure1.

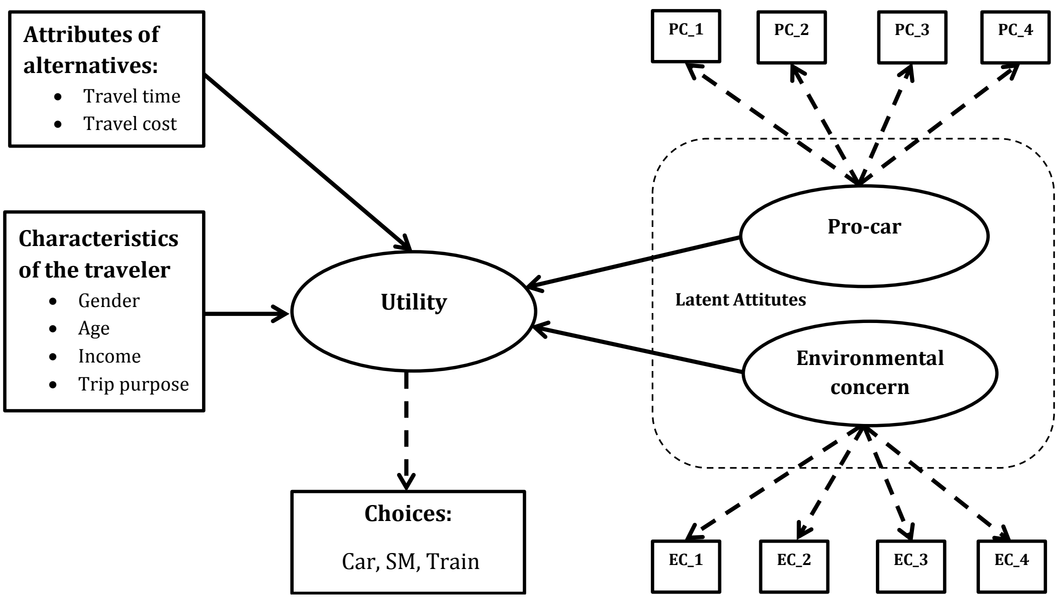


Figure 1. Theoretical framework

7 Models and results

In this section, firstly the results of SEM is described, and then the integration of the psychological factors into the DCM is explained.

7.1 Estimation of the latent variables

Table 7 shows the parameters estimates and critical ratios (C.R.) for the indicators of the two latent constructs. Confirmatory factor analysis (CFA) was used to show relationships between the latent variables and their indicators. The C.R. is the ratio of parameter estimate and standard error. The model was estimated using maximum likelihood estimator in the lavaan package (R). The negative values for the parameter estimates are related to negatively-phrased items.

The models fit statistics are also provided in the Table 7. The goodness of fit measures in terms of RMSEA equals to 0.03 and CFI equals to 0.998 shows that the model fits the observed data well (Muthén & Muthén, 2009).

Table 7. Estimates of the measurement equations of the latent constructs

Pro-car	**Estimate**	**C.R.**
Driving car gives me a feeling of freedom	0.94	-
I do not like to change mode when I travel	0.94	188.3
Driving a car is a cool way to travel	0.94	191.5
I do not feel more independent when I drive a car	-0.83	-132.7
Environmental concern	**Estimate**	**C.R.**
I am concerned about global warming	0.87	-
We should increase the price of gasoline to reduce air pollutions	0.88	123.8
We must act and take decisions to limit CO2 emissions	0.89	126.5
The environmental impacts of using car are not very important	-0.84	-112.7
Test statistics		
Comparative Fit Index (CFI): 0.998		
Root Mean Square Error of Approximation (RMSEA): 0.03		
Standardized Root Mean Square Residual (SRMR): 0.005		
$\chi^2 = 206.93$, *df= 19*		

7.2 Integration of the latent variables in the mode choice

The microeconomic approach to mode choice is based on the concept that individuals q make a trade-off between the attributes of travel alternatives j (i.e. TT and TC) and choose the alternative gives them the highest utility (U). The socioeconomic characteristics (SE) of individual influence their preferences. This approach was extended to explain the fact that individual choices can be also influenced by the latent effects (LV).

In this study, the data were collected through a stated preference choice experiment. One feature of the data set is the presence of multiple observations on the choice responses for each individual. This means that the potential for correlated responses across observations is a violation of the independence of observations assumption in classical choice model estimation. In order to deal with this issue, mixed logit model were used in which the model is specified in such a way that the choice sets can be correlated across each individual.

based on what mentioned above, the model takes the following form:

$$U_{jqt} = ASC_j + \beta_{TT}TT_{jqt} + \beta_{TC}TC_{jqt} + \beta_{SE}SE_q + \beta_{LV}LV_q + \mu_{jq} + \varepsilon_{jqt}$$

ASC_j is a constant specific for each alternative j, $\mathcal{E}_{jq}$ is an error term distributed identically and independently extreme value, and βs are the parameters, associated to each attribute, to be estimated. μ_{jq} is a normally distributed error component that captures the correlation among choice tasks t answered by the same individual.

The results of the model estimated are presented in Table 8 where SM was maintained as the base. For purpose of comparison, the mixed logit model without any latent effects and the hybrid choice model placed next to each other.
As expected, the coefficients of LOS attributes in both models are negative and significant (p-value<0.01). However, the significance level of the parameters in the HCM are slightly smaller than those in the model without latent variables. The negative signs indicate that the utility decrease if any of the attributes increases. In other words, if travel time or travel cost increases the probability of selecting that travel mode decreases. Hence, individuals seek to balance the attributes by choosing the travel mode provides them with the highest overall utility. In our sample, travel time and cost are equally important for the respondents as the coefficients are almost identical.

Table 8 also shows that the mode choice varies considerably with trip purpose, age class of individuals and their sex. In both models, the mode-specific socioeconomic dummies, generally display the same pattern in terms of the sign, magnitude and significance. I this regards, there is no significant difference between male and female to choose car rather than SM. However, females show higher propensity to travel by train rather than SM.

Compare to commuting trips, on one side, individuals are less likely to choose car rather than SM for shopping trips. Furthermore, individuals are more likely to choose car rather than SM for business and leisure trips. On the other side, travelers are more likely to choose train rather than SM for shopping and business trips while as for leisure trips, there is not significant difference between the choice of train and SM (p-value < 0.05).

Compare to age class 1, age classes 2 to 4 show lower propensity to travel by car and train rather than SM. When it comes to age class 5, they are more likely to choose train and less likely to choose car rather than SM.

Table 8. Robust ML parameter estimates for the traditional and hybrid model (SM is the reference)

	Mixed Logit model		HCM_Mixed logit model	
	Estimate	**t-test**	**Estimate**	**t-test**
Intercept (Car)	-1.449	-5.028***	-0.872	-3.063***
Intercept (Train)	1.077	6.503***	1.012	5.864***
St.dev (Car)	15.51	38.317***	11.18	33.677***
St.dev (Train)	8.45	31.361***	7.77	30.333***
Cov (Car-Train)	3.86	14.256***	3.21	13.592***
LOS variables				
Travel time	-3.144	-54.744***	-2.960	-49.419***
Travel cost	-3.120	-40.118***	-2.770	-34.924***
SE variables				
Gender (Car)	-0.008	-0.074	-0.087	-0.841
Gender (Train)	-1.310	-13.407***	-1.281	-13.105***
Purpose 2 (Car)	-0.187	-1.026	-0.391	-2.170**
Purpose 2 (Train)	0.763	5.209***	0.672	4.632***
Purpose 3 (Car)	2.727	20.841***	2.101	16.600***
Purpose 3 (Train)	0.375	2.787***	0.259	1.932*
Purpose 4 (Car)	4.454	28.711***	2.923	20.611***
Purpose 4 (Train)	-0.054	-0.313	-0.174	-0.964
Purpose 5 (Car)	-2.288	-5.929***	-1.540	-4.047***
Purpose 5 (Train)	0.731	3.249***	0.685	3.063***
Age class 2 (Car)	-2.066	-7.189***	-1.750	-6.233***
Age class 2 (Train)	-2.610	-18.127***	-2.540	-17.621***
Age class 3 (Car)	-1.277	-4.591***	-1.289	-4.698***
Age class 3 (Train)	-2.716	-18.149***	-2.626	-17.667***
Age class 4 (Car)	-1.847	-6.359***	-1.884	-6.560***
Age class 4 (Train)	-1.929	-11.467***	-1.864	-10.981***
Age class 5 (Car)	-0.144	-0.474	-0.517	-1.714*
Age class 5 (Train)	0.804	4.007***	0.890	4.392***
Latent effects				
Pro-car (Car)			0.748	8.934***
Pro-car (Train)			-0.054	-0.485
Env. concern (Car)			-0.610	-5.894***
Env. concern (Train)			0.168	1.206
Summary				
Sample size	10,710		10,710	
Number of draws	500		500	
RHO^2	0.4		0.410	
Adjusted RHO^2	0.398		0.408	
Null Log-likelihood	-10083.33		-10083.33	
Final Log-likelihood	-6050.4		-5945.2	

Note:
* p<0.1
** p<0.05
*** p<0.01

The latent variables part shows the direct influence of the psychological factors in the choice of travel mode. As expected, the parameters for car choice are significant meaning that who are pro car compared to who are not, are more likely to select car rather than SM and who are environmentally concern compared to who are not, are less likely to select car rather than SM. In other words, individuals who choose car gain utility and disutility of respectively having pro-car and pro-environmental attitudes. The parameters for train choice are statistically insignificant (p-value < 0.05) meaning that having such attitudes does not influence the individuals' choice between train and SM. In Appendix II, the results of the model estimated are presented in Table 9 where Train was maintained as the base. Considering the latent effects in both Table 8 and 9, it could be concluded that individuals who are pro-car are more likely to choice car rather than public transport and who are environmentally concern are more likely to choice public transport rather than car.

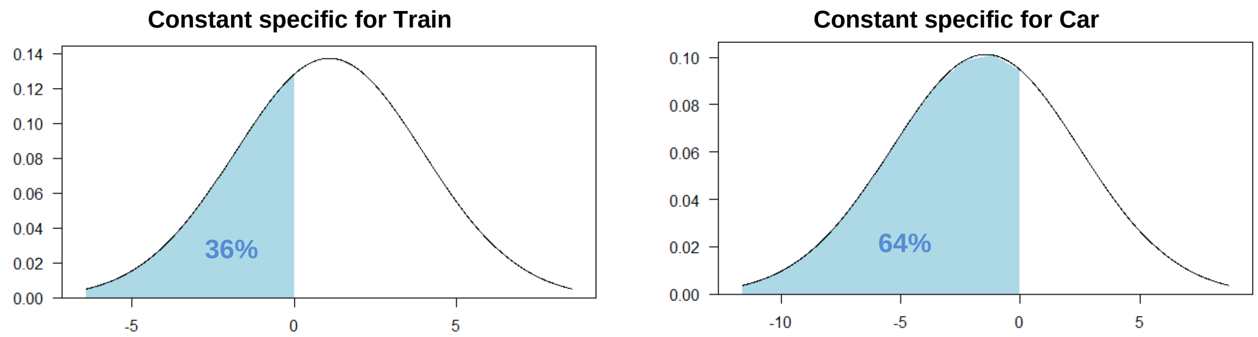


Figure 2. Distribution of the random parameters for the model without latent variables

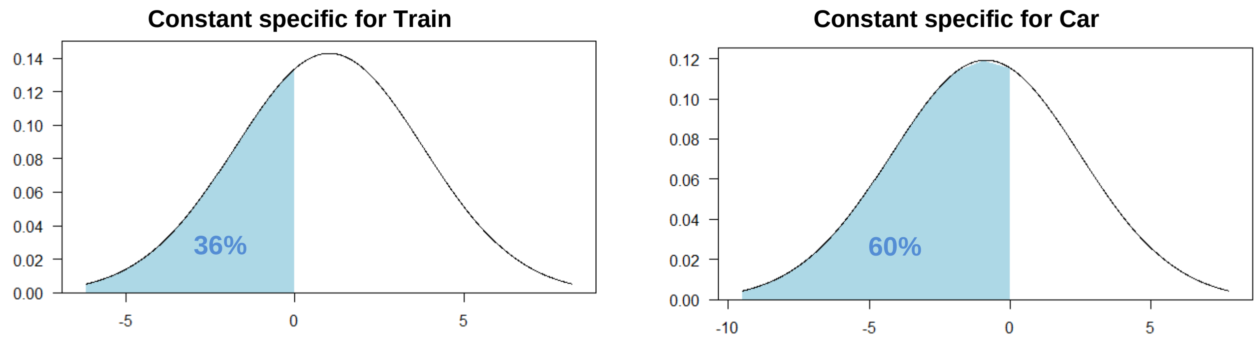


Figure 3. Distribution of the random parameters for the HCM model

To test Hypotheses 1-3, a likelihood ratio test (Ben-Akiva & larman, 1985) was performed for the HCM against the traditional model and found that the model fit for the HCM significantly improved. Therefore, the results support the hypothesis.

Test statistic: $\chi^2 = -2(logL(\hat{\beta}_R) - logL(\hat{\beta}_U)) = 210.4$

$P(\chi^2_{df=4} > 210.4)$ = 2.2e-16.

8 Conclusion

The results of the study show that the understanding and modelling of travel mode choice can be improved by the inclusion of relevant psychological factors in to a DCM. Two psychological factors namely, pro-car and environmental concern were selected and it was assumed that having these attitudes play a role in individuals travel mode choice.

It was demonstrated that by including these two psychological effects, the model estimation were significantly improved compared to a DCM model without latent variables. Furthermore, the specific effects of the psychological variables in the utility of travel modes were found. As expected, the pro-car attitude increase the utility of car and the environmental concern increases the utility of public transport so that the individuals who are sensitive to environmental issues use public transport more.

The negative consequences of private car use are increasingly being recognized, especial ly the link between polluting emissions and global warming. In recent years, research has increasingly focused the attention on measures and policies that affect individuals' travel behavior in order to achieving car-use reduction. The results of this study indicate the importance of making people aware of the environmental issues which could motivate individuals to voluntarily switch their car travel to more sustainable travel modes.

Further research could consist of the inclusion of more attitudinal variables as well as a better characterization of their indicators in order to explain more carefully individuals travel behavior and realize such motivations.

9 Reference

Anable, J. (2005). Complacent car addicts' or 'aspiring environmentalists? Identifying travel behaviour segments using attitude theory. *Transport Policy*, *12*(1), 65-78.

Antonini, G., Gioia, C., & Frejinger, E. (2007). Swissmetro: description of the data.

Atasoy, B., Glerum, A., & Bierlaire, M. (2013). Attitudes towards mode choice in Switzerland. *disP-The Planning Review*, *49*(2), 101-117.

Bahamonde-Birke, F., Raveau, S., Yáñez, M. F., & Ortúzar, J. D. D. (2010). The role of tangible attributes in hybrid discrete choice models. In *European Transport Conference* (pp. 11-13).

Belgiawan, P. F., Schmöcker, J. D., Abou-Zeid, M., Walker, J., & Fujii, S. (2015). The role of expectation of others on students' likelihood to buy a car. *management*, *18*, 19.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. *Perpetual motion: Travel behaviour research opportunities and application challenges*, 431-470.

Cantillo, V., Arellana, J., & Rolong, M. (2015). Modelling pedestrian crossing behaviour in urban roads: a latent variable approach. *Transportation research part F: traffic psychology and behaviour*, *32*, 56-67.

Daly, A., Hess, S., Patruni, B., Potoglou, D., & Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation*, *39*(2), 267-297.

Daziano, R. A., & Bolduc, D. (2013). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica A: Transport Science*, *9*(1), 74-106.

De Vos, J., Derudder, B., Van Acker, V., & Witlox, F. (2012). Reducing car use: changing attitudes or relocating? The influence of residential dissonance on travel behavior. *Journal of Transport Geography*, *22*, 1-9.

Domencich, T. A., & McFadden, D. (1975). *Urban travel demand-a behavioral analysis*. North Holland, Amsterdam.

Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(2), 427-448.

Fleischer, A., Tchetchik, A., & Toledo, T. (2015). Does it pay to reveal safety information? The effect of safety information on flight choice. *Transportation Research Part C: Emerging Technologies*, *56*, 210-220.

Glerum, A., Stankovikj, L., Thémans, M., & Bierlaire, M. (2013). Forecasting the demand for electric vehicles: accounting for attitudes and perceptions. *Transportation Science*, *48*(4), 483-499.

Kamargianni, M., Ben-Akiva, M., & Polydoropoulou, A. (2014). Incorporating social interaction into hybrid choice models. *Transportation*, *41*(6), 1263-1285.

Kamargianni, M., Dubey, S., Polydoropoulou, A., & Bhat, C. (2015). Investigating the subjective and objective factors influencing teenagers' school travel mode choice–An integrated choice and latent variable model. *Transportation Research Part A: Policy and Practice*, *78*, 473-488.

Kim, J., Bae, Y., & Chung, J. H. (2012). Effects of personal pro-environmental attitudes on mode choice behavior: new ecofriendly water transit system in Seoul, South Korea. *Transportation Research Record: Journal of the Transportation Research Board*, (2274), 175-183.

La Paix Puello, L. C., & Geurs, K. T. (2015). Modelling observed and unobserved factors in cycling to railway stations: application to transit-oriented-developments in the Netherlands. *European journal of transport and infrastructure research*, *15*(1), 27-50.

Mabit, S. L., Cherchi, E., Jensen, A. F., & Jordal-Jørgensen, J. (2014). Hybrid choice modeling allowing for reference-dependent preferences: estimation and validation for the case of alternative-fuel vehicles. In *Transportation Research Board 93rd Annual Meeting* (No. 14-0705).

Muthén, B., & Muthén, B. O. (2009). *Statistical analysis with latent variables*. Hoboken: Wiley.

Raveau, S., Álvarez-Daziano, R., Yáñez, M., Bolduc, D., & de Dios Ortúzar, J. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record: Journal of the Transportation Research Board*, (2156), 131-139.

Paulssen, M., Temme, D., Vij, A., & Walker, J. L. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation*, *41*(4), 873-888.

Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., & Lee, A. (2011). Importance of traveler attitudes in the choice of public transportation to work: findings from the Regional Transportation Authority Attitudinal Survey. *Transportation*, *38*(4), 643-661.

Prato, C. G., Bekhor, S., & Pronello, C. (2012). Latent variables and route choice behavior. *Transportation*, *39*(2), 299-319.

Scagnolari, S., Walker, J., & Maggi, R. (2015). Young drivers' night-time mobility preferences and attitude toward alcohol consumption: A Hybrid Choice Model. *Accident Analysis & Prevention*, *83*, 74-89.

Schüssler, N., & Axhausen, K. W. (2011). *Investigating the influence of the environmentalism and variety seeking on mode choice*. Eidgenössische Technische Hochschule Zürich, IVT Institute for Transport Planning and Systems.

Soto, J. J., Cantillo, V., & Arellana, J. (2014). Hybrid Choice Modelling of Alternative Fueled Vehicles in Colombian Cities Including Second Order Structural Equations. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4545).

Sottile, E., Cherchi, E., & Meloni, I. (2015). Measuring soft measures within a stated preference survey: the effect of pollution and traffic stress on mode choice. *Transportation Research Procedia*, *11*, 434-451.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice*, *39*(2), 147-162.

Temme, D., Paulssen, M., & Dannewald, T. (2008). Incorporating latent variables into discrete choice models—a simultaneous estimation approach using SEM software. *BuR-Business Research*, *1*(2), 220-237.

Thorhauge, M., Haustein, S., & Cherchi, E. (2016). Accounting for the Theory of Planned Behaviour in departure time choice. *Transportation Research Part F: Traffic Psychology and Behaviour*, *38*, 94-105.

Train, K. E., McFadden, D. L., & Goett, A. A. (1987). Consumer attitudes and voluntary rate schedules for public utilities. *The Review of Economics and Statistics*, 383-391.

Walker, J. (2002). Mixed logit (or logit kernel) model: Dispelling misconceptions of identification. *Transportation Research Record: Journal of the Transportation Research Board*, (1805), 86-98.

Yazdanpanah, M., & Hosseinlou, M. H. (2016). The influence of personality traits on airport public transport access mode choice: A hybrid latent class choice modeling approach. *Journal of Air Transport Management*, *55*, 147-163.

10 Appendix I

10.1 Generation of the latent variables:

I wanted to prove that having pro-car attitude increases car utility and conversely, having pro-environmental decreases its utility. The latent construct related to pro-car has four indicators as described in Table 5. I generated data for PC_1, PC_2 and PC_3 in such a way that if the traveler's choice was car then they agreed or completely agreed with the statement. When it comes to PC_4, since it is a negative phrase, they disagreed or completely disagreed with the statement.

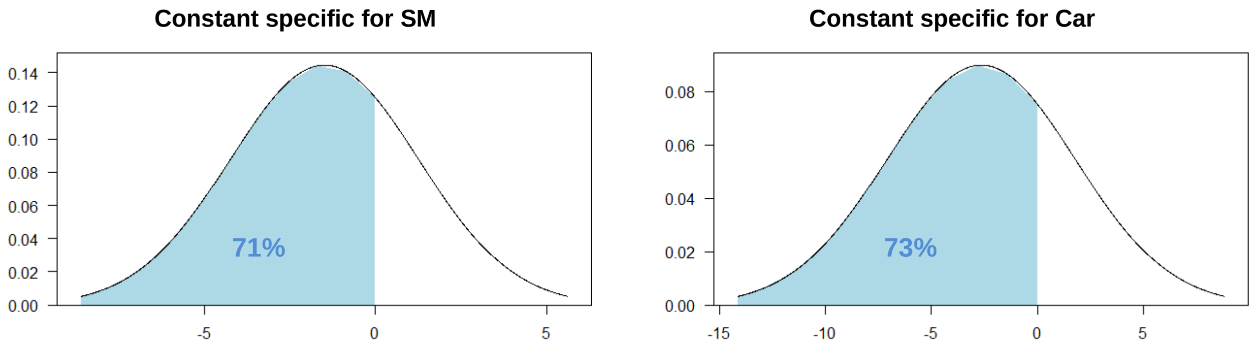
I wanted also to prove that having pro-car attitude decreases train and SM utilities and conversely, having pro-environmental increase their utilities. The latent construct related to environmental concern has four indicators as described in Table 5. I generated data for EC_1, EC_2 and EC_3 in such a way that if the traveler's choice was either train or SM then they agreed or completely agreed with the statement. When it comes to EC_4, since it is a negative phrase, they disagreed or completely disagreed with the statement.

11 Appendix II

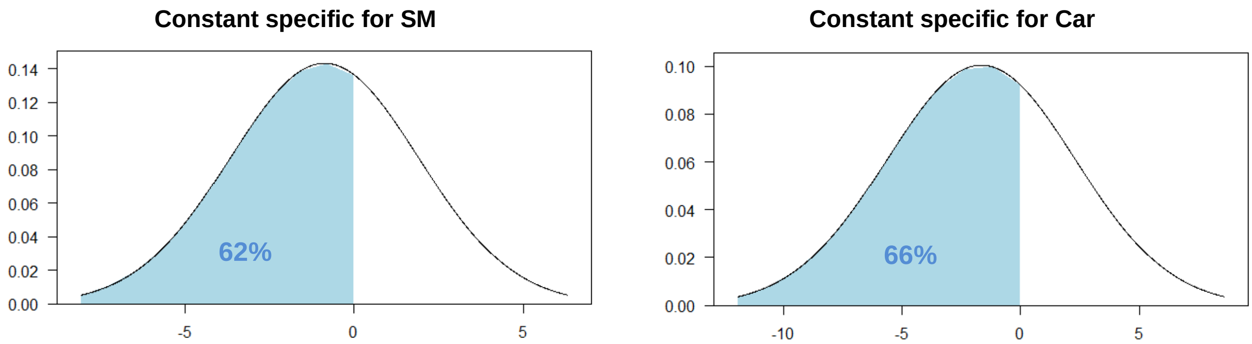
Table 9. Robust ML parameter estimates for the traditional and hybrid models (Train is the reference)

	Mixed Logit model		HCM_Mixed logit model	
	Estimate	t-test	Estimate	t-test
Intercept (Car)	-2.656	-8.276***	-1.664	-5.237***
Intercept (SM)	-1.497	-8.990***	-0.878	-5.101***
St.dev (Car)	19.739	34.498***	15.777	32.058***
St.dev (SM)	7.617	38.980***	7.748	35.580***
Cov (Car-SM)	4.377	12.492***	5.953	18.081***
LOS variables				
Travel time	-3.246	-56.082***	-2.988	-51.700***
Travel cost	-3.216	-41.368***	-2.831	-36.432***
SE variables				
Gender (Car)	1.336	10.004***	1.180	8.889***
Gender (SM)	1.293	13.417***	1.168	12.127***
Purpose 2 (Car)	-0.785	-3.532***	-1.137	-5.155***
Purpose 2 (SM)	-0.500	-3.458***	-0.630	-4.348***
Purpose 3 (Car)	2.533	14.041***	1.805	10.313***
Purpose 3 (SM)	-0.231	-1.721*	-0.376	-2.812***
Purpose 4 (Car)	4.844	21.537***	3.040	14.122***
Purpose 4 (SM)	0.310	1.791*	0.043	0.240
Purpose 5 (Car)	-3.525	-8.327***	-2.795	-6.691***
Purpose 5 (SM)	-0.503	-2.239**	-0.581	-2.593***
Age class 2 (Car)	0.169	0.564	0.443	1.498
Age class 2 (SM)	2.701	19.080***	2.566	18.077***
Age class 3 (Car)	1.095	3.714***	1.200	4.116***
Age class 3 (SM)	2.818	19.271***	2.752	18.731***
Age class 4 (Car)	-0.157	-0.502	-0.340	-1.090
Age class 4 (SM)	2.032	12.316***	1.884	11.254***
Age class 5 (Car)	-1.094	-3.185***	-1.713	-4.974***
Age class 5 (SM)	-0.544	-2.719***	-0.735	-3.623***
Latent effects				
Pro-car (Car)			0.842	6.439***
Pro-car (SM)			0.177	1.590
Env. concern (Car)			-0.835	-5.148***
Env. concern (SM)			-0.090	-0.647
Summary				
Sample size	10,710		10,710	
Number of draws	500		500	
RHO^2	0.40		0.41	
Adjusted RHO^2	0.398		0.408	
Null Log-likelihood	-10083.3		-10083.3	
Final Log-likelihood	-6050.4		-5945	

Note:
* $p<0.1$
** $p<0.05$
*** $p<0.01$



Distribution of random parameters for the model without latent variables



Distribution of random parameters for the HCM model