Bahamonde-Birke, Francisco J.; Kunert, Uwe; Link, Heike; de Dios Ortúzar, Juan

**Working Paper**

# About attitudes and perceptions: Finding the proper way to consider latent variables in discrete choice models

DIW Discussion Papers, No. 1474

**Provided in Cooperation with:**
German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Bahamonde-Birke, Francisco J.; Kunert, Uwe; Link, Heike; de Dios Ortúzar, Juan (2015) : About attitudes and perceptions: Finding the proper way to consider latent variables in discrete choice models, DIW Discussion Papers, No. 1474, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/110318

DIW BERLIN

1474

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2015

# About Attitudes and Perceptions – Finding the Proper Way to Consider Latent Variables in Discrete Choice Models

Francisco J. Bahamonde-Birke, Uwe Kunert, Heike Link, and Juan de Dios Ortúzar

IMPRESSUM



DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

# About Attitudes and Perceptions – Finding the Proper Way to Consider Latent Variables in Discrete Choice Models

**Francisco J. Bahamonde-Birke**[*]

**Uwe Kunert, Heike Link**[†]

**Juan de Dios Ortúzar**[‡]

**Abstract**

We provide an in-depth theoretical discussion about the differences between attitudes and perceptions, as well as an empirical exercise to analyze its effects. This discussion is of importance, as the large majority of papers considering attitudinal latent variables, just consider those as attributes affecting directly the utility of a certain alternative while systematic taste variations are rarely taken into account and perceptions are normally completely ignored.

The results of our case study show that perceptions may indeed affect the decision making process and that they are able to capture a significant part of the variability that is normally explained by alternative specific constants. In the same line, our results indicate that attitudes may be a reason for systematic taste variations, and that a proper categorization of the latent variables, in accordance with the underlying theory, may outperform the customary assumption of linearity.

***Keywords:*** *Hybrid Discrete Choice Modelling, Latent Variables, Attitudes, Perceptions*

***JEL****: C50*

[*] Energy, Transportation and Environment Department, Deutsches Institut für Wirtschaftsforschung – Berlin, and Technische Universität Berlin. E-mail: bahamondebirke@gmail.com

[†] Energy, Transportation and Environment Department, Deutsches Institut für Wirtschaftsforschung – Berlin. E-mails: ukunert@diw.de; hlink@diw.de

[‡] Department of Transport Engineering and Logistics, Centre for Urban Sustainable Development (CEDEUS) Pontificia Universidad Católica de Chile. E-mail: jos@ing.puc.cl

## 1. INTRODUCTION

The last decades have seen discrete choice models (DCM) become a key element in travel demand modelling and forecasting (Ortúzar and Willumsen, 2011). Their current state-of-practice considers objective characteristics of the alternatives and the individuals as explanatory variables and yield as output individual probabilities of choice between different alternatives. It is also well known that attitudes and perceptions play a role in the decision making process, and the usual approach to take these into account considers the estimation of a Multiple Indicator MultIple Cause (MIMIC) model, as suggested by Bollen (1989). The joint use of MIMIC models and DCM leads to state-of-the-art hybrid discrete choice (HDC) models (Ben-Akiva *et al.*, 2002; Bolduc and Alvarez-Daziano, 2010; Bahamonde-Birke and Ortúzar, 2014a).

In the last years, the literature has provided abundant empirical and theoretical evidence about the advantages of this approach and the use of HDC models has gained popularity (v. Acker *et al.*, 2011; Ashok *et al.*, 2002; Bahamonde-Birke *et al*., 2010; Raveau *et al.,* 2012; Daziano and Bolduc, 2013; among others). Notwithstanding, attitudes and perceptions are usually addressed as a whole, not considering that both are expressions of different value judgments. This way, attitudes express a characteristic of the individuals toward life, society, etc. and are intrinsically related to them, playing a role in every decision made. Perceptions, instead, are exclusively related to the way certain alternatives are perceived. This being the case, an attitude resembles a socio-economic characteristic of the individual, while a perception is intrinsically associated with an alternative.

This difference has important implications and the way in which both should be treated in a discrete choice model is completely different. In turn, this issue affects not only the fashion in which latent variables are estimated but almost all hypotheses concerning them. Hence, different assumptions will have an effect on both the way the latent variables are constructed

through the MIMIC model as well as on the manner in which these constructs are reflected in the utility function of the DCM.

The rest of the paper is organized as follows. Section 2 offers a theoretical overview of HDC models, while Section 3 presents an extensive discussion about the different ways to consider latent variables in DCMs. Section 4 describes an experiment carried out to test the hypotheses of the previous section, and its results are discussed in section 5. Finally, section 6 summarizes our conclusions.

## 2. THEORETICAL BACKGROUND

Under the assumption that individuals are rational decision makers, it can be postulated that individuals $q$ facing a set of available alternatives $A(q)$, will choose the alternative $i$ that maximizes their perceived utility. In accordance with Random Utility Theory (Thurstone, 1927; McFadden, 1974), it is possible to depict this utility as the sum of a representative component ($V_{iq}$) and an error term ($\varepsilon_{iq}$), which leads to the following expression (Ortúzar and Willumsen, 2011):

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{2.1}$$

The representative utility ($V_{iq}$), considering all attributes that can be quantified by an observer, is usually characterized through concrete and measurable properties of the alternatives and the individuals; the error term, in turn, is considered to take into account all unknown or abstract elements affecting the decision.

When considering a Hybrid Discrete Choice (HDC) modelling framework (Ben-Akiva *et al.*, 2002), the modeller attempts to depict abstract attributes as measurable variables in order to include them as part of the systematic utility. Hereby, immaterial constructs, known as latent variables ($\eta_{liq}$), are also included into the modelling. These variables are supposed to represent attitudes and/or perceptions of the individuals and, as they cannot be directly

observed, they must be constructed as a function of positively observed variables. The usual approach to construct these latent variables relies on a MIMIC structure (Zellner, 1970; Bollen, 1989). Here, 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, at the same time, a set of attitudinal and/or perceptual indicators ($y_{ziq}$), previously gathered from the individuals, through so called *measurement equations*. This framework can be represented through the following equations:

$$\eta_{liq} = \sum_r \alpha_{lri} \cdot s_{riq} + \upsilon_{liq} \tag{2.2}$$

$$y_{ziq} = \sum_l \gamma_{lzi} \cdot \eta_{liq} + \varsigma_{ziq} \tag{2.3}$$

where the indices $i$, $q$, $r$, $l$ and $z$ refer to alternatives, individuals, exogenous variables, latent variables and indicators, respectively. The error terms $v_{liq}$ and $\zeta_{ziq}$ can follow any distribution but they are typically assumed to distribute Normal with mean zero and a certain covariance matrix. Finally, $\alpha_{lri}$ and $\gamma_{lzi}$ are parameters to be jointly estimated.

If we assume a linear specification in $V_{iq}$, the utility function can be expressed as (2.4). This specification can be understood as a first-order Taylor expansion of any multi-variable complex function (and therefore it is always valid in the neighbourhood of the estimation point); further, if the attributes are also assumed to be linear, the estimated parameters $\theta_{ik}$ and $\beta_{il}$ (related to the tangible attributes and latent variables, respectively) can be directly interpreted as marginal utilities:

$$U_{iq} = \sum_k \theta_{ki} \cdot X_{kiq} + \sum_l \beta_{li} \cdot \eta_{liq} + \varepsilon_{iq} \tag{2.4}$$

Under the assumption that the error terms $\varepsilon_{iq}$ in (2.1) 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 $\lambda$, leading to the well-known Multinomial Logit (MNL) model (Domencich and McFadden, 1975); in this case, the probability of choosing alternative *i* is given by:

$$P_{iq} = \frac{e^{\lambda \cdot V_{iq}}}{\sum_j e^{\lambda \cdot V_{jq}}} \tag{2.5}$$

and $\lambda$ is inversely related to the standard deviation of the error terms:

$$\lambda = \frac{\pi}{\sigma\sqrt{6}} \tag{2.6}$$

However, as the scale factor cannot be estimated (assuming a linear function as usual), it is customary to *normalize* it to one (Walker, 2002).

The estimation of both parts of the model should be performed simultaneously, as a sequential estimation considering first the MIMIC part as an isolated system and evaluating afterwards the expected values for the latent variables cannot guarantee consistent and unbiased estimators (Train *et al*., 1987; Ben-Akiva *et al*., 2002). However, empirical evidence sustains the thesis that the sequential estimation produces no major discrepancies regarding the ratios between the estimated parameters and, therefore, the marginal rates of substitution (Raveau *et al*., 2010; Bahamonde-Birke *et al.*, 2010). Nevertheless Bahamonde-Birke and Ortúzar (2014a, 2014b) prove that the estimators may indeed be affected by a significant deflation bias (affecting all estimated parameters).

An intermediate alternative between the simultaneous and classical sequential estimation consists of estimating the model sequentially, but taking into account the variability of the latent variables. Despite the fact that this approach also requires integrating over the domain of the latent variables, it offers significant advantages in terms of computation costs. This approach leads to consistent but inefficient estimators (Ben-Akiva *et al*., 2002) and avoids the bias described by Bahamonde-Birke and Ortúzar (2014b).

## 3. ABOUT ATTITUDES AND PERCEPTIONS

Prior to discussing the different ways in which latent variables may be considered in a HDC-model, it is necessary to understand the difference between attitudes and perceptions. The former may be considered as a mind-set or a tendency to act in a particular way based on the individual's experience and temperament (Allport, 1935; Pickens, 2005). Therefore, in our approach indicators representing these attitudes depend only on the individuals and are considered constant for all alternatives. Thus, one set of attitudinal indicators will be enough to represent all decisions taken by the individual in question. Contrariwise, perceptions (although closely related to attitudes) may be interpreted as the process by which individuals experience their environment (Lindsay and Norman, 1972) and depend, therefore, on both the person and the stimuli (Pickens, 2005). As a corollary, perceptual indicators should be a function of both the individual and the alternatives. Even more, any variation in the alternatives may lead to a different valuation of them, as every detail may affect the way in which the population perceives the various alternatives. Therefore, in order to work with perceptions and perceptual indicators, it is necessary to gather a new set of indicators for every alternative the individual faces.

This issue can lead to a significant increase in the information required, as normally the alternatives would consist of different attributes that are subject to variations. Therefore, it is mandatory to make certain simplifying assumptions. Firstly, it may be assumed that certain

attributes will not affect the way in which a given alternative is perceived and, consequently, this dimension may be excluded from the design (e.g. price and accessibility indicators). In the same line, it can be assumed that the model is valid across individuals (avoiding the need that everyone states his/her perceptions for every combination of possible attributes, as long as other individuals are faced with the remaining combinations).

Once the indicators are gathered it is possible to construct the latent variables. Obviously, attitudinal variables will be related to attitudinal indicators and *vice versa*. Thus, while attitudinal variables must be solely explained by characteristics of the individuals (as no variation across alternatives will be observed), perceptions should be also explained by the attributes of the different alternatives considered in the experimental design.

For estimation purposes, different sets of perceptual indicators associated with different alternatives (in terms of the attributes that constitute them) may be treated jointly or separately. If the sets of indicators are considered jointly, the estimated MIMIC model will hold for any variation of the alternatives and the estimated latent variables should be comparable among each other. An isolated estimation would offer a better goodness-of-fit, but the estimated latent variables would not be longer comparable. This must be taken into account when estimating the utility function as it can lead to considerable trouble if the alternatives cannot be isolated in the DCM. Note that using common estimators (and a common structure) for the measurement equations overcomes the difference between both approaches. When more than one set of indicators per person is considered, correlation among individuals must be taken into account.

Further, the treatment of both kinds of variables (attitudes and perceptions) in the DCM should not be equal and some attitudes, just as socio-economic variables, should be considered through systematic taste variations and not directly in the utility function, as they affect the way in which the attributes of the alternatives are perceived.

That being the case, it is possible to identify three kinds of latent variables:

a) Non-alternative related attitudes: Most researchers working with HDC models consider this kind of variables (Bolduc and Alvarez-Daziano, 2010; Acker *et al.,* 2011, among many others). They represent general attitudes of the individuals toward their social and physical environment, such as a more ecological mind-set or a higher valuation of social status. Even when using variables that may be understood as perceptions, such as comfort, security, etc., the modeller is, in fact, dealing with a non-alternative related attitude, as in this case these variables stand for the importance assigned by the individuals to these aspects and not for a perception of the alternative itself. This way, inferences such as "Alternative A is perceived as more comfortable" would not be accurate but rather "individuals caring for comfort favour alternative A", which is not equivalent. Chorus and Kroesen (2014) argue (rightly) that this kind of models does not allow deriving policy implications, as these attitudinal variables are intrinsic characteristics of the individuals (like sex or age) and are therefore not sensitive to changes in the alternatives.

   On the other hand, as these variables resemble socio-economic characteristics, for the DCM to be identified they must be considered together with alternative specific attributes in the utility function. However, in most reported cases they are just considered in conjunction with alternative specific constants (Vredin-Johansson *et al.,* 2006; Bolduc *et. al*, 2008); this restriction may neglect important aspects of the decision, as it can be expected that individuals with different attitudes toward life exhibit a different valuation of the attributes of the alternatives, and therefore systematic taste variation should be allowed for (Ortúzar and Willumsen, 2011, page 279).

   Further, as in the case of socio-economic variables, it is not clear that attitudes should have a linear impact over the utility. Therefore, a categorization of the latent variables should be considered. This way, for instance, it is plausible that a low or

intermediate appraisal of security or safety may have no effect over the decision whatsoever, but a high concern could lead to a significantly different valuation. If this is the case, treating the variable linearly would not properly reflect the behaviour of the individuals.

Categorizing the latent variables offers also significant advantages in terms of flexibility, as it allows estimating different utility functions for every category, resembling a latent class model, but expanding it in order to account for the behavioural information.

b) Alternative related attitudes: These variables are similar to the abovementioned variables with the exception that attitudes are unequivocally related to a given alternative. Thus, these variables must be considered in conjunction with the alternative specific constant. For instance, Daziano and Barla (2012) considered the effect of a favourable predisposition toward automobiles or transit systems in this way. As in the previous case, systematic taste variations (within the same alternative) and a possible categorization should also be analysed.

c) Perceptions: These variables are alternative related, i.e. they exhibit a different valuation depending on the alternative considered and as such they resemble observed attributes of the alternatives such as price or travel time; hence, both kind of variables should be treated in the same fashion. Alternative specific and generic estimators may be considered.

Regarding the model's identifiability, necessary and sufficient conditions have not yet been developed (Ben-Akiva *et al.*, 2002). This way, most studies relying on HDC models achieve identification by not letting some explanatory variables impact the utility of a given alternative, both directly and through a latent variable (Bhat, 2014). This is indeed a sufficient but very restrictive condition and, especially when dealing with perceptual latent variables, the modeller may be forced to employ the same attributes in the structural equations as well as in the utility functions to represent behaviour properly (e.g. air conditioning may have an

effect over perceived comfort, but still have a direct impact on the decision due to other considerations). Under these circumstances identification must be analysed on a case-by-case basis (see Ben-Akiva *et al.*, 2002, for a good discussion on the identifiability of HDC models).

## 4. STUDY CASE

Departing from usual practice (which considers, typically, only attitudes), we developed an experiment considering both attitudinal and perceptual indicators in a transport choice framework. This allows testing for the more appropriate manner to consider both kinds of latent variables in a DCM, regarding the underlying theoretical concerns.

We conducted a stated choice (SC) experiment where respondents were asked to choose between different interurban public transport alternatives in Germany (regional[§] and intercity trains, and interurban coaches). The experiment was carried out in three waves (January 2014, March 2014 and April/May 2014), contacting both students and employees of the two universities in Berlin (the Technische Universität Berlin and the Humboldt-Universität zu Berlin), as well as employees of member institutions of the Leibniz-Gemeinschaft[**]. After data cleaning the survey yielded a total of 1,832 responses.

The questionnaire had four parts. In the first, respondents were asked to describe the main characteristics (fare, travel time, number of transfers, etc.) of their last trips with the regional and intercity trains of Deutsche Bahn. At the end of this module and based on their experience travelling with Deutsche Bahn (i.e. considering the same kind of trains and the same number of transfers of the journey described), participants were required to state their level of agreement with the following statements:

[§] Regional trains should not be confused with commuter rail. In Germany, regional trains operate over long interurban distances, stopping more and over shorter distances than intercity trains. It is possible to travel across the country using only regional trains.

[**] The Leibniz-Gemeinschaft is one of the shelter associations of publicly funded research institutes in Germany.

| | |
|---|---|
| *I was able to relax during the trip ($y_{11}$)* | ***Relax*** |
| *I felt secure from thefts and losses ($y_{12}$)* | ***Security*** |
| *Traveling with heavy luggage was (would have been) uncomplicated ($y_{13}$)* | ***Luggage*** |
| *The departure time was reliable ($y_{14}$)* | ***Departure*** |
| *The arrival time was reliable ($y_{15}$)* | ***Arrival*** |
| *It was possible to use the travel time productively ($y_{16}$)* | ***Productivity*** |
| *The station was easily accessible ($y_{17}$)* | ***Station*** |
| *Purchasing the ticket was uncomplicated ($y_{18}$)* | ***Tickets*** |

In the same line, respondents were also asked to state their level of agreement with these statements under the assumption that a bus carrier with no transfers would offer the service. The level of agreement was stated on a scale which ranged from strongly disagree (1) to strongly agree (10).

The second part of the survey gathered travel behaviour data as well as indicators related to traveller's attitudes toward current political issues discussed in Germany. Hereby, the respondents had to state the level of agreement with the following sentences:

| | |
|---|---|
| *I agree with the nuclear power phase-out ($y_{21}$)* | ***NuclearPhaseOut*** |
| *Environment protection is more important than economic growth ($y_{22}$)* | ***Environment*** |
| *I am willing to pay a 25% surcharge on my electric bill to reduce $CO_2$ emissions from coal power plants ($y_{23}$)* | ***ElectricSurcharge*** |
| *Highway tolls should be introduced to compensate $CO_2$ emissions ($y_{24}$)* | ***HighwayTolls*** |
| *Automobiles with higher engine power should pay more taxes ($y_{25}$)* | ***CarTax*** |
| *Investing on the development of high-speed trains should be encouraged ($y_{26}$)* | ***HSTrains*** |
| *New highways or additional lanes to the existing ones should be built ($y_{27}$)* | ***Highways*** |
| *New high-speed rail lines should be built ($y_{28}$)* | ***RailLines*** |
| *I agree with the introduction of speed limits on highways ($y_{29}$)* | ***SpeedLimits*** |

The third part of the questionnaire was the SC experiment itself. Here, respondents were required to choose between a first pivotal alternative, representing the trip previously described, and a new travel alternative. Altogether, respondents were confronted with twelve choice situations, where the first six used a pivotal alternative based on the trip with the Deutsche Bahn regional trains and the last six considered a trip with Deutsche Bahn intercity trains. Alternatives were described in terms of their travel time, fare, number of transfers,

mode of transport - regional trains (RE), intercity trains (FVZ) and coaches (LB) - and a safety level (represented through the number of severely injured passengers and the number of fatalities in the overall network over a year). Finally the fourth part of questionnaire gathered socioeconomic information about the respondents.

## 5. MODEL ESTIMATION

### 5.1 Model Structure

Before starting with the estimation of HDC models, it was necessary to establish the structure of the MIMIC-model considered. For this, the indicators were analysed using factor analysis to guarantee a correct specification of the latent variables (LV). This way, it was possible to identify three components explaining 70% of the variance of the perceptual indicators ($y_{11}$ to $y_{18}$). In the same way, it was possible to establish that two variables captured 54% of the variability associated with the attitudinal indicators ($y_{21}$ to $y_{29}$). Table 1 presents the rotated component matrices for both types of indicators. On the basis of these results, we constructed five latent variables, as highlighted in Table 1. The first was identified as “Comfort”, as it was exclusively related to comfort indicators. The second component was called “Stress-free”, as it was associated with situations causing tension during the trip. Finally, the third component was identified as “Reliability”.

Table 1 – Rotated Component Matrix of Perception and Attitudinal Indicators

| Indicator | Comfort | StressFree | Reliability |
|---|---|---|---|
| *Relax* | **0.548** | **0.591** | 0.171 |
| *Security* | 0.144 | **0.782** | 0.132 |
| *Luggage* | 0.061 | **0.810** | 0.178 |
| *Departure* | 0.117 | 0.245 | **0.892** |
| *Arrival* | 0.280 | 0.125 | **0.867** |
| *Productivity* | **0.663** | 0.432 | 0.099 |
| *Station* | **0.810** | 0.064 | 0.177 |
| *Tickets* | **0.711** | 0.059 | 0.153 |

| Indicator | Green | TrainFan |
|---|---|---|
| *NuclearPhaseOut* | 0.688 | -0.029 |
| *Environment* | 0.726 | -0.074 |
| *ElectricSurcharge* | 0.704 | 0.030 |
| *HighwayTolls* | 0.658 | 0.214 |
| *CarTax* | 0.686 | 0.192 |
| *HSTrains* | 0.114 | **0.860** |
| *Highways* | -0.546 | 0.365 |
| *RailLines* | 0.046 | **0.891** |
| *SpeedLimits* | 0.610 | 0.082 |

Regarding the attitudinal indicators, the first component was associated with a "Green" attitude, including a negative predisposition toward automobiles ($y_{24}$, $y_{25}$, $y_{27}$ and $y_{29}$). The second component related to individuals who have great appreciation for the development of trains and rail lanes (for this reason, this LV was called "TrainFan"). These results are interesting for our analysis as it was possible to identify a non-alternative ("Green") and an alternative ("TrainFan") related attitude.

### 5.2 MIMIC models

Given the complex structure and size of the data set (1,832 individuals; 3,900 sets of perceptual indicators; eight latent variables and 13,138 observed decisions), it was not computationally possible to perform a simultaneous estimation of the HDC model[††]. In addition, we wanted to analyse the effect of attitudinal latent variables both as continuous and as categorized variables, which complicated the structure of the model even more.

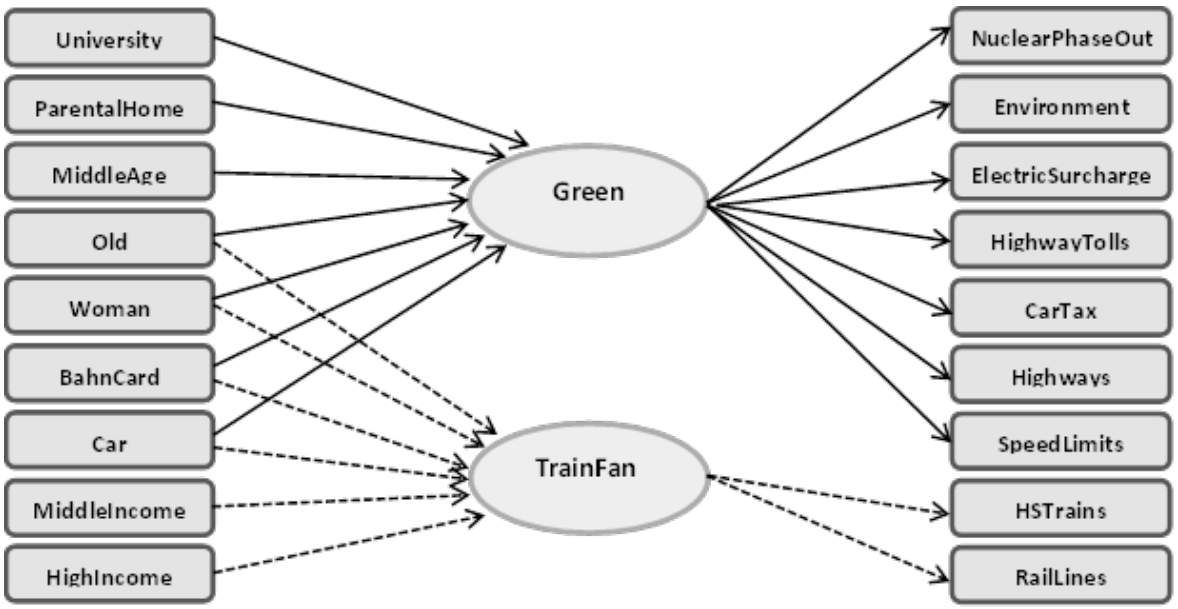


Figure 1 – Structure of the Attitudinal MIMIC model[‡‡]

[††] Attempts in this regard have been carried out, observing that the optimization algorithms interrupt the computation without reaching a convergence after approximately three weeks (*unsuccessful linesearch*).
[‡‡] The different line types are only used to ease the understanding of the figure.

Therefore, a sequential estimation considering the own variability of the latent variables and integrating over their domain was attempted. Thus, the MIMIC model was estimated first and the latent variables considered in the DCM component were constructed accordingly to these estimates. In fact, it was necessary to estimate two different MIMIC models. First, one for the attitudinal variables, which only considered individual characteristics as explanatory variables. Figure 1 presents the final structure of the selected model (several specifications were considered).

In this case, "University" is associated with workers holding this educational degree. "Parental Home" only applies to students and it indicates that the individual still lives at the parental home. "BahnCard" indicates that the individual holds a Deutsche Bahn yearly discount card (which is common in Germany due to the price discrimination policies adopted by Deutsche Bahn AG), while "Car" indicates automobile ownership. The remaining variables are self-explanatory. Table 2 presents their estimated parameters.

Table 2 – Estimated Parameters for the Attitudinal MIMIC model

| Explanatory Variable | Estimate | t-test | Attitudinal Indicator | Estimate | t-test |
|---|---|---|---|---|---|
| ***Green Attitude*** | | | ***Green Attitude*** | | |
| ***University*** | 0.258 | 4.134 | ***NuclearPhaseOut*** | 1.463 | 44.416 |
| ***ParentalHome*** | -0.181 | -2.815 | ***Environment*** | 1.178 | 49.917 |
| ***MiddleAge*** | 0.298 | 6.124 | ***ElectricSurcharge*** | 1.666 | 51.02 |
| ***Old*** | 0.497 | 3.713 | ***HighwayTolls*** | 2.14 | 51.02 |
| ***Woman*** | 0.287 | 5.917 | ***CarTax*** | 1.628 | 46.489 |
| ***BahnCard*** | 0.334 | 6.565 | ***Highways*** | -1.053 | -37.228 |
| ***Car*** | -0.524 | -10.075 | ***SpeedLimits*** | 2.282 | 51.902 |
| ***TrainFan*** | | | ***TrainFan*** | | |
| ***Old*** | 0.282 | 2.048 | ***HSTrains*** | 2.088 | 49.695 |
| ***Woman*** | -0.282 | -5.639 | ***RailLines*** | 2.12 | 49.937 |
| ***BahnCard*** | 0.333 | 6.356 | | | |
| ***Car*** | -0.058 | -1.116[§§] | | | |
| ***MiddleIncome*** | 0.141 | 2.807 | | | |
| ***HighIncome*** | 0.128 | 1.66[***] | | | |

[§§] The variable was kept in the model as it is considered a policy variable and has the proper sign.
[***] As the signs of the estimators were known *a priori*, a one-tailed test was performed ($\alpha_{5\%}$=1.645).

As stated above, a second MIMIC model was estimated for the perceptual indicators. In this case, not only the characteristics of the individuals but also the attributes of the transport modes were considered as explanatory variables. It is also important to consider interactions between these two kinds of variables, as different population groups perceive differently the attributes of the alternatives (i.e. systematic taste variations).

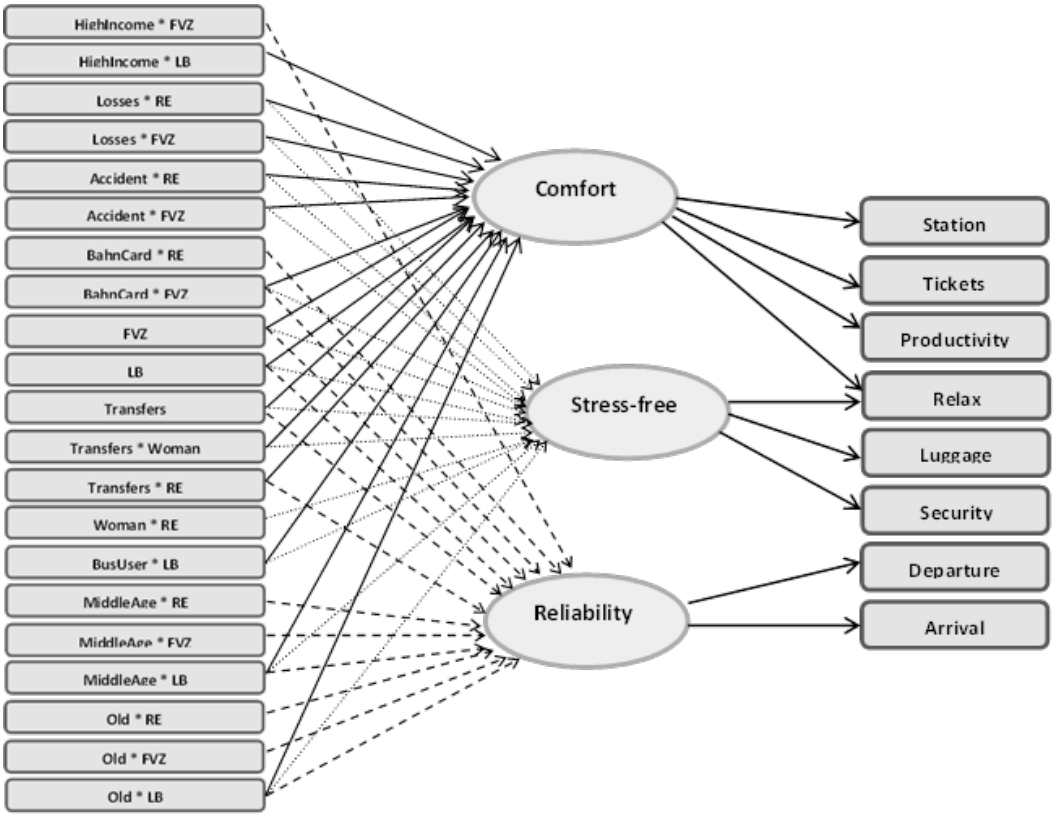


Figure 2 – Structure of the Perception MIMIC model

The structure of the estimated model is shown in Figure 2. Here, “Losses” and “Accidents” indicate that the individual had suffered losses during a trip in the past or had been involved in a train accident, respectively. The number of transfers is represented by a discrete variable ranging between zero and four, while “BusUser” indicates whether the individual had undertaken at least one trip with coach services during the last three years. The estimation results are presented in Table 3.

Table 3 – Estimated Parameters for the Perception MIMIC model

| Explanatory Variable | Estimate | t-test | Attitudinal Indicator | Estimate | t-test |
|---|---|---|---|---|---|
| ***Comfort*** | | | ***Comfort*** | | |
| ***HighIncome * LB*** | -0.249 | -2.958 | ***Station*** | 1.403 | 64.12 |
| ***Losses * RE*** | -0.28 | -3.577 | ***Ticket*** | 1.136 | 57.733 |
| ***Losses * FVZ*** | -0.173 | -2.028 | ***Productivity*** | 1.935 | 72.199 |
| ***Accident * RE*** | -0.243 | -3.527 | ***Relax*** | 1.36 | 62.262 |
| ***Accident * FVZ*** | -0.217 | -2.816 | | | |
| ***BahnCard * FVZ*** | 0.341 | 7.199 | | | |
| ***FVZ*** | 0.471 | 12.704 | | | |
| ***LB*** | -0.907 | -23.639 | | | |
| ***Transfers*** | -0.161 | -8.05 | | | |
| ***Transfers * Woman*** | -0.06 | -1.765[†††] | | | |
| ***Transfers * RE*** | 0.059 | 2.448 | | | |
| ***BusUser * LB*** | 0.338 | 8.011 | | | |
| ***MiddleAge * LB*** | -0.282 | -5.917 | | | |
| ***Old * LB*** | -0.522 | -3.796 | | | |
| ***Stress-free*** | | | ***Stress-free*** | | |
| ***Losses * RE*** | -0.539 | -6.808 | ***Relax*** | 0.719 | 36.669 |
| ***Losses * FVZ*** | -0.468 | -5.403 | ***Luggage*** | 2.119 | 72.975 |
| ***Accident * RE*** | -0.188 | -2.704 | ***Security*** | 1.546 | 64.834 |
| ***Accident * FVZ*** | -0.246 | -3.162 | | | |
| ***BahnCard * FVZ*** | 0.283 | 5.928 | | | |
| ***FVZ*** | 0.311 | 8.35 | | | |
| ***LB*** | 0.399 | 10.674 | | | |
| ***Transfers*** | -0.12 | -5.965 | | | |
| ***Transfers * Woman*** | -0.144 | -4.193 | | | |
| ***Woman * RE*** | -0.207 | -4.56 | | | |
| ***BusUser * LB*** | 0.104 | 2.438 | | | |
| ***MiddleAge * LB*** | -0.128 | -2.67 | | | |
| ***Old * LB*** | -0.462 | -3.327 | | | |
| ***Reliability*** | | | ***Reliability*** | | |
| ***LB*** | -0.227 | -6.209 | ***Departure*** | 2.115 | 72.855 |
| ***FVZ*** | 0.063 | 1.724[7] | ***Arrival*** | 2.284 | 74.573 |
| ***Transfers*** | -0.158 | -7.921 | | | |
| ***Transfers * Woman*** | -0.099 | -2.934 | | | |
| ***HighIncome * FVZ*** | -0.173 | -2.07 | | | |
| ***BahnCard * RE*** | 0123 | 4.155 | | | |
| ***BahnCard * FVZ*** | 0.083 | 2.583 | | | |
| ***BusUser * LB*** | 0.185 | 1.773[7] | | | |
| ***MiddleAge * RE*** | 0.133 | 2.974 | | | |
| ***MiddleAge * FVZ*** | 0.083 | 1.754[7] | | | |
| ***MiddleAge * LB*** | -0.103 | 2.185 | | | |
| ***Old * RE*** | 0.133 | 0.974[‡‡‡] | | | |
| ***Old * FVZ*** | 0.166 | 1.212[8] | | | |
| ***Old * LB*** | -0.347 | -2.536 | | | |

In line with our hypotheses, all explanatory variables affecting perceptions are directly related with the specific alternatives for which they were calculated, whether considering the attributes directly or through systematic taste variations (e.g. *Transfers * Woman*). It is

[†††] As the signs of the estimators were known *a priori*, a one-tailed test was performed ($\alpha_{5\%}$=1.645).

[‡‡‡] The variables were kept in the model, despite their low significance, as the signs as well as the magnitudes of the estimated parameters were consistent with the values obtained for the other age related estimators.

important to note that not considering the latter works as well, but provides a worse goodness-of-fit and therefore, a worse representation of the way in which the alternatives are apparently perceived by the individuals.

### 5.3 Discrete Choice

This section reports the results of the estimation of the discrete choice component of the model. In addition to the previously described latent variables, socioeconomic characteristics of the individuals and attributes of the alternatives (transport mode, price, travel time, number of transfers and safety level) were considered. In addition, an inertia variable taking the value of one when individuals chose their revealed preference option in spite of the advantages of new alternatives was introduced.

Altogether, there were 13,138 observations available for estimation. The potential correlation between the responses of a given individual (panel effect) was analysed but was not included in the final models since it was statistically insignificant[§§§]. The latent variables "Green" and "TrainFan" were considered both linearly as well as categorized into different levels (with the *+67%* sub-index indicating that individuals belong to the upper third of the categorization).

Estimation was performed using PythonBiogeme (Bierlaire, 2003). Table 4 presents the estimation results for five different specifications. The results of the t-test for statistical significance are presented in parenthesis and the reported value for the log-likelihood refers only to the discrete choice component.

As can be observed, two of the three perceptual indicators were found to be statistically significant. This way, both the perception of reliability and comfort affect positively the utility

[§§§] A plausible explanation for the absence of correlation relies in the fact that the alternatives presented were not related to a specific transport mode.

ascribed to a certain alternative. On the contrary, the perception of a stress-free travel appears not to be statistically important in the decision making process.

Table 4 – Estimated Parameters for the Discrete Choice model

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| *Inertia* | 0.443<br>(9.21) | 0.465<br>(6.36) | 0.415<br>(10.25) | 0.43<br>(9.92) | 0.363<br>(13.22) |
| *LB* | -0.626<br>(-3.01) | -0.592<br>(-2.51) | -0.741<br>(-3.42) | -0.768<br>(-3.37) | -1.47<br>(-12.46) |
| *RE* | -0.0147<br>(-0.14) | 0.134<br>(0.98) | 0.0201<br>(0.19) | 0.0338<br>(0.31) | -0.407<br>(-9.28) |
| *Travel Time* | -0.0341<br>(-9.85) | -0.0384<br>(-6.04) | -0.0233<br>(-12.12) | -0.0332<br>(-10.97) | -0.0266<br>(-16.25) |
| *Travel Time * LV Green* | - | - | 0.00892<br>(6.05) | - | - |
| *Travel Time * LV Green (+67%)* | 0.0271<br>(7.31) | 0.0311<br>(5.25) | - | 0.0266<br>(7.89) | 0.0206<br>(9.04) |
| *Ln(Price) * Very Low Income* | -7.34<br>(-11.97) | -8.17<br>(-6.64) | -6.77<br>(-14.5) | -7.09<br>(-13.71) | -5.91<br>(-28.69) |
| *Ln(Price) * Low Income* | -6.46<br>(-11.34) | -7.13<br>(-6.47) | -5.95<br>(-13.52) | -6.23<br>(-12.8) | -5.1<br>(-24.3) |
| *Ln(Price) * Middle Income* | -5.04<br>(-9.41) | -5.56<br>(-5.92) | -4.65<br>(-10.71) | -4.85<br>(-10.36) | -3.89<br>(-14.49) |
| *Ln(Price) * High Income* | -3.83<br>(-6.85) | -4.15<br>(-5.01) | -3.5<br>(-7.24) | -3.69<br>(-7.22) | -2.9<br>(-8.18) |
| *Safety Level* | -0.00556<br>(-4.63) | -0.00616<br>(-4.01) | -0.00519<br>(-4.78) | -0.00545<br>(-4.8) | -0.00419<br>(-4.77) |
| *Transfers* | -0.418<br>(-8.78) | -0.434<br>(-4.7) | -0.355<br>(-7.85) | -0.369<br>(-7.94) | -0.49<br>(-17.65) |
| *LV Comfort* | 0.84<br>(3.9) | 1.15<br>(3.03) | 0.661<br>(3.47) | 0.679<br>(3.41) | - |
| *LV Reliability* | 0.354<br>(1.51) | 0.145<br>(0.61) | 0.368<br>(1.69) | 0.395<br>(1.68) | - |
| *LV Stress-free* | - | 0.23<br>(1.39) | 0.204<br>(1.4) | 0.236<br>(1.56) | - |
| *FVZ * LV TrainFan* | 0.575<br>(3.2) | - | 0.54<br>(3.25) | 0.554<br>(3.19) | 0.625<br>(5.52) |
| *FVZ * LV TrainFan (+67%)* | - | -0.176<br>(-1.02) | - | - | - |
| *Log-Likelihood* | -7 413.747 | -7 416.368 | -7 424.263 | -7 412.272 | -7 443.765 |

Note that when the perceptual attributes are omitted, the mode specific constants are highly significant (Model 5), but when perceptions are integrated they capture a large part of the variability previously described by the former that become either statistically insignificant (RE), or their impact on the decision decreases (LB). Finally, everything is accompanied by a significant improvement in goodness-of-fit.

These findings are in accordance with theory, in the sense that mode specific constants normally capture the omitted information regarding a specific alternative; when perceptions

are considered, this information is enriched and therefore the importance of the constants decreases.

Regarding the attitudinal latent variables, it was found that our non-alternative related attitude ("Green") affects the way in which travel time is perceived (i.e. it was possible to identify a systematic taste variation related to this attitude). As can be observed, this variable is statistically significant both when considered linearly and when it is categorized, reflecting the importance of the systematic taste variation. However, the categorization of the variable ($LV\ Green_{+67\%}$) is associated with a substantially superior goodness-of-fit (Models 3 and 4), suggesting a considerably better representation of the behaviour. This finding is in line with the perception that shorter travel times imply higher speeds and, therefore, more $CO_2$ emissions and a larger damage to the environment. Also, the fact that the effect of this variable is not linear, is in agreement with the notion that only highly environmentally concerned individuals are willing to accept larger travel times in order to reduce the ecological harm.

Finally, as expected, our alternative related attitude ("TrainFan") is statically significant in conjunction with intercity trains. It was possible to detect a social group of train enthusiasts willing to favour the railways in spite of the apparent advantages of other alternatives. However, this favouritism does not extend to regional trains. In this case, models considering the variable in a categorized fashion (e.g. Model 2) do not outperform the linear specification. It was also not possible to identify a systematic taste variation within the alternative intercity trains.

## 6. CONCLUSIONS

Despite the significant technical and methodological improvements in the estimation of HDC models during the last decade, this has not led to a significantly better understanding of the way in which perceptions affect the decision making process, as these aspects are usually

ignored by modellers. Even in the case of attitudes, which have been widely studied, the specification of latent variables has tended to be rather simplistic and rarely depart from the linearity assumptions (fortunately latent class models have been an alternative in this regard), while the analysis of systematic taste variations in association with attitudes appears to be practically inexistent.

This reticence may be related to deeper concerns about artificial constructs, such as latent variables and the information that can be acquired from them. Nevertheless, it should not be forgotten that we as modellers aim to depict reality in the best way possible, and therefore, if we decide to work with latent variables, we should guide our efforts to represent as accurately as possible the decision making process and the fashion in which the different variables take part in it.

This paper gives an overview of the different ways in which attitudes and perceptions may affect the decision making process as well as providing practical recommendations about data collection and estimation issues.

Our empirical analysis provides evidence sustaining the fact that perceptions affect the way in which individuals ascribe a utility to a certain alternative. In the same line, our evidence shows that perceptions may explain a significant portion of the variability that is normally captured by alternative specific constants, offering significant improvements in goodness-of-fit for the whole model. Also, our results sustain our hypotheses, in the sense that attitudes may indeed be related with systematic taste variations and that attitudinal latent variables should be treated in the same way as socio-economic variables.

Although we were able to identify systematic taste variations as well as a categorization for latent variables that outperform the linearity assumption, this does not imply that every attitudinal latent variable should be considered in this way. Prior to estimation, or even better, prior to constructing the experiment, the analyst should study which variables take part on

the decision making process and decide the way in which they are considered in accordance with the underlying theory.

Although not treated in this paper, interactions among attitudes and perceptions represent an issue that should be considered in further research, as cognitive dissonance might affect the way in which different alternatives are perceived, leading to correlation and more complex error structures.

**ACKNOWLEDGMENTS**

We gratefully acknowledge the support of BecasChile given by the Chilean Council for Scientific and Technological Research (CONYCIT), the Millennium Institute in Complex Engineering Systems (ICM: P05-004F; FONDECYT: FB016), the All Latitudes and Cultures BRT Centre of Excellence funded by the Volvo Research and Educational Foundations, the Alexander von Humboldt Foundation and the Centre for Sustainable Urban Development, CEDEUS (Conicyt/Fondap/15110020).

**REFERENCES**

van Acker, V., Mokhtarian, P.L. and Witlox, F. (2011). Going soft: on how subjective variables explain modal choices for leisure travel. *European Journal of Transport and Infrastructure Research* **11**, 115-147.

Allport, G.W. (1935). Attitudes. In Murchison, C. (ed.). *A handbook of social psychology*. Clark University Press, Worcester, MA, USA.

Ashok, K., Dillon, W. and Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research* **39**, 31-46.

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

Bahamonde-Birke, F.J. and Ortúzar, J. de D. (2014a). On the variability of hybrid discrete choice models. *Transportmetrica* **10**, 74-88.

Bahamonde-Birke, F.J. and Ortúzar, J. de D. (2014b). Is sequential estimation a suitable second best for estimation of hybrid choice models? *93rd Annual Meeting of the Transportation Research Board 2014*, Washington, D.C., 12-16, January 2014.

Bhat, C.R. (2014). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Working paper, University of Texas at Austin.

Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H.S. Mahmassani (ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bolduc, D., Boucher, N. and Alvarez-Daziano, R. (2008). Hybrid choice modeling of new technologies for car choice in Canada. *Transportation Research Record: Journal of the Transportation Research Board*, ***2082*****(1)**, 63-71.

Bolduc, D. and Alvarez-Daziano, R. (2010). On the estimation of hybrid choice models. In S. Hess and A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice*, 259–288. Emerald, Bingley.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, Chichester.

Chorus, C. and Kroesen, M. (2014). On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transport Policy*, 10.1016/j.tranpol.2014.09.001.

Daziano, R.A. and Barla, P. (2012). On the way to sustainable commuting to campus: stated, revealed, and attitudinal response using a mixed logit model. *Kuhmo Nectar Conference of ITEA 2012*, Berlin, 21-22, June, 2012.

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

Domencich, T. and McFadden, D. (1975). *Urban Travel Demand – A Behavioural Analysis.* North Holland, Amsterdam.

Lindsay, P.H. and Norman, D.A. (1972). *Human information processing: An introduction to psychology*. Academic Press, NY, USA.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers in Econometrics*, 105-142. Academic Press New York.

Ortúzar, J. de D. and Willumsen, L.G (2011). *Modelling Transport*. Fourth Edition, John Wiley and Sons, Chichester.

Pickens, J. (2005). *Organizational Behavior in Health Care.* Jones and Bartlett Publishers, Sudbury, MA, USA.

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

Raveau, S., Yáñez, M.F. and Ortúzar, J. de D. (2012) Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research* **46B**, 1374-1383.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review* **34**, 273 286.

Train, K.E., McFadden, D.L. and Goett, A.A. (1987). Consumer attitudes and voluntary rate schedules for public utilities. *Review of Economics and Statistics* **64**, 383-91.

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

Walker, J.L. (2002). The mixed logit (or logit kernel) model: dispelling misconceptions of identification. *Transportation Research Record* **1805**, 86-98.

Zellner A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review* **11**, 441-454