Adjectives qualifying individuals' perceptions impacting on transport mode preferences

Aurélie Glerum * Bilge Atasoy * Alberto Monticone †

Michel Bierlaire *

Second International Choice Modeling Conference
July 2011

Abstract

In this paper, we present an analysis of the impact of perceptions on transport mode choice. Traditional revealed preference surveys very often only involve questions leading to the collection of quantitative data, discrete or continuous, but little attention is given to qualitative data. In the survey on which this research is based, such data appears as adjectives describing a series of transport modes, freely reported by respondents. The calibration of an integrated choice and latent variable model has shown that travelers' perception of comfort in public transports has a significant impact on their mode preferences. Moreover the prediction of individuals' choices by such model is much more accurate than by a logit model with multiple alternatives.

Key words

Discrete choice models, latent variables, attitudes, perceptions, qualitative data, transportation, demand analysis, market shares, elasticities.

*ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR), {aurelie.glerum, bilge.atasoy, michel.bierlaire}@epfl.ch

†POLITECNICO DI TORINO, Department of hydraulics, transport and civil infrastructures (DITIC), al.monticone@gmail.com

Contents

1 Introduction

Individuals' mobility is complex and sometimes hard to understand. In this research, we are particularly interested in understanding the transport mode choices of inhabitants of low-density areas, in order to figure out their needs in terms of mobility. Our assumption is that their mode preferences are not only driven by classical variables such as cost or time, but that some aspects related to their attitudes or perceptions play a major role in their decisions and need to be taken into account to analyze their transport mode choices.

This motivates the development of survey designs, involving new ways of phrasing questions in order to obtain different types of answers. To capture individuals' attitudes or perceptions, the techniques used in the literature include showing a certain number of statements where respondents have to indicate a rating on a five-point scale. For example, Vredin Johansson et al. (2006) collect ratings of the importance of some attitudinal questions related to comfort, convenience and flexibility of transport modes, Abou-Zeid et al. (2008) ask respondents to indicate their satisfaction when they commute either by public transports or by car, Hurtubia et al. (2010) and Atasoy et al. (2010) request individuals to rate their agreement on a list of statements related to a set of topics, including environmental concern or lifestyle.

The design of diverse data collection techniques generates a need for the development of new demand models. In particular, we want to address the issue of assessing the impact on choice of latent attitudes or perceptions. So far, this has been performed by integrating a choice model with a latent variable model in a hybrid framework (see Walker, 2001, Walker and Ben-Akiva, 2002 and Ben-Akiva et al., 2002). In the literature, the effect on choice of several latent variables has been evaluated. In particular, Espino et al. (2006) assess the effect of the latent variable of comfort on the choice between bus and car. Vredin Johansson et al. (2006) also analyze the impact of comfort on mode preferences as well as the one of flexibility and care for environment.

We tackle the issue of collecting indicators of perception by using a particularly new method which consists of asking respondents to report several adjectives describing best a variable of interest (Kaufmann et al., 2001 and Kaufmann et al., 2010). In the case of the analysis of transport mode preferences, this variable can be any available mode in the context of the study.

The subsequent step is to model the impact of perception on mode choice. Perception being a concept which is not directly available from the data, we model it as a latent variable. Hence, the framework we will use in this modeling context is an integrated choice and latent variable model.

The field of possible answers for a question asking to report adjectives describing people's perception of a transport mode is very wide and therefore allows for the analysis of a variety of themes related to perception. For instance, one could model perception of flexibility, environmental impact or reliability of a particular transport mode. In this paper the analysis of the impact on choice of individual's perception of comfort in public transports is presented.

The data which was used in this research comes from a revealed preference survey performed in low-density areas of Switzerland as a joint project between PostBus, one of the major companies operating in such regions, and EPFL's Transportation Center (TraCe). In this survey, inhabitants were asked to describe all trips performed on a particular day as well as a the chosen transport modes.

In this paper, we first present the data collection phase. Then we describe the integrated choice and latent variable model and explain how we take into account qualitative *adjective* data in the framework. In order to test the validity of the model, a forecasting analysis is performed and presented in a subsequent section. We finally conclude by presenting some possible further developments.

2 Data collection

In order to analyze the transport mode choices of inhabitants of regions which are little connected, a *revealed preference* survey was set up. Such survey aims at collecting data about individuals' real choices.

This section presents the context preceding this survey as well as its structure.

2.1 Complete survey

The revealed preference survey on which the present research is based was designed using results of a qualitative survey conducted on inhabitants of rural areas or suburban areas. The latter consisted on interviews coupled with GPS recordings of the respondents' trips. A complete description of the qualitative survey is given in Doyen (2010).

The outcomes of the qualitative survey enabled us to design the questions asked in the revealed preference survey. In particular, this is the case of some of the questions related to people's opinions or attitudes such as their concern for environmental.

2.2 Revealed preference survey

The qualitative survey was followed by a quantitative one, i.e. the revealed preference survey, which was conducted between 2009 and 2010. Two exemplars of a questionnaire were sent to each household of 57 towns or villages connected by post busses. The towns and villages were selected in order to be representative of the whole network of PostBus and respondents of 16 years and over were asked to answer the questionnaire. In total, $1'763$ valid questionnaires were collected. Each of them consisted of six parts:

Description of trips: The respondents had to report all trips they performed in one day. In particular, they had to mention in the mode(s) they used for each trip, the activity they performed at the destination, the trip duration and the cost of fuel or public transport ticket. To summarize, this section contained the central information necessary for the calibration of a discrete choice model.

Opinions: A series of statements was shown to the respondents and they had to rate their agreement on a five-point Likert scale, ranging from a total disagreement to a total agreement. The sentences related to different topics such as environmental concern, mobility, residential choice and lifestyle were defined on the basis of results of the qualitative survey. Examples of statements related to each respective theme are reported below:

> *The price of gasoline should be increased in order to reduce traffic congestion and air pollution.*
>
> *Taking the bus helps making a town more comfortable and welcoming.*
>
> *Accessibility and mobility conditions are important in the choice of an accommodation.*
>
> *I always plan my activities a long time in advance.*

The answers to these statements build the data necessary for the inclusion of latent variables in the choice model.

Mobility habits: In order to understand the underlying concepts that drive individuals' mode preferences, a section of the questionnaire was dedicated to the collection of information on people's mobility habits. For instance, it included questions about the transport modes used for certain trip types (work, shopping, leisure, etc.) or during childhood.

Perception of transport modes: Respondents were asked to report three adjectives describing best each transport mode of a given list. The latter included the following modes: car, train, bus/metro/tram, post bus, bike or walk. This section of the survey is fundamental to our analysis of perception of transport modes, as it provides a very rich data set made of words freely reported by respondents. But in order to include such variables into a model, they need to be quantified. This procedure is explained in details in section 2.3.

Household description: As mode choices are often results of common decisions taken within a household and not necessarily relative to the respondent's own choice, a section of the questionnaire was designed to collect information about the respondent's household, such as its number of cars or its total income.

Personal data: Finally, classical socio-economic data about the respondent was collected.

The data collected in all sections above provided a very complete basis for the calibration of a discrete choice model. Let us remark that due to the inaccuracy of the durations and costs reported by the respondents for each of their trips, we used times and costs given on the websites of the Swiss railways (SBB) `http://www.cff.ch` and of ViaMichelin `http://fr.viamichelin.ch`. The same websites were used to infer the times and costs for the non-chosen alternatives.

2.3 Adjective data

In the section of the questionnaire on the perception of the transport modes, respondents had to report three adjectives describing best the following set of transport modes: car, train, bus/metro/tram, post bus, bike or walk. As an illustation, Table 1 presents the exact way the question was formulated.

For each of the following transport modes, give three adjectives that describe them best according to you.

		Adjective 1	Adjective 2	Adjective 3
1	The car is:			
2	The train is:			
3	The bus, the metro and the tram are:			
4	The post bus is:			
5	The bicycle is:			
6	The walk is:			

Table 1: Question on the perception of several transport modes, as it appeared in the questionnaire.

These adjectives were grouped into several themes including comfort, perception of time, perception of cost, difficulty of access, flexibility, efficiency, reliability, environmental impact, appreciation, look, etc. The adjectives classified within each theme provide information which we assume to reflect closely each respondent's perception of each topic, as they are freely reported.

In the model presented in this paper (see section 4 for the specification), we are interested in evaluating the effect of the perception of one of the characteristics of transport modes listed above, that is *comfort in public transports*, on the mode choice. Hence, we use the adjective data relative to modes train, bus/metro/tram and post bus as indicators of this particular attribute. But first the adjective data must be quantified, i.e. we need to find a scale of 'comfort'. In order to have enough variability in this scale, we ranked the adjectives describing comfort on a five-point scale from -2 to 2. For example, adjective 'bumpy' would be rated as -2, adjective 'tiring' as -1, adjective 'empty' as 1, adjective 'relaxing' as 2. All adjectives which are not related to comfort were coded as 0. The adjectives related to the theme comfort with their corresponding scale are shown in Table 2[1]. Let us note that none of the adjectives of the table are coded with value 0 as they are all related to the perception of comfort, in a positive or negative way. Adjectives with scale 0 are related to other themes. Example of such adjectives could be 'precise', 'expensive' or 'healthy' which do not give any information on the perception of comfort.

Let us note that a drawback of the coding of adjectives presented here is that it is subjective, as some variation when assigning the scales could occur from one modeler to the other. Improvements on that aspect are planned in future research.

[1]Let us remark that some of the words reported by respondents are actually not adjectives. They were nevertheless included in the analysis.

Comfort	Scale
hardly full	1
packed	-1
bumpy	-2
comfortable	1
hard	-1
irritating	-2
tiring	-1
unsuitable with bags	-1
uncomfortable	-1
bad air	-2
unsuitable with strollers	-1
difficult	-2
full	-1
relaxing	2
restful	2
without stress	2
shaking	-2
stressful	-1
suffocating	-1
empty	1

Table 2: List of adjectives related to the perception of comfort, together with the corresponding scale.

As respondents reported three adjectives for each of the three public transports (train, bus/metro/tram and post bus), the coding described above implied the creation of nine variables, which are indicators of the perception of comfort in public transports. These indicators will be used as measurements of the latter.

3 Integrated choice and latent variable model

In this section, we present the modeling framework that is used to explain and predict individuals' mode preferences.

As we aim at explaining people's mode choices with attitudes or perceptions, which are not directly available from the data, we use a modeling framework allowing for the inclusion of latent concepts, i.e. an *integrated choice and latent variable model* (see Walker, 2001). A scheme of such framework is shown in Figure 1. It consists of two models:

- A discrete choice model.
- A latent variable model.

Two sorts of variables build this framework:

Measurable variables: These variables (represented by rectangles) are either *explanatory variables* X, which aim at explaining individuals' choices or latent attitudes, *indicators* I of a latent variable, often in the form of answers to questions about opinions, or *indicators* y of the respondents' actual choices. We use indicators of a latent variable to quantify it, as it is not directly available from the data.

Latent variables: The latent variables X^* (represented by ovals) can be attitudes or perceptions, which cannot be directly obtained from the data, or utilities U, which measures how useful the different choices are for each individual.

An integrated choice and latent variable model is made of two types of relationships:

Structural equations: These equations aim at explaining an unobservable variable by some observable explanatory variables.

Measurement equations: These relationships express a measurable indicator by latent variables, the latter being not directly available from the data.

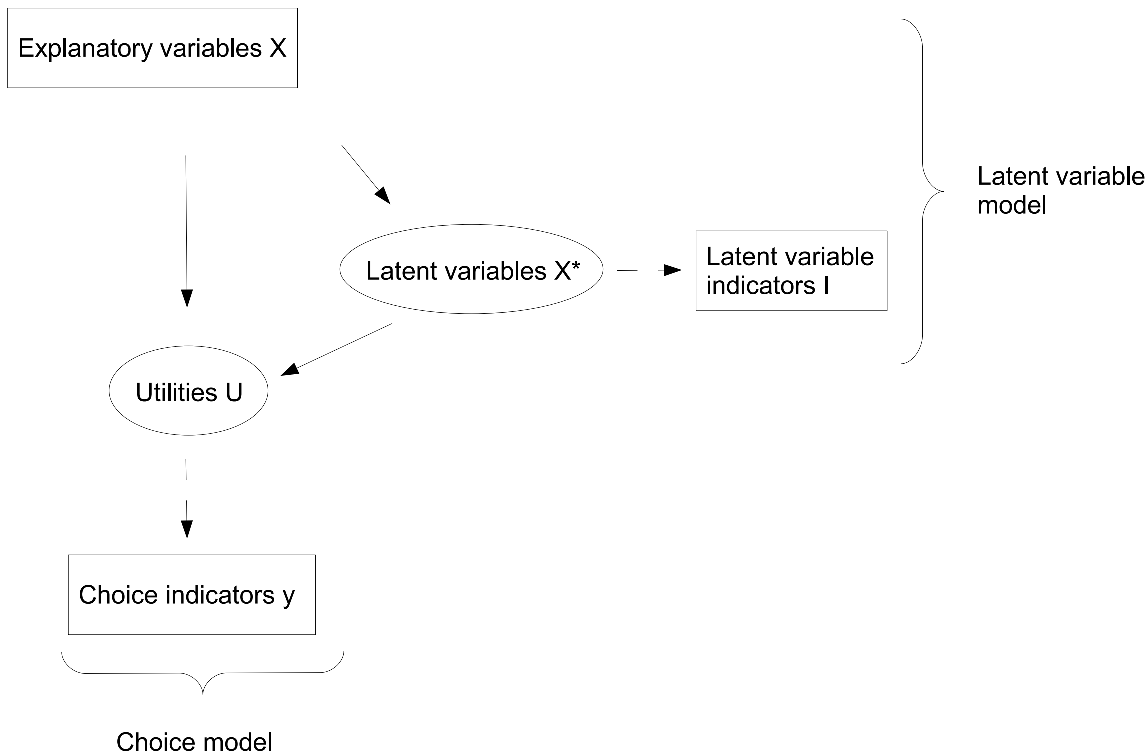

Figure 1: The integrated choice and latent variable model (Walker, 2001). Variables in rectangles are observable and variables in ovals are latent. The continuous arrows represent structural equations and the dashed ones are measurement equations.

Two types of structural equations are specified, one for the choice model and one for the latent variable model.

The structural equation for the choice model is given by the following relationship:

$$U_{in} = V_{in} + \varepsilon_{in}, \qquad \text{with } \varepsilon_{in} \sim \text{EV}(0,1) \tag{1}$$

In equation (1), U_{in} represents the utility of alternative i for respondent n. According to the random utility theory, each utility U_{in} is the sum of a deterministic term V_{in} and a random variable ε_{in} of an extreme value distribution with location parameter 0 and scale parameter 1. The deterministic part is a function $V : \mathbb{R}^{2p} \longrightarrow \mathbb{R}$ of observable explanatory variables X_{in}, latent variables X_n^* and a p-dimensional vector of parameters β. In other terms, we have

$$V_{in} = V(X_{in}, X_n^*; \beta). \tag{2}$$

The structural equation for the latent variable model is given by the following formula:

$$X_n^* = h(X_{in}; \lambda) + \omega_{in}, \qquad \text{with } \omega_{in} \sim \mathcal{N}(0, \sigma_\omega) \tag{3}$$

The latent variable X_n^* of equation (3) is expressed as a function $h : \mathbb{R}^{2l} \longrightarrow \mathbb{R}$ of observable explanatory variables X_{in} for an alternative i and an individual n, and an l-dimensional vector λ of coefficients. Similarly as for the structural equation of the choice model, we specify a random term ω_{in}. In this case, it is normally distributed with mean 0 and standard deviation σ_ω.

In the integrated choice and latent variable model, we specify measurement equations for the latent variables, as the latter cannot be obtained directly from the data. For each individual n and each alternative i, these equations express an indicator I_n as a function $m : \mathbb{R}^{2a} \longrightarrow \mathbb{R}$ of latent explanatory variables X_n^* and of a a-dimensional vector α, where a is the number of indicators of latent variables X_n^*. A random term ν_n of Logistic distribution with location parameter 0 and scale parameter 1 is added to function m. To summarize, the measurement equations are of the following form:

$$I_n = m(X_n^*; \alpha) + \nu_n, \qquad \text{with } \nu_n \sim \text{Logistic}(0,1) \tag{4}$$

In the model presented in this paper, we use *discrete indicators* for the latent variable. The reason is that the indicators used for measuring the latent variable can only take five integer values from -2 to 2 (see section 2.3 for more details).

For a discrete indicator with k levels $i_1, \ldots, i_k$, such that $i_1 < i_2 < \ldots < i_k$, the measurement equations are given by threshold functions:

$$I_n = \begin{cases} i_1 & \text{if } -\infty < X_n^* \leq \tau_1 \\ i_2 & \text{if } \tau_1 < X_n^* \leq \tau_2 \\ \cdots & \\ i_k & \text{if } \tau_{k-1} < X_n^* \leq +\infty \end{cases} \tag{5}$$

for every individual n. In other terms, we specify an *ordered logit model* for latent variable X_n^*. Parameters $\tau_1, \ldots, \tau_{k-1}$ are thresholds which need to be estimated.

The vectors of parameters β, λ, σ_ω and α, as well as threshold values $\tau_1, \ldots, \tau_{k-1}$ are estimated by maximum likelihood techniques. For each individual n, the joint probability of choosing alternative i and observing indicator I_n is given as follows:

$$f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega) = \int_{X_n^*} P(y_{in} | X_{in}, X^*; \beta, \sigma_\omega) \cdot f(I_n | X_{in}, X_n^*; \alpha, \sigma_\omega) \cdot f(X_n^* | X_n; \lambda, \sigma_\omega) dX_n^*,$$

where choice indicator y_{in} is defined as follows:

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} = \max_j U_{jn} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The following likelihood function is then estimated:

$$\mathcal{L} = \prod_{n=1}^{N} f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega). \tag{7}$$

Values for parameters β, λ, σ_ω, α and $\tau_1, \ldots, \tau_{k-1}$ are obtained by maximizing $\mathcal{L}$ and enable the modeler to evaluate the effect of characteristics of the alternative and/or the respondent on the choice. The software used for the estimation of the model of this paper is BIOGEME (Bierlaire, 2003). We used its extended version which is presented in Bierlaire and Fetiarison (2009).

4 Model specification and estimation

In section 3, we described the modeling framework we wish to apply in order to analyze the mode preferences of the individuals. In this section, we present its specification and estimation results.

4.1 Choice variable

Before providing the specification of the discrete choice model, we need to define the choice variable carefully. The latter is defined as the transport mode used by the respondents. It can be one of the three following categories:

- *Public transport modes*, such as bus, train, etc. This category is denoted as PT in the choice model.
- *Private transport modes*, such as car, motorbike, etc. This category is denoted as PM in the choice model.
- *Soft modes*, such as walk or bike. This category is denoted as SM in the choice model.

The respondents' mode choice are analyzed on *loops* and not on single trips, that is, one observation corresponds to one sequence of trips starting from each respondent's home and ending at the same place. For example, one simple loop can consist of a series of starting points and destinations *home-work-home.* A longer loop such as *home-work-shopping-home* can include an additional trip for shopping. In the data, $2'265$ loops were identified from the $1'763$ valid questionnaires.

Hence, the choice variable is defined as the set of modes, i.e. public transports, private modes or soft modes, used on each loop.

4.2 Model specification

The integrated choice and latent variable model is made of two models: a discrete choice model and a latent variable model. For the discrete choice model, the deterministic parts of the utility functions are given as follows:

$$V_{\mathrm{PM}} = \mathrm{ASC}_{\mathrm{PM}} + \beta_{\mathrm{cost}} \cdot \mathrm{cost}_{\mathrm{PM}} + \beta \mathrm{work} \cdot \mathrm{work} + \beta_{\mathrm{French}} \cdot \mathrm{French} + \beta_{\mathrm{time}_{\mathrm{PM}}} \cdot \mathrm{time}_{\mathrm{PM}}$$

$$V_{\mathrm{PT}} = \mathrm{ASC}_{\mathrm{PT}} + \beta_{\mathrm{cost}} \cdot \mathrm{cost}_{\mathrm{PT}} + \beta_{\mathrm{time}_{\mathrm{PT}}} \cdot \mathrm{time}_{\mathrm{PT}} + \beta_{\mathrm{comfort}} \cdot \mathrm{PerceptionComfortPT} \cdot \mathrm{HighUsagePT}$$

$$V_{\mathrm{SM}} = \mathrm{ASC}_{\mathrm{SM}} + \beta_{\mathrm{distance}} \cdot \mathrm{distance}$$

All three utility functions include constant terms, that is, parameters $\mathrm{ASC}_{\mathrm{PM}}$, $\mathrm{ASC}_{\mathrm{PT}}$ and $\mathrm{ASC}_{\mathrm{SM}}$ for the private, public and soft transport modes, respectively. The constant term $\mathrm{ASC}_{\mathrm{PT}}$ is fixed to 0. We assume that the deterministic utilities V_{PM} and V_{PT} of the private and public transport modes are influenced by the travel times $\mathrm{time}_{\mathrm{PM}}$ and $\mathrm{time}_{\mathrm{PT}}$, and travel costs $\mathrm{cost}_{\mathrm{PM}}$ and $\mathrm{cost}_{\mathrm{PT}}$. For the soft modes, a distance term distance is included.

In addition to the characteristics of the transportation alternatives, some socio-economic variables are assumed to have an impact on the transport mode choice, that is, a variable work which is an indicator that the respondent peformed home-work-home loops and a variable French indicating that the respondent resides in a French-speaking region of Switzerland.

In the deterministic utility V_{PT}, a latent explanatory variable PerceptionComfortPT is also included and accounts for the image people have of comfort in public transports. It is interacted with a variable HighUsagePT which indicates if the respondent uses public transportation at least once a week. By interacting the perception of comfort in public transports and a frequent use of the latter, we assume that travelers who often take public transports have a better image of their comfort.

The perception of comfort in public transports PerceptionComfortPT is described by the following structural equation:

$$\begin{aligned} \mathrm{PerceptionComfortPT} &= \lambda_{\mathrm{mean}} + \lambda_{\mathrm{German}} \cdot \mathrm{German} + \lambda_{\mathrm{age50}} \cdot \mathrm{age50} + \lambda \mathrm{active} \cdot \mathrm{active} \\ &+ \lambda \mathrm{cars} \cdot \mathrm{cars} + \omega, \qquad \text{with } \omega \sim \mathcal{N}(0, \sigma) \end{aligned} \tag{8}$$

In equation (8), we specify an intercept λ_{mean} and assume that several socio-economic variables have an effect on individual's perception of comfort in public transportation. The latter include an indicator German of a residence in a German-speaking region, a variable age50 indicating that the respondent is younger than 50 years, a variable active equal to 1 if the respondent has a full-time / part-time job, and 0 for another working status, and a variable cars indicating if the respondent's household owns at least 2 cars. A random variable ω of mean 0 and standard deviation σ is also added to the structural equation, in order to take into account effects which the deterministic part does not model.

As latent variable PerceptionComfortPT cannot be directly quantified by a survey question, measurement equations that relate it with indicators are specified. The indicators are the values ranging from -2 to 2 which were assigned to each adjective respondents reported for a list of transport modes, according to the procedure explained in section 2.3. In particular, we asked respondents to report three adjectives for the each of the following three types of public transports: train, bus/metro/tram and post bus.

This enables us to define 9 indicators, which can be written as follows, according to equation (5), for $k = 1, \ldots, 9$:

$$I^k = \begin{cases} -2 & \text{if } -\infty < \mathrm{PerceptionComfortPT} \leq \tau_1 \\ -1 & \text{if } \tau_1 < \mathrm{PerceptionComfortPT} \leq \tau_2 \\ 0 & \text{if } \tau_2 < \mathrm{PerceptionComfortPT} \leq \tau_3 \\ 1 & \text{if } \tau_3 < \mathrm{PerceptionComfortPT} \leq \tau_4 \\ 2 & \text{if } \tau_4 < \mathrm{PerceptionComfortPT} \leq +\infty \end{cases} \tag{9}$$

In practice, we do not estimate each τ_r, with $r = 1, \ldots, 4$, but define other variables δ_s, with $s = 1, \ldots, 3$, such that

$$\tau_2 = \tau_1 + \delta_1$$
$$\tau_3 = \tau_2 + \delta_2$$
$$\tau_4 = \tau_3 + \delta_3$$

with $\delta_s \geq 0 \ \forall s = 1, \ldots, 3$. Variables δ_s, with $s = 1, \ldots, 3$ are hence defined in order to have $\tau_1 < \tau_2 < \tau_3 < \tau_4$. Then the only parameters which must be estimated are τ_1 and all variables δ_s, with $s = 1, \ldots, 3$.

Each indicator I^k defined in equation (9), with $k = 1, \ldots, 9$, is related to latent variable PerceptionComfortPT via a measurement equation specified as follows:

$$I^k = \alpha_k \cdot \text{PerceptionComfortPT} + \nu^k, \qquad \text{with } \nu^k \sim \text{Logistic}(0, 1)$$

4.3 Estimation results

Parameters ASC_h with $h \in \{\text{PM}, \text{PT}, \text{SM}\}$, β_i, with $i \in \{\text{cost}, \text{time}_{PM}, \text{time}_{PT}, \text{distance}, \text{work}, \text{French}, \text{comfort}\}$, λ_j, with $j \in \{\text{mean}, \text{German}, \text{age50}, \text{active}, \text{cars}\}$, σ, α_k with $k = 1, \ldots, 9$, and finally τ_1 are estimated by maximizing likelihood function specified in equation (7).

Let us note that for normalization purposes, α_1 is fixed to 0.

The estimation results of the integrated choice and latent variable model are reported in Table 3. A first observation is that all parameters are significant at a 95% level of confidence, except τ_1. The significance of the latter has no importance, as it is only the value of a threshold. The following conclusions can be drawn from the estimates of the integrated choice and latent variable model:

- Regarding the choice model, the alternative specific constant ASC_{PT} was set to 0. This means that given the positive signs of constants ASC_{PM} and ASC_{SM}, individuals have an a priori preference for private transport modes, such as car, and soft modes over public transport modes. Moreover, ASC_{PM} is larger than ASC_{SM}, which shows that people have a greater preference for private modes over the soft ones.

- The negative sign of parameter β_{cost} shows that the higher the price of one loop is, the lower its utility is. The time coefficients $\beta_{\text{time}_{PM}}$ and $\beta_{\text{time}_{PT}}$ both have negative signs as well, indicating that travel durations affect the utilities of private modes and public transports in a negative way. Moreover, the time coefficient $\beta_{\text{time}_{PT}}$ for public transportation is smaller than the one for private modes. This shows that travel time is less important in the choice of public transports than in the choice of private modes. Finally, the total distance of a sequence of trips affects the choice of soft modes in a negative way. This can be seen in the negative sign of parameter β_{distance} and is consistent with the fact that travelers do not want to choose soft modes when they need to perform long distances.

- Some socio-economic variables have a significant effect on the choice of transport modes. First, when loops only include a trip from home to work and back from work to home are performed, public transportation is preferred. This can be seen in the negative sign of coefficient β_{work}. Second, the positive sign of β_{French} shows that, in French-speaking regions, inviduals have a preference for private modes. This could be due to the fact that the public transportation offer is not as developed as in German-speaking regions.

- The positive sign of coefficient PerceptionComfortPT shows that a good perception of comfort in public transports drives their selection, when the traveler is a frequent user of such modes. Let us recall that parameter PerceptionComfortPT was interacted with variable HighUsagePT, which indicates that the respondent uses public transportation at least once a week.

From the analysis of the estimates of the parameters of the structural equation of latent variable PerceptionComfortPT, we notice that individuals with full- or part-time jobs have a negative perception of comfort in public transports. Age is also a factor affecting the perception of comfort is such modes, i.e. people below 50 years have a more negative image of it. Leaving in a German-speaking region is positively correlated with a good perception of comfort in public transports. Finally, respondents with at least two cars in the household have a negative image of comfort in public transports.

	Name	Value	Robust std error	Robust t-test	p-value
1	$\mathrm{ASC}_{\mathrm{PM}}$	2.20	0.19	11.64	0.00
2	$\mathrm{ASC}_{\mathrm{SM}}$	1.74	0.34	5.16	0.00
3	β_{cost}	−0.02	0.01	−2.54	0.01
4	$\beta_{\mathrm{distance}}$	−0.21	0.05	−4.07	0.00
5	β_{work}	−0.47	0.12	−3.96	0.00
6	β_{comfort}	0.39	0.14	2.89	0.00
7	β_{French}	0.70	0.15	4.56	0.00
8	$\beta_{\mathrm{time}_{\mathrm{PM}}}$	−0.02	0.01	−3.91	0.00
9	$\beta_{\mathrm{time}_{\mathrm{PT}}}$	−0.01	0.00	−3.19	0.00
10	$\lambda_{\mathrm{active}}$	−0.29	0.08	−3.67	0.00
11	λ_{age50}	−0.28	0.07	−4.04	0.00
12	λ_{mean}	7.44	2.43	3.06	0.00
13	$\lambda_{\mathrm{German}}$	0.14	0.06	2.17	0.03
14	σ	0.22	0.06	3.82	0.00
15	λ_{cars}	−0.19	0.07	−2.70	0.01
16	δ_1	3.07	0.12	24.80	0.00
17	δ_2	4.72	0.06	80.11	0.00
18	δ_3	2.05	0.07	27.99	0.00
19	α_2	0.92	0.03	29.72	0.00
20	α_3	0.87	0.05	18.26	0.00
21	α_4	0.83	0.06	14.69	0.00
22	α_5	0.77	0.08	9.74	0.00
23	α_6	0.79	0.08	10.46	0.00
24	α_7	0.95	0.02	48.32	0.00
25	α_8	0.88	0.04	20.19	0.00
26	α_9	0.85	0.06	15.44	0.00
27	τ_1	0.73	2.46	0.30	0.77

Table 3: Estimates of the coefficients of the integrated choice and latent variable model that aims at evaluating demand for the three transportation modes, with corresponding standard errors, values of t-test and p-values.

5 Validation

From the estimations results presented in section 4.3, we can conclude that the perception of comfort in public transports has a significant impact on mode choice preferences.

Now we would like to assess the validation power of the model. As no other similar data set is available, we first estimate the model described in section 4.2 on 80% of the data and in a second phase, we apply it on the remaining 20%. The purpose of this procedure is to see if the choice probabilities computed for each observation of the 20% of the data are predicted correctly. By choice probability, we mean the probability predicted by the model that the respondent chooses the transport mode he reported.

For comparison purposes, a simple logit model with multiple alternatives was also estimated on the same data set with 80% of the observations and validated on the data set with 20% of the observations. Histograms of the choice probabilities predicted on 20% of the data are plotted in Figure 2 for both the logit model and the integrated choice and latent variable model. A choice probability close to 1 indicates that the alternative selected by the respondent is predicted by the model with a high confidence.

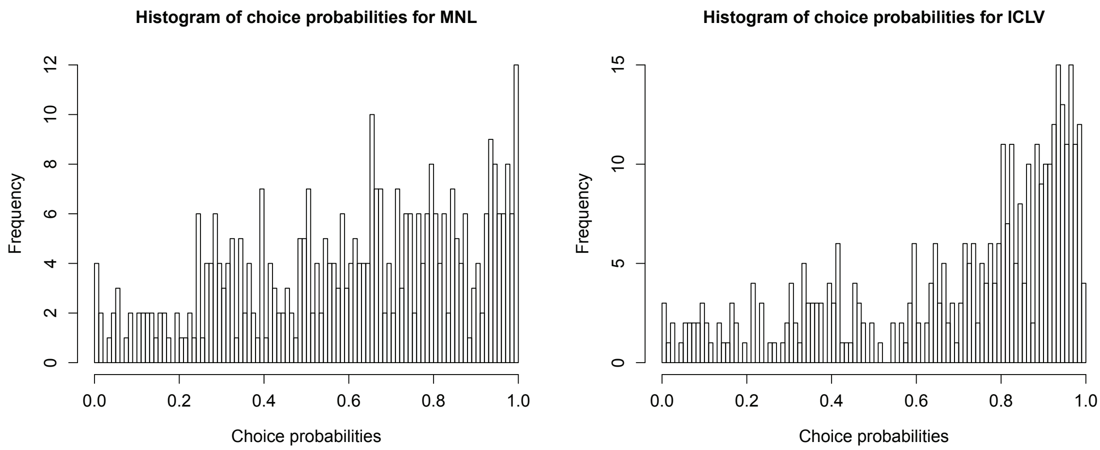

Figure 2: Histogram of the choice probabilities for the logit model with multiple alternatives (left) and for the integrated choice and latent variable model (right), resulting from the estimation of each model on 80% of the data and application on the remaining 20%.

We notice from Figure 2 that the integrated choice and latent variable model improves the predictive power of the model as more choice probabilities have moved from the left to the right of the graph. Table 4 shows the frequency of choice probabilities which are higher than 0.5 and 0.9 for both the logit model with multiple alternatives and the integrated choice and latent variable model. The inclusion of the latent variable of comfort in public transports has increased the proportion of choice probabilities above 0.5 from 67.2% to 74.9% and the proportion of choice probabilities above 0.9 from 17.9% from 30.1%.

Threshold	MNL	ICLV
0.5	67.2%	74.9%
0.9	17.9%	30.1%

Table 4: Frequency of choice probabilities higher than 0.5 and 0.9 both for the logit model with multiple alternatives (MNL) and for the integrated choice and latent variable model (ICLV).

An additional indicator of validity of the integrated choice and latent variable model is the value of $\bar{\rho}^2$. Similarly as in Table 4, the values of $\bar{\rho}^2$ together with the final values of log-likelihood are reported in Table 5. We notice that the fit of the integrated choice and latent variable model is better than the one of the logit model, which additionally supports the introduction of the perception of comfort in the choice model.

Value	MNL	ICLV
Log-likelihood	-1206.31	-9818.49
$\bar{\rho}^2$	0.420	0.554

Table 5: Values of final log-likelihood and $\bar{\rho}^2$ for the logit model (MNL) and the integrated choice and latent variable model (ICLV).

5.1 Analysis of demand indicators

The mode choice model presented in section 4.2 enables us to perform an analysis of demand for public and private transports. To do so, we compute two sorts of aggregate indicators: market shares and elasticities (see Table 6). These indicators were corrected in order to be representative of the whole population of the peri-urban regions where the survey was conducted. In particular, persons with a high education level, with ages between 40 and 79 years, or men had a higher response rate than other categories. This bias was corrected by introducing weights in the computation of market shares and elasticities. These weights were computed using the *iterative proportional fitting (IPF)* algorithm.

We report two types of aggregate elasticities, that is, *direct elasticities* and *cross elasticities*. We are interested in analyzing the percent change in the market shares of the alternatives of private, public and soft modes with respect to changes in a particular quantity $x_j \in \{\mathrm{cost_{PM}}, \mathrm{cost_{PT}}, \mathrm{time_{PM}}, \mathrm{time_{PT}}\}$.

Aggregate direct elasticities are computed using the following formula:

$$E^i_{x_i} = \frac{\sum_{n=1}^N w_n P_n(i) E^i_{x_{in}}}{\sum_{n=1}^N w_n P_n(i)},$$

where w_n is the corrective weight of observation n which was computed using the IPF algorithm, $P_n(i)$ is the probability that the individual who performed loop n chooses alternative i and $E^i_{x_{in}}$ is the disaggregate direct elasticity of the demand for observation n for variations in individual quantity x_{in}. This disaggregate elasticity is computed using the following expression:

$$E^i_{x_{in}} = \frac{\partial P_n(i)}{\partial x_{in}} \frac{x_{in}}{P_n(i)}.$$

To summarize, an aggregate direct elasticity denotes the percent change in the market share for alternative i with respect to a change of 1% in the value of an attribute x_i of i.

Aggregate cross elasticities are given by the following expression:

$$E^i_{x_j} = \frac{\sum_{n=1}^{N} w_n P_n(i) E^i_{x_{jn}}}{\sum_{n=1}^{N} w_n P_n(i)},$$

where $E^i_{x_{jn}}$ is the elasticity of the demand for observation n for variations in individual quantity x_{jn}. This disaggregate elasticity is computed using the following expression:

$$E^i_{x_{jn}} = \frac{\partial P_n(i)}{\partial x_{jn}} \frac{x_{jn}}{P_n(i)}.$$

An aggregate direct elasticity hence represents the percent change in the market share for alternative i with respect to a change of 1% in the value of an attribute x_j of another alternative j.

Table 6 shows that private modes have the highest market share. They are followed by public transports.

From an analysis of the elasticities, we can conclude that demand is little elastic to changes in travel time or cost. The cost elasticity for public transportation is very low (-0.07) indicating that an increase of 1% in the cost of travel fares results in a decrease of 0.07% of the market share of public transports. This is also the case for the cost elasticity for private modes (-0.02), which shows that an increase of 1% in the travel costs (e.g. which could be driven by an increase of the petrol price) only induces a decrease of 0.02% in the market shares of private modes. Similarly, cross elasticities with respect to changes in cost of private and public modes are rather low, ranging between 0.01 and 0.05.

The demand elasticities for public transports (-0.32) and private modes (-0.15) relative to time are higher than the ones relative to cost, but still very low. Indeed, an increase of 1% in the duration of a loop of public transports results in a decrease of 0.32% of their market share. For private modes, the same increase of the travel time induces a decrease of 0.15% in their market share. The cross elasticities of demand relative to changes in travel time in private or public modes are also higher than the ones relative to changes in cost. For instance, the market share of soft modes can increase of 0.14% if the travel cost of public modes increases of 1%.

Indicator	PM	PT	SM
Market share	65.2%	28.5%	6.28%
Elasticity for cost of PM	-0.02	0.05	0.02
Elasticity for cost of PT	0.03	-0.07	0.01
Elasticity for time of PM	-0.15	0.32	0.14
Elasticity for time of PT	0.14	-0.32	0.05

Table 6: Market shares for private modes, public transports and soft modes, and direct and cross elasticities of demand relative to variations in cost and time for private modes and public transports.

5.2 Impact of an increase of the comfort level

After observing the significant impact of the perception of comfort in public transports (see section 4.3), we would like to test how the market shares of the different transport modes vary with respect to a change in this perception. As an example, we analyze the variation of the market shares when the individuals' perception of comfort is increased by 50%. The obtained market shares are reported in Table 7, as well as the ones predicted by the model without any increase of the perception of comfort. We notice that such an increase shifts the market share of public transports from 28.5% to 36.1% and hence shows that changes in the perception of comfort in these modes can impact on their choice in a non-negligible way.

Perception of comfort	PM	PT	SM
Without increase	65.2%	28.5%	6.28%
With a 50% increase	58.3%	36.1%	5.64%

Table 7: Market shares for private modes, public transports and soft modes, without and with an increase of 50% in the perception of comfort in public transports.

6 Conclusion and further works

With the model presented in this paper, we could show that in addition to travel times and costs, and socio-economic information of the respondent, perceptions can have a significant impact on mode choice. This is the case for the latent variable of a positive perception of comfort in public transports.

The novelty of this approach also lies in the fact that the data used to model perception are originally adjectives freely reported by respondents. In order to quantify them, a classification into categories followed by an attribution of a perception scale were performed.

The validation of such integrated choice and latent variable shows a large improvement over the validation of a logit model with multiple alternatives, which assesses a good prediction power.

Despite some promising modeling results, the transformation of the qualitative data consisting of adjectives into quantitative indicators is subjective. Different modelers might indeed give slighly different ratings to the same adjective. In our future research we plan to improve the mapping from the adjective data to the indicator values.

In this paper, we analyzed the impact of the perception of comfort in public transports. In addition to this latent variable, the perception of comfort in private modes could also be introduced. Moreover, further improvements could consist of the inclusion of other latent perceptions, such as the image of reliability, of security or of the environmental impact.

The perception of comfort in public transports was interacted with a frequent use of public transports. In order to better capture differences between frequent riders and other travelers, a latent class model could be developed. Such resulting model could hence combine the integrated choice and latent variable model with a latent class model such as in the generalized random utility framework (Walker and Ben-Akiva, 2002).

7 Acknowlegdements

The authors would like to thank CarPostal for their advice and feedback on the collaborative project which was set up and for funding it. They also would like to acknowledge the Urban Sociology Laboratory (LaSUR) and the Urban and Regional Planning Laboratory

(CEAT) who co-conducted the qualitative and the revealed preference surveys which serve as a basis to this research.

References

M. Abou-Zeid, M. Ben-Akiva, and M. Bierlaire. Happiness and travel behaviour modification. In *European Transport Conference (ETC)*, October 2008.

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

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

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

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

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

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

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

V. Kaufmann, C. Jemelin, and J.-M. Guidez. *Automobile et modes de vie urbains quel degré de liberté?* Documentation Française, 2001.

V. Kaufmann, K. Tabaka, N. Louvet, and Guidez J.-M. *Et si les Français n'avaient plus seulement une voiture dans la tête.* Certu, 2010.

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

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

J. L. Walker. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables.* PhD thesis, Massachusetts Institute of Technology, 2001.