Chicken or the egg? Modeling simultaneous transportation choices: a joint model of mode choice, residential location, car and transit pass ownership

Abstract

Keywords:

JEL:

1 INTRODUCTION

An average person makes several hundred or even several thousand choices every day. These choices, most of which are certainly unconscious, affect all areas of life (health, food, transport...). Some choices are short-term and are repeated almost daily (mode choice, dress code...). Other choices are medium-term and are repeated every year or every five to ten years (car purchase, furniture, job...). Other choices are long-term and are made only one or two times in a life (residential location, spouse...). This continuum of choices is
Despite the recognition of the importance of taking into account the simultaneity of choices to avoid erroneous parameter's estimates (Pinjari et al., 2011; Bhat et al., 2014; Bhat, 2015; Bhat et al., 2016) most research still models transportation related choices independently of each other and considers the other choices as exogenous.

For instance, mode choice is one of the most studied choices in the transportation field because of its environmental impacts (e.g. air pollution, climate change, space consumption) and the heavy and costly infrastructure some transport modes require. According to a literature review (De Witte et al., 2013), variables used to explain mode choice include number of cars, variables linked to the residential location (density, frequency of public transport, proximity to infrastructures and services) or travel cost, which is linked to the decision to have a transit pass. Usually, these variables are considered to be explanatory variables for mode choice, which makes it difficult to identify the true causal relationship and to distinguish a causal relationship from a purely associative relationship. While one may hypothesize that household car ownership has a significant impact on mode choice, it is also possible that a household has several cars because all adults in the household commute by car. In the same way, the characteristics of the available transport modes are strongly linked to the residential location. For instance, some commuters may have no public transport modes available because they chose to live in a low density zone. The car dependency may then be partly due to a residential self-selection effect.

The transportation choices include long-, medium- and short-term choices. To develop effective and relevant public policies, it is necessary to take into account all the adaptations made by people along this continuum of choices, and therefore to consider both long-term and medium-term choices as endogenous to travel models, including mode choice models (Pinjari et al., 2011). Such an approach has the advantage of taking into account intermediate effects of medium-term choices intervening effects of medium-term in the interconnections between the long- and short-term choices. Concerning short-term choices, we focus on mode choice which is decisive in terms of public policies for the reasons mentioned above. Concerning medium-term choices, we focus on equipment and investment decisions, such as car, bicycle or transit-pass ownership, which determine the range of choices available, as well as the marginal cost of using each option. Concerning long-term choices, we consider the household residential location choice, whose simultaneity with short-term travel has been recognized by the field of integrated land-use transport modeling (Cervero, 2002; Timmermans, 2003).

(Pinjari et al., 2011) summarize the various interdependencies between this continuum of choices in four categories. First, long-term choices have a **causal effect** on short-term choices (e.g., someone living in rural areas does not have access to the same travel modes as someone living in urban areas). Second, the **residential self-selection** effect is due to the selection of individuals in residential areas based on lifestyle preferences, which are also related to car or transit-pass ownership or to mode choice. Third, bicycle, car or transit-pass **ownership** is **endogenous** with respect to mode choice (e.g. individuals commuting by train and buying an annual transit-pass). Fourth, some choices are **associative** (and not causal) and can be explained by common unobserved latent variables (e.g. a household whose member has a transit-pass will have fewer cars available, may be because of positive perceptions of public transport modes or a high environmental concern).

In statistical terms, ignoring the interdependence of the transportation related choices and estimating separate models is misleading for several reasons (Bhat, 2015). First, it is inefficient in estimating covariate effects for each endogenous choice because it fails to borrow information from the other endogenous choices. Second, it results in inconsistent estimation of the effects of one endogenous choice on the other. Third, the use of a joint model facilitates the use of global tests, which increases the power of statistical test and allow for a better control of type I error rates. Failing to consider the "bundling" of choices therefore lead to a spurious interpretation of the results and, consequently, to potentially inappropriate public policies.

In order to model the simultaneity of choices, a key element is to take into account the latent psychological variables underlying the different choices, that is to measure these latent variables through indicators and integrate them as explanatory variables in the choice equations as well as consider the unobserved covariance among the multidimensional outcomes. For instance, Bhat (2015) shows that there is a substantial degradation of parameter recovery if the latent psychological variables are ignored away, and especially those associated with the endogenous variable effects.

On the basis of the work of Pinjari et al. (2011), we model the travel choices continuum. Compared to this work, our contribution to the literature is three-fold. First, we explicitly model latent variables by means of a structural equation model to analyze their simultaneous impact on several choice dimensions and on residential self-selection. This work is made possible by the Generalized Heterogeneous Data Model (Bhat, 2015). Second, mode choice is modeled explicitly with the description of travel cost, time and quality of service.

2 MODEL

The model developed in this article is based upon the Generalized Heterogeneous Data Model (GHDM) (Bhat, 2015), which is a generalization of Bhat and Dubey (2014) probit-based approach of Integrated Choice Latent Variables (ICLV) model. In this model, the use of normally distributed error terms is a mean to 1) reduce the estimation time with alternative estimation methods (Maximum Approximate Composite Marginal Likelihood - MACML), and 2) enable more flexible substitution patterns across alternatives by specifying a general covariance matrix. GHDM jointly handles mixed types of dependent variables, including nominal discrete variables and count variables. The covariance relationships between the dependent variables are expressed as a function of the latent variables. To estimate this model, instead of using a simulation-based approach as it is usually the case with latent variables model, we use the Maximum Approximate Composite Marginal Likelihood (MACML) developed by Bhat (2011).

2.1 Model formulation

The notations used in this section are drawn from Bhat (2015). For ease of notation, we consider a cross-sectional model and omit the index q for decision-makers ($q = 1, 2, \cdots, Q$). The overall framework is made of two components. The first component is the latent variable structural model which relates the latent psychological variables (e.g. environmental concern, attitudes, perceptions...) to each other (based on psychological models, such as the theory of planned behavior-Ajzen, 1985- or the transtheoric model -Prochaska and DiClemente, 1986) and to the observed socio-demographic variables. The second component of the GHDM models each of the dependent variables with observed and latent variables (in a reduced form). It is the presence of the latent variables in the modeling of all outcomes that estimates jointness among the outcomes.

2.1.1 Structural equations model

Let consider L latent variables $z_l^\star$ $(l = 1, 2, \cdots, L)$. Each latent variable may be written as a linear function of observed covariates w:

$$z_l^\star = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(D \times 1)$ vector of covariates (excluding a constant? If yes, we have to standardize all items?). $\boldsymbol{\alpha}_l$ is the $(D \times 1)$ vector of coefficients associated with the covariates. η_l is a normally distributed random error term. We also define the $(L \times D)$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_L)'$ as well as the $(L \times 1)$ vectors $\boldsymbol{z}^\star = (z_1^\star, z_2^\star, \cdots, z_L^\star)'$ and $\boldsymbol{\eta}^\star = (\eta_1^\star, \eta_2^\star, \cdots, \eta_L^\star)'$. $\boldsymbol{\eta}$ follows a MultiVariate Normal (MVN) distribution to accommodate interactions among the latent variables: $\boldsymbol{\eta} \rightsquigarrow \text{MVN}_L[\boldsymbol{O_L}, \boldsymbol{\Gamma}]$, with $\boldsymbol{O_L}$ a $(L \times 1)$ vector composed of zeros and $\boldsymbol{\Gamma}$ the $(L \times L)$ variance-covariance matrix. With $\boldsymbol{\Gamma}$, instead of specifying *a priori* relationships between the latent variables (as in psychological models), we specify a general covariance structure between the latent variables. This explains why $z_l^\star$ depends only on covariates $\boldsymbol{w}$ and not on other latent variables $z_k^\star$ with $l \neq k$. In matrix form, Equation 1 may be re-written:

$$\boldsymbol{z}^\star = \boldsymbol{\alpha} \boldsymbol{w} + \boldsymbol{\eta}. \tag{2}$$

2.1.2 Structural model of the mixed outcome system

Compared to (Bhat, 2015) and subsequent papers on GHDM, we restrict the mixed outcome system to both unordered qualitative variables (discrete choice variables) and count variables.

Modeling non-orderered qualitative variables We consider three unordered qualitative variables: mode choice, transit pass ownership and residential location, indexed by g $(g = 1, \cdots, G)$. I_g is the number of alternatives corresponding to the gth variable $(I_g \in \{2, 3\})$ and i_g is the corresponding index $(i_g = 1, 2, \cdots, I_g)$.
BHAT 2015 considers $I_g \geq 3$. Why not also considering a binomial probit model with $I_g = 2$?
Let define U_{i_g} the utility associated to the ith alternative of unordered qualitative variable g, which is modeled with a linear combination of observable and latent variables:

$$U_{gi_g} = \boldsymbol{b}'_{gi_g} \boldsymbol{x} + \boldsymbol{\nu}_{gi_g} (\boldsymbol{\beta}_{gi_g} \boldsymbol{z}^\star) + \zeta_{gi_g}, \tag{3}$$

with $\boldsymbol{x}$ an $(A \times 1)$ vector of observed exogeneous variables (including a constant) as well as the observed values of the other endogenous variables (either unordered qualitative or count variables).[1] $\mathbf{b}_{gi_g}$ is an $(A \times 1)$ vector of coefficients to estimate, and ζ_{gi_g} is a standard normal random error term. $\boldsymbol{\beta}_{gi_g}$ is an $(N_{gi_g} \times L)$ matrix of variables interacting with the latent variables to influence the utility of alternative i_g.
To write Equation 3 in the matrix form, let define:

- an $(I_g \times 1)$ vector $\boldsymbol{\zeta}_g = (\zeta_{g1}, \cdots, \zeta_{gI_g})'$, with $\boldsymbol{\zeta}_g \rightsquigarrow \text{MVN}_{I_g}\big(\mathbf{0}, \boldsymbol{\Lambda}_g\big)$
- an $(I_g \times 1)$ vector $\boldsymbol{U_g} = (U_{g1}, U_{g2}, \cdots, U_{gI_g})'$
- a $(I_g \times A)$ matrix $\boldsymbol{b}_g = (\boldsymbol{b}_{g1}, \boldsymbol{b}_{g2}, \cdots, \boldsymbol{b}_{I_g 1})'$
- a $(\sum_{i_g=1}^{I_g} N_{gi_g} \times L)$ matrix $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, \cdots, \boldsymbol{\beta}'_{gI_g})'$

[1] $\boldsymbol{x}$ and $\boldsymbol{w}$ may contain different or similar variables.

- a $(I_g \times \sum_{i_g=1}^{I_g} N_{gi_g})$ matrix which contains the $\boldsymbol{\nu}'_{gi_g}$ row vectors
- a $(I_g \times L)$ matrix $\boldsymbol{\varpi} = (\boldsymbol{\nu}_g \times \boldsymbol{\beta}_g)$
- a $(\sum_{g=1}^{G} I_g \times 1)$ vector $\boldsymbol{U} = (\boldsymbol{U}'_1, \boldsymbol{U}'_2, \cdots, \boldsymbol{U}'_G)'$
- a $(\sum_{g=1}^{G} I_g \times \sum_{g=1}^{G} I_g)$ matrix $\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & 0 & & 0\cdots & 0 \\ 0 & \boldsymbol{\Lambda}_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{\Lambda}_G \end{pmatrix}$
- a $(\sum_{g=1}^{G} I_g \times 1)$ vector $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \cdots, \boldsymbol{\zeta}_G)'$, with $\zeta \rightsquigarrow \text{MVN}(\mathbf{0}, \boldsymbol{\Lambda})$
- a $(\sum_{g=1}^{G} I_g \times A)$ matrix $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, \cdots, \boldsymbol{b}'_G)'$
- a $(\sum_{g=1}^{G} I_g \times L)$ matrix $\boldsymbol{\varpi} = (\boldsymbol{\varpi}'_1, \boldsymbol{\varpi}'_2, \cdots, \boldsymbol{\varpi}'_G)'$
- a vector $\boldsymbol{\nu} = \text{Vech}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \cdots, \boldsymbol{\nu}_G)$.

In matrix form, Equation 3 may be rewritten:

$$\boldsymbol{U} = \boldsymbol{b}\boldsymbol{x} + \boldsymbol{\varpi}\boldsymbol{z}^\star + \boldsymbol{\zeta}. \tag{4}$$

Modeling count variables The count variable, number of cars owned by the household, is modeled as a generalized ordered probit; see Bhat (2015) for more details. Let c be the count variable and $y^\star$ be the associated latent continuous propensity variable. $y^\star$ may be written as a linear function of observed covariates $\boldsymbol{x}$ and latent variables $\boldsymbol{z}^\star$:

$$y^\star = \boldsymbol{\gamma}\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{z}^\star + \epsilon. \tag{5}$$

$\boldsymbol{\gamma}$ is a $(A \times 1)$ vector of observed covariate loadings on the count outcome and $\boldsymbol{d}$ is an $(L \times 1)$ vector of latent variable loadings on the count outcome. ϵ is a standard normal random error term.
$y^\star$ is mapped to the observed count variable c by threshold parameters ψ_m, with m the observed value of the count variable. The threshold parameters are determined by the function:

$$\psi_m = \Phi^{-1}\left[\frac{(1-c)^\theta}{\Gamma(\theta)} \sum_{t=0}^{m} \left(\frac{\Gamma(\theta+t)c^t}{t!}\right)\right] + \phi_n \tag{6}$$

with

$$c = \frac{exp(\tilde{\gamma}'\tilde{\boldsymbol{x}})}{exp(\tilde{\gamma}'\tilde{\boldsymbol{x}}) + \theta}. \tag{7}$$

In Equations 6 and 7, θ is a dispersion parameter and ϕ a flexibility parameter that allow flexible count distribution modeling. In Equation 6, Φ^{-1} is the inverse function of the cumulative normal distribution function and Γ is the traditional gamma function. In Equation 7, $\boldsymbol{x}$ is a $(G \times 1)$ vector defined as above and $\tilde{\boldsymbol{\gamma}}$ is a $(G \times 1)$ vectors of parameters. For identification issues, we set $\psi_{-1} = -\infty$ and $\psi_0 = 0$. We also define $m^\star$, a count value above which $\psi_{m>m^\star}$ is held fixed ($\psi_{m>m^\star} = \psi_{m^\star}$) to allow the count model to predict beyond the range available in the estimation sample.

The model system The model system is composed of the Equations 1, 4 and 5. In its reduced form, replacing $\boldsymbol{z}^\star$ by its expression, the system can be re-written:

$$\begin{cases} \boldsymbol{U} & = \boldsymbol{b}\boldsymbol{x} + \boldsymbol{\varpi}\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\varpi}\boldsymbol{\eta} + \boldsymbol{\zeta} \\ y^\star & = \boldsymbol{\gamma}'\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{d}\boldsymbol{\eta} + \epsilon \end{cases} \tag{8}$$

The $(C \times \sum_{g=1}^{G} I_g)$ vector $y^\star \boldsymbol{U} \rightsquigarrow \text{MVN}(\boldsymbol{B}, \boldsymbol{\Omega})$, with

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\gamma}'\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{\alpha}\boldsymbol{w} \\ \boldsymbol{b}\boldsymbol{x} + \boldsymbol{\varpi}\boldsymbol{\alpha}\boldsymbol{w} \end{pmatrix}$$

and

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{d}'\boldsymbol{\Gamma}\boldsymbol{d} & \boldsymbol{d}'\boldsymbol{\Gamma}\boldsymbol{\varpi} \\ \boldsymbol{\varpi}'\boldsymbol{\Gamma}\boldsymbol{d} & \boldsymbol{\varpi}\boldsymbol{\Gamma}\boldsymbol{\varpi}' + \boldsymbol{\Lambda}. \end{pmatrix}$$

2.2 Estimation

The MACML approach allows consistent estimators to be obtained without being computationally intensive, even in high dimensional mixed multivariate model systems. It only requires the evaluation of bivariate or univariate cumulative normal distribution functions regardless of the number or latent variables or the number and type of dependent variable outcomes. To provide us with a good starting point throughout estimation steps, we start from individual models (for each dependent variable), then the models estimated by combinations of two and then all three together. Then, we add the SEM section and combine everything to Measurement equations model.

3 DATA

Stated Preferences (SP) data were collected between January and April 2015 in the Rhône-Alpes Region (France). In addition to the choice experiment questions, the originality of this survey is that it includes questions about environmental concern, motives for car use, as well as attitudes to and perceptions of public transport modes.

3.1 The survey

The survey methods consisted of face-to-face and web-based interviews. The sample of surveyed travelers was compiled from two sources. We first sampled respondents from a large revealed preferences travel survey carried out in the same region. This database of more than 37,000 travelers is geographically stratified (Hurez and Tébar, 2014). Those travelers who declared that they had used the train as a mode of transport on one of their reported trips were asked to answer the web survey. Due to the low rate of regular train users in the population, they were oversampled with a face-to-face survey carried out in regional trains using the quota sampling method (sex, age, motive, travel time and train line).

Respondents were first asked to describe in detail (time, cost, purpose, origin and destination) a journey they had made by coach, train or car during the last month within the area of the Rhône-Alpes Region. This reference journey was then used to tailor the choice questions. Such a strategy is known to minimize the hypothetical bias.

Only respondents living in the Rhône-Alpes Region, aged 18 or over, having a car and a driving license and whose trip was made or could have been made by train or coach, were asked to answer the choice questions. The availability of the alternatives was checked by creating a database with travel time by public transport and car for each of the 8.6 million origin-destination pairs in the Rhône-Alpes Region, within a radius of 10 km around train stations. In total, 1799 persons answered the whole SP survey (both choice and attitudinal questions, see hereafter).

Table 1 reports the descriptive statistics for all the variables used in the models. It is important to note that the survey was not designed to be representative of the entire population of travelers in the region but rather to analyze drivers of mode choice. For more details on the survey, see Bouscasse (2017).

3.2 Choice questions

The choice experiment focused on mode choice. Each respondent had to choose between three transport modes: train, coach or car. Alternatives were described in terms of travel mode, cost, time, probability and time delay, frequency, clock-face timetable and comfort. To avoid a cognitive burden, variations of the attributes describing the proposed journeys were split into three choice exercises. We focus here on the one that describes modes of transport that differ with respect to travel time, travel cost and level of comfort. Figure 1 depicts one choice exercise. Travel time was defined from origin to destination (sum of access time, egress time, waiting time and in-vehicle time). Travel cost included public transport ticket or pass, gasoline, parking cost and toll. Comfort is defined as a dummy variable that models whether a seat is guaranteed (comfort= 1) or not guaranteed (comfort= 0). Respondents were faced with a series of four choice questions. Since a few respondents did not answer all of them, XXX observations are available.

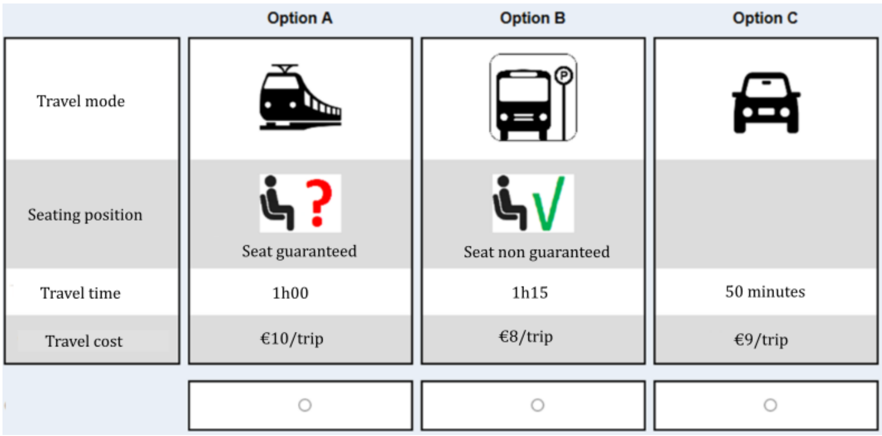

FIGURE 1 Example of choice question in the choice experiment

One of the modes of transport is defined as a status-quo alternative. For the mode of transport reported in the reference journey, the actual travel time and cost attributes were systematically proposed in the choice experiment. For the other alternatives, the levels of the time and cost attributes are pivoted around the collected reference values. To improve the efficiency of the design, a Bayesian efficient design was implemented (Rose et al., 2008). *A priori* weights of attributes were taken from the literature and adjusted during the pilot tests.

3.3 Attitudinal variables

The last part of the questionnaire is dedicated to the collection of additional socio-economic characteristics and quantitative information to capture attitudes and psychological constructs about traveling habits. This part of the questionnaire attempts to measure three sets of attitudinal variables: environmental concern, motives for car use and perception of comfort in public transport. A first survey, dedicated to the measurement of these latent constructs, allowed us to refine the phrasing and selection of the measurement items.

To investigate the simultaneity in short, medium and long term travel-related choices, we focus

TABLE 1 Descriptive statistics

Variable definition	Label	Mean	S.D.	Min	Max
Alternative-specific variables					
Travel cost by train (in euros)	Train cost	8.96	7.91	1.00	62.00
Travel cost by coach (in euros)	Coach cost	8.91	8.05	1.00	78.00
Travel cost by car (in euros)	Car cost	10.41	8.93	1.00	62.00
Travel time by train (in minutes)	Train time	72.42	52.73	7.00	325.00
Travel time by coach (in minutes)	Coach time	72.87	54.28	7.00	325.00
Travel time by car (in minutes)	Car time	59.13	38.54	4.00	330.00
Comfort in train (1 if a seat is guaranteed, 0 otherwise)	Comfort	0.51			
Individual variables					
Age (in years)	Age	46.15	15.63	19.00	83.00
Number of cars in the household	Cars	1.68	0.72	1.00	5.00
Gender (1 if man, 0 if woman)	Gender	0.52	0.50	0.00	1.00
Presence of children in the household (1 if yes, 0 otherwise)	Child	0.34	0.47	0.00	1.00
Monthly income above 4,000 euros (1 if yes, 0 otherwise)	Income_h	0.29	0.45	0.00	1.00
Car user for the reference trip (1 if yes, 0 otherwise)	Car_user	0.49	0.50	0.00	1.00
Perceived behavioral control					
I'm not comfortable when I travel with people I don't know well.	Pbc1	3.67	1.03	1.00	5.00
It's hard to take public transport when I travel with my children.	Pbc2	2.80	1.10	1.00	5.00
It's hard to take public transport when I travel with bags or luggage.	Pbc3	2.10	1.02	1.00	5.00
Perceived time					
I like seeing people and having other people around me.	Ptime1	3.30	0.90	1.00	5.00
It's time I put up with and I just wait for it to pass.	Ptime2	3.21	1.10	1.00	5.00
I use the time to rest and relax.	Ptime3	3.83	0.89	1.00	5.00
I use the time to do things I wouldn't necessarily do elsewhere.	Ptime4	3.28	1.05	1.00	5.00
I just want to be on my own and undisturbed.	Ptime5	2.85	1.04	1.00	5.00
Given my commutes, the time is too short: I don't have time to do anything.	Ptime6	3.54	0.89	1.00	5.00
It's wasted time.	Ptime7	3.55	1.03	1.00	5.00
Feelings					
I feel a sense of freedom.	Feel1	2.36	1.00	1.00	4.00
It puts me in a good mood.	Feel2	2.50	0.77	1.00	4.00
I feel comfortable and at ease.	Feel3	2.55	0.77	1.00	4.00
I feel I could meet people and get into conversation with them.	Feel4	2.12	0.80	1.00	4.00
I feel I'm doing something, I feel useful.	Feel5	1.83	0.86	1.00	4.00
I find the people, noise and smells disagreeable.	Feel6	3.09	0.68	1.00	4.00
I feel stressed.	Feel7	3.60	0.64	1.00	4.00
I feel harassed.	Feel8	3.73	0.53	1.00	4.00

Notes: The perceived behavioral control and perceived time items are measured on the basis of 5-point Likert scales, which range from "completely disagree" (1) to "completely agree" (5). feelings experienced in public transport are measured on 4-point Likert scales ranging from "never" (1) to "always" (5).

on the variables that pertain to 1) environmental concern; 2) motives for car use; 3) perception of comfort in public transport. Environmental concern is composed of three dimensions: cognitive with items from the New Environmental Paradigm Scale (Dunlap et al., 2000), affective with items describing to which extent the respondent feel concerned by problems caused by the car (Steg, 2003), and conative with items describing transport public policies (Hurtubia et al., 2014; **?**; Steg, 2003). See Bouscasse et al. (2018) for more details. Motives for car use is measured with items related to instrumental, affective and symbolic motives (Steg, 2005; Lois and López-Sáez, 2009). Perceived comfort during travel by public transport modes model three main features: perceived time in interurban public transport, feelings experienced during journeys made by public transport and Perceived Behavioral Control (PBC) on using interurban public transport.[2]

Table 1 lists all the items presented in the survey to measure these three latent variables. The internal consistency of the *Perceived Time* latent variable improves without the item *ptime5*. This item is thus dropped for further analysis. The measurement for *Perception of time* and *Feelings* is based on a local study carried out on public transport in Lyon (Casals, 2012). The items used to represent PBC are based on Atasoy et al. (2011) and Morikawa et al. (1996).

4 RESULTS

5 CONCLUSION

[2]The notion of PBC is part of the theory of planned behavior (Ajzen, 1985). This theory is based on the idea that behavior is driven by internal mental states rather than external conditions, with the assumption that behavior is the outcome of a deliberative conscious process (Savage et al., 2011). Behavior is supposed to be determined by intention, which is, in turn, determined by a combination of three factors: attitudes, social norms and PBC. PBC is defined as the perceived ease or difficulty with which an individual performs a particular behavior, here traveling by public transport.

REFERENCES

Ajzen, I. (1985). *From Intentions to Actions: A Theory of Planned Behavior*. SSSP Springer Series in Social Psychology. Springer Berlin Heidelberg.

Atasoy, B., Glerum, A., and Bierlaire, M. (2011). Attitudes towards mode choice in switzerland. Technical Report TRANSP-OR 110502, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (macml) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, 45(7):923–939.

Bhat, C. R. (2015). A new generalized heterogeneous data model (ghdm) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological*, 79:50–77.

Bhat, C. R., Astroza, S., Bhat, A. C., and Nagel, K. (2016). Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: Application to residential self-selection effects analysis in an activity time-use behavior model. *Transportation Research Part B: Methodological*, 91:52–76.

Bhat, C. R., Astroza, S., Sidharthan, R., Alam, M. J. B., and Khushefati, W. H. (2014). A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B: Methodological*, 68:31–51.

Bhat, C. R. and Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, 67:68–85.

Bouscasse, H. (2017). *Essays on travel mode choice modeling - A discrete choice approach to the interactions between economic and behavioral theories*. PhD thesis.

Bouscasse, H., Joly, I., and Bonnel, P. (2018). How does environmental concern influence mode choice habits? a mediation analysis. *Transportation Research Part D: Transport and Environment*, 59:205–222.

Casals, J. (2012). Étude exploratoire sur les perceptions du temps et les activités dans les transports en commun - rapport de l'enquête quantitative. Technical report, Nova 7.

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, 7(4):265–284.

De Witte, A., Hollevoet, J., Dobruszkes, F., Hubert, M., and Macharis, C. (2013). Linking modal choice to motility: A comprehensive review. *Transportation Research Part A: Policy and Practice*, 49:329–341.

Dunlap, R., Liere, K., Mertig, A., and Jones, R. E. (2000). Measuring endorsement of the new ecological paradigm: A revised nep scale. *Journal of Social Issues*, 56(3):425–442.

Hurez, C. and Tébar, M. (2014). The Mobility of Rhône-Alpes Inhabitants: the Construction of a New Travel Survey Protocol. In *Transport Research Arena-Paris*.

Hurtubia, R., Nguyen, M., Glerum, A., and Bierlaire, M. (2014). Integrating psychometric indicators in latent class choice models. *Transportation Research Part A: Policy and Practice*, 64:135–146.

Lois, D. and López-Sáez, M. (2009). The relationship between instrumental, symbolic and affective factors as predictors of car use: A structural equation modeling approach. *Transportation Research Part A: Policy and Practice*, 43(9-10):790–799.

Morikawa, T., Ben-Akiva, M., and McFadden, D. (1996). Incorporating psychometric data in econometric choice models. Technical report, Massachusetts Institute of technology.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., and Waddell, P. A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6):933.

Prochaska, J. O. and DiClemente, C. C. (1986). Toward a comprehensive model of change. In *Treating addictive behaviors*, pages 3–27. Springer.

Rose, J. M., Bliemer, M. C., Hensher, D. A., and Collins, A. T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological*, 42(4):395–406.

Savage, B., Knight, T., Bacon, J., Millington, A., Bullock, H., and Buckland, J. (2011). Behavioural insights toolkit. Technical report, Social Research and Evaluation Division, UK Department for Transport, London, UK.

Steg, L. (2003). Can public transport compete with the private car? *IATSS Research*, 27(2):27–35.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice*, 39(2-3):147–162.

Timmermans, H. (2003). The saga of integrated land use-transport modeling: how many more dreams before we wake up. In *Keynote paper, Moving through nets: The Physical and social dimension of travel, 10th International Conference on Travel Behaviour Research, Lucerna, www. ivt. baug. ethz. ch/allgemein/pdf/timmermans. pdf*.