View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Infoscience - École polytechnique fédérale de Lausanne

Capturing response behavior heterogeneity in the quantification of semi-open questions: a mixed discrete-continuous approach

Aurélie Glerum *   Michel Bierlaire *

5th March 2014

Report TRANSP-OR 140305
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

Abstract

This paper presents a method to account for response behavior heterogeneity in the quantification of adjectives reported in semi-open questions. Semi-open questions are useful to capture psychological constructs such as perceptions. However, due to the qualitative nature of adjectives, it is difficult to assign an objective numerical value to them. The proposed model contributes to (1) account for response behavior heterogeneity in the evaluation of a new data type and (2) investigate the use of both discrete and continuous scales. As application example, the estimated ratings of the adjectives are integrated into a hybrid choice model of transportation mode preferences.

Key words

Response style, latent variable models, semi-open questions, adjectives, discrete choice models, hybrid choice models, perceptions.

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

1 Introduction

The measurement of psychological constructs is a complex task due to their latent nature. Traditional measurement techniques have relied on the design of statements to be rated on a Likert-scale (Likert, 1932). However studies in social sciences (Krosnick et al., 2005; Kaufmann et al., 2001) and in psychology (Mossholder et al., 1995; Potkay and Allen, 1973) have emphasized on the importance of considering responses to open questions to fully capture respondents' conception of a construct. Such data type has also been used to measure attitudes in political science studies (Holbrook et al., 2001). Despite more reliable (Krosnick et al., 2005), accurate (Churchill, 1991) and diverse (Schuman and Presser, 1996) answers, the information from open questions is difficult to translate into a quantitative measure. This complexity is partially reduced in the case where respondents are asked to elicit words or short expressions. Methodologies designed to provide quantitative inference based on such data have been increasingly used in the recent years (Arentze et al., forthcoming; Glerum et al., 2014; Horeni et al., 2010; Ter Hofstede et al., 1998).

In this research we focus on quantifying adjectives measuring perceptions. More specifically, we present a novel methodology that allows to assign an objective score to an adjective. This process is not straightforward and therefore several key aspects need to be taken into consideration.

First any rating process involves subjectivity. In the case of opinion statements rated on a Likert scale, two individuals with a similar opinions might assign slightly different ratings to the same sentence. Response behavior heterogeneity is a well-known phenomenon and a variety of different response behaviors have been identified in the literature (Clarke III, 2001). They include yea-saying, nay-saying, centrism or extremism / extreme response behavior (ERS). In the case of statements on a Likert scale, models accounting for 'differing response category interpretation' (Javaras, 2004) while capturing individuals' attitudes have been investigated (Javaras and Ripley, 2007; Johnson, 2003). Such differences can be explained by socio-demographic factors and in particular, ERS varies importantly across cultures (Dolnicar and Grün, 2007; Leung and Bond, 1989; Chun et al., 1974; Zax and Takahashi, 1967). Variations in response behavior needs not to be neglected since differences in scale usage may bias analyses (Johnson and Bolt, 2010; Rossi et al., 2001). In this reseach, we show that heterogeneous response behavior occurs in the rating process of a new type of psychometric indicators (the adjectives) and take this aspect into account in the model we develop.

Second the rating scale should be chosen adequately. Discrete scales are highly used in the literature and theories diverge regarding the optimal number of items that can be chosen from. Five-point (e.g. Atasoy et al., 2013) and seven-point (e.g. Sarman et al., 2013; Finstad, 2010) scales are very common. Recent studies also recommend using scales with a higher number of points. For example, Cummins and Gullone (2000) recommend using a ten-point scale. Shaftel et al. (2012) show that the reliability of the analysis increases monotonically as a function of the number of response categories in a given scale. While some researchers highlight that psychological constructs such as attitudes can be represented as a 'continuum that runs from extremely anti-object to extremely pro-object' (Javaras, 2004), continuous scales have only been used to a minor extent. In an early study by Arnold (1981), subjects are asked to indicate their preference on a drawn line and the distance between edges is measured with a ruler. Such a method is also used by Russell and Bobko (1992) who however indicate that the measurement process is cumbersome, but that it should be eased by computer-assisted methods. At present the latter techniques are available and we are making use of them in this paper. Russell and Bobko (1992) also provide a comparative analysis between discrete and continuous scale regarding attitude statements. They conclude that five-point Likert scales are too coarse and do not allow to capture correctly interactions. In the context of this research, it is difficult to determine whether a continuous or a discrete scale is more appropriate for the rating of the adjectives. As stated by Russell and Bobko (1992), experiments should allow as many response items as there exists in the theoretical response domain. In our case, it is not straightforward to define what the response domain is. Therefore we investigate the use of both discrete and continuous scales. In addition we also explore the benefits of a joint discrete-continuous approach.

The contributions of this paper are two-fold. First of all we account for response behavior heterogeneity in the evaluation process of a new type of data. Second we develop a modeling framework that accounts both for discrete and continuous ratings of the same items. We compare our mixed approach with discrete only or continuous only models.

In addition to the methodological contributions, we demonstrate the applicability of the model on a concrete case study. It comes from a survey about transportation mode preferences (Bierlaire et al., 2011). Among other variables, a set of adjectives characterizing transportation modes were collected and classified into themes. By selecting the subset of adjectives related to comfort, we first show that the method can be used to assign a numerical value to each adjective. Second, we use these numerical scores as indicators of the perception of comfort in public transportation in a hybrid choice model (HCM) (Walker, 2001; Walker and Ben-Akiva, 2002; Ben-Akiva et al., 2002) of transportation modes.

The paper is structured as follows. Section 2 presents the theoretical modeling framework. Section 3 shows an application of the methodology on a case study. In particular, we give a description of the available data, we show an application of the quantification procedure to several adjectives describing comfort and we present a comparative analysis of modeling approaches involving discrete, continuous or mixed ratings of adjectives. Finally we propose an illustration of how the inferred values of the adjectives can be used to measure perception of comfort of public transportation in transportation mode choice model. Section 4 discusses implications of the present research.

2 The integrated model framework

The aim of this research is to develop a model that can explain heterogeneity in the rating process of adjectives on the scale of a perceptional variable. We use this model to infer a numerical value for each adjective, removing potential bias using socio-economic characteristics of the respondents. We denote this model as the 'quantification model'. As an illustration of the benefits of the model, we show how the inferred values can be integrated into a discrete choice model.

In this section we describe the type of data necessary to apply the proposed method, the specification and estimation procedure of the quantification model and its integration into the HCM framework.

2.1 Adjectives

The context of our work is the use of semi-open questions in surveys, where respondents are asked to report adjectives characterizing an object, a service, or an experience (Kaufmann et al., 2010, 2001). It is called 'open' as the adjectives are reported spontaneously (as opposed to selected in a predefined list), and 'semi' because the respondent is constrained to report only adjectives (as opposed to any sentence). The adjectives reveal us the perception of the respondent about some aspects of the object, the service, or the experience. For example, we may have collected adjectives reporting perceptions of the size of a car (e.g. 'small', 'large', 'compact'), the efficiency of a delivery service (e.g.

‘fast’, ‘unreliable’, ‘good’), or the taste of a candy (e.g. ‘sweet’, ‘sour’, ‘strong’). In the case study presented below, we consider the perception of comfort in public transportation. We use also this example to describe the methodology, in order to avoid being too abstract. But the methodology is clearly not restricted to this example. At this point, we assume that we have access to a catalog of reported adjectives related to the concept of interest (i.e. comfort).

2.2 Evaluators

In order to associate numerical values with these adjectives, we recruit a set of evaluators, and ask them to rate each adjective on the scale of comfort. In our experiment, some evaluators were asked to rate the adjective on a scale ranging from -2 to 2, and some others on a scale ranging from -1000 to 1000. In the following, we call the former a ‘discrete scale’ and the latter a ‘continuous scale’, and the large range of values has been designed to decrease the limitations of the discrete scale. We denote by $J^D_{\ell m}$ the discrete scale reported by evaluator m for adjective ℓ, and $J^C_{\ell m}$ the continuous scale.

2.3 Quantification model

The quantification model is designed to assign values of adjectives related to a given concept.

Let c_ℓ be the real value of an adjective ℓ on the scale of comfort, unknown to the analyst. Each evaluator m has a different subjective perception of this value, denoted by $J^*_{\ell m}$ and also unknown to the analyst.

The quantification model of the adjectives is a latent variable model (LVM). It consists of a structural equation and two types of measurement equations: a measurement equation based on the discrete scale and another based on the continuous scale.

2.3.1 Specification

The structural equation explains the value $J^*_{\ell m}$ of adjective ℓ as perceived by evaluator m using socio-economic variables X_m as shown in equation (1).

$$J^*_{\ell m} = c_\ell + g(X_m;\gamma) + \sigma_\gamma \xi_\gamma, \text{ with } \xi_\gamma \sim \mathcal{N}(0,1), \tag{1}$$

where γ and σ_γ are parameters to estimate and g is a function of X_m and γ.

We have two types of measurement equations. The continuous measurement $J^C_{\ell m}$ is expressed by the following measurement equation:

$$J^C_{\ell m} = J^*_{\ell m} + \sigma_C \xi_C, \text{ with } \xi_C \sim \mathcal{N}(0,1), \tag{2}$$

where the error term $\sigma_C \xi_C$ captures various measurement errors, such as the bias due to the phrasing of the question, and the nature of the continuous scale. Let us note that for numerical convenience, $J^C_{\ell m}$ is equal to the continuous measure of adjective ℓ reported by respondent m, after a division by 500.

In order to capture heterogeneity in the response behavior of individuals, σ_C is assumed to be a function of socio-economic characteristics X_m. The reason for this specification comes from social science studies, where it is shown that some survey respondents tend to provide more extreme answers, while others are more moderate, to express the same opinion (Schuman and Presser, 1996). It has been shown in previous research that heterogeneity of response behavior can be captured by considering different scale parameters for different individuals (Glerum and Bierlaire, 2012). In the present work, we consider a structural function s_C for the scale:

$$\sigma_C = s_C(X_m; \beta_C), \tag{3}$$

where β_C is a list of parameters to estimate.

Similarly, we define $J^D_{\ell m}$ as the discrete measurement of intrinsic value $J^*_{\ell m}$. The corresponding measurement equation is given as follows:

$$J^D_{\ell m} = \begin{cases} -2 & \text{if } J^*_{\ell m} + \sigma_D \xi_D \leq \tau_{1\ell} \\ -1 & \text{if } \tau_{1\ell} < J^*_{\ell m} + \sigma_D \xi_D \leq \tau_{2\ell} \\ 0 & \text{if } \tau_{2\ell} < J^*_{\ell m} + \sigma_D \xi_D \leq \tau_{3\ell} \\ 1 & \text{if } \tau_{3\ell} < J^*_{\ell m} + \sigma_D \xi_D \leq \tau_{4\ell} \\ 2 & \text{if } \tau_{4\ell} < J^*_{\ell m} + \sigma_D \xi_D \end{cases} \tag{4}$$

where $\tau_{1\ell}$, $\tau_{2\ell}$, $\tau_{3\ell}$ and $\tau_{4\ell}$ are adjective-specific thresholds to estimate and the error term $\sigma_D \xi_D$ captures various measurement errors, such as the bias due to the phrasing of the question, and the nature of the discrete scale.

Parameter σ_D is defined as a function s_D of socio-economic characteristic of the respondent, similarly as in the continuous case:

$$\sigma_D = s_D(X_m; \beta_D), \tag{5}$$

where β_D is a list of parameters to estimate.

2.3.2 Estimation

A graphical representation of the quantification model of an adjective ℓ is provided in Figure 1. The parameters of the model are estimated using maximum likelihood techniques. The likelihood function $\mathscr{L}_\ell$ for adjective ℓ is given by the following formula:

$$\mathscr{L}_\ell = \prod_{m=1}^{M} \int_{J^*_{\ell m}} f(J^C_{\ell m}|J^*_{\ell m}, X_m; \beta_C) P(J^D_{\ell m}|J^*_{\ell m}, X_m; \tau_{1\ell}, \tau_{2\ell}, \tau_{3\ell}, \tau_{4\ell}, \beta_D) f(J^*_{\ell m}|c_\ell, \gamma, \sigma_\gamma) dJ^*_{\ell m}, \tag{6}$$

where M is the total number of evaluators, β_C, β_D, $\tau_{1\ell}$, $\tau_{2\ell}$, $\tau_{3\ell}$, $\tau_{4\ell}$, c_ℓ, γ, σ_γ are parameters to estimate, $f(J^C_{\ell m}|J^*_{\ell m}, X_m; \beta_C)$ and $P(J^D_{\ell m}|J^*_{\ell m}, X_m; \tau_{1\ell}, \tau_{2\ell}, \tau_{3\ell}, \tau_{4\ell}, \beta_D)$ are respectively the the PDF and PMF of $J^C_{\ell m}$ and $J^D_{\ell m}$, and $f(J^*_{\ell m}|c_\ell, \gamma, \sigma_\gamma)$ is the PDF of $J^*_{\ell m}$.

The estimate of parameter c_ℓ provides the value of adjective ℓ, for $\ell = 1, \ldots L$, where L is the total number of evaluated adjectives. In the next section, we discuss how these scores are used as measurement indicators of a latent perceptional variable in the HCM framework.

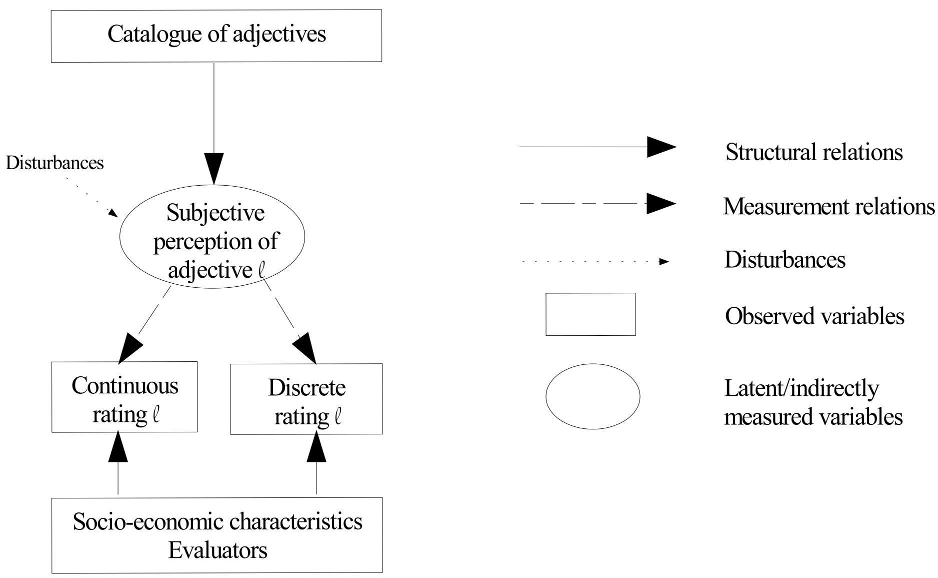

Figure 1: Quantification model.

2.4 Hybrid choice model

The values c_ℓ of the adjectives (see section 2.3) can be used as indicators of a latent perceptional variable. In this section, we present the specification and estimation procedure of the HCM that integrates these values.

2.4.1 Discrete choice model

We consider a standard DCM, where the choice of the alternatives is additionally influenced by the effect of a latent variable, e.g. a perception.

The utility U_{in} of an alternative i for a decision-maker n is expressed as a function V of observed attributes X_{in} of i and n and of a latent attribute X^*_{in}:

$$U_{in} = V(X_{in}, X^*_{in}; \beta) + \varepsilon_{in}, \text{ with } \varepsilon_{in} \sim \text{EV}(0,1), \tag{7}$$

where β is a list of parameters to estimate.

2.4.2 Latent variable model

The latent variable X^*_{in} cannot be directly observed and must be indirectly measured by means of indicators. Therefore a LVM relates X^*_{in} to a list of K_i adjectives I^*_{kin} reported by decision-maker n, for $k = 1, \ldots, K_i$.

The structural component of the LVM is defined by a function h, relating socio-economic information X_n of individual n to the latent variable X^*_{in}:

$$X^*_{in} = h(X_n; \lambda) + \sigma_\omega \omega_n, \text{ with } \omega_n \sim \mathcal{N}(0,1), \tag{8}$$

where λ and σ_ω are parameters to estimate.

The measurement component of the LVM is described by a set of functions r_{ki}, relating the latent variable X^*_{in} to its indicators I^*_{kin}:

$$I^*_{kin} = r_{ki}(X^*_{in}; \eta_{ki}) + \sigma_{ki} \nu_{kin}, \text{ with } \nu_{kin} \sim \mathcal{N}(0,1), \tag{9}$$

where η_{ki} and σ_{ki} are parameters to estimate, for $k = 1, \ldots, K_i$. Note that we allow these parameters to vary with k, since the order in which the adjectives are reported may reveal some aspect of the latent perception.

2.4.3 Estimation

Figure 2 shows a diagram of the integrated model framework, which summarizes the three components introduced in this section, that is, the discrete choice model, the LVM that characterizes the perceptional variable and the LVM that quantifies the adjectives (quantification model of Figure 1). Building upon the framework developed by Walker and Ben-Akiva (2002), latent variables are represented by ovals, observed variables by rectangles, structural relations by straight arrows, measurement relations by dashed arrows and disturbances are related to the latent variables by dotted arrows.

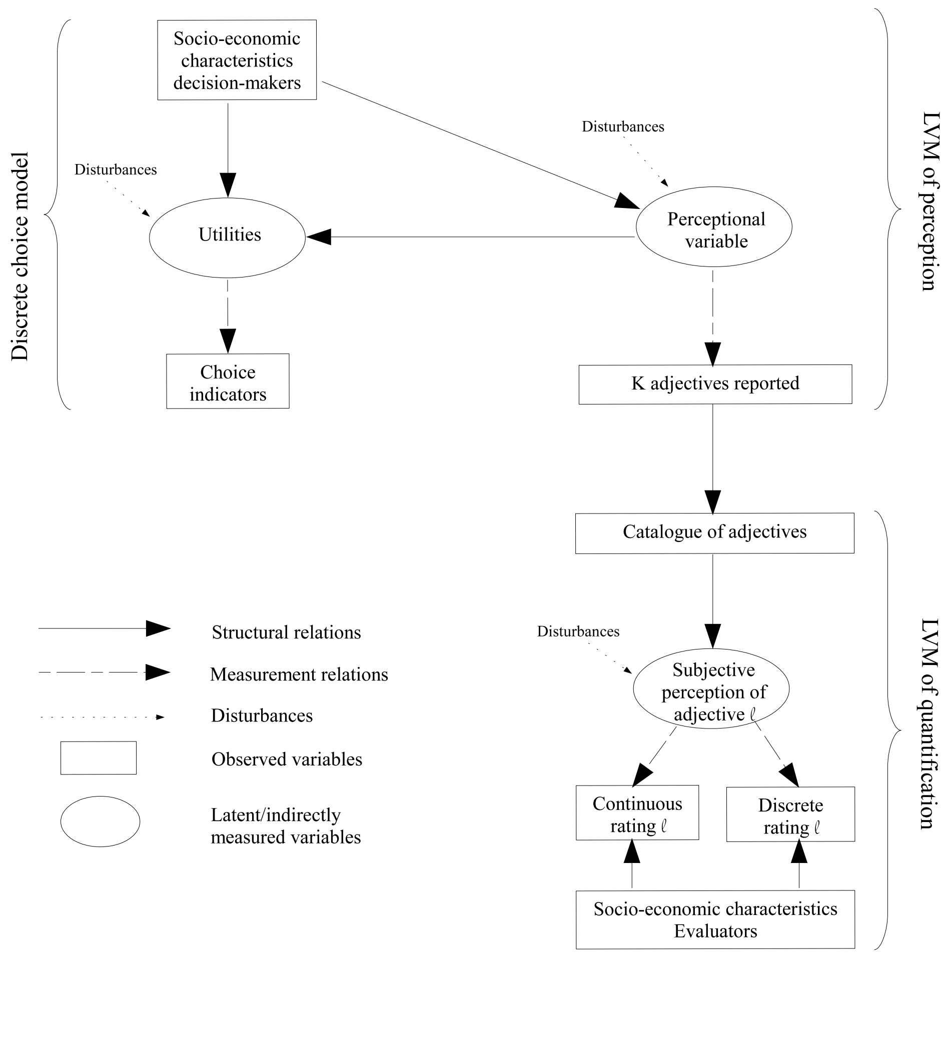

Figure 2: Integrated model framework.

To estimate the parameters in the framework, two methods are possible: the full information estimation technique or the sequential estimation technique. We describe both techniques hereafter.

The expression of the full information likelihood is the following:

$$\mathscr{L}_{\text{full}} = \prod_{n=1}^{N} \int_{X_{in}^*} \prod_{i=1}^{J_n} P(i|X_{in}, X_{in}^*; \beta)^{y_{in}} \cdot f(X_{in}^*|X_n; \lambda, \sigma_\omega) \cdot \prod_{k=1}^{K_i} f(I_{kin}^*|X_{in}^*; \eta_{ki}, \sigma_{ki}) \prod_{\ell=1}^{L} \mathscr{L}_\ell^{H_{ki\ell n}} dX_{in}^*, \tag{10}$$

where J_n is the number of choice alternatives for n, $P(i|X_{in}, X_{in}^*; \beta)$ is the probability that respondent n chooses alternative i, y_{in} is an indicator that i is chosen by n, $f(X_{in}^*|X_n; \lambda, \sigma_\omega)$ is the PDF of X_{in}^* and $f(I_{kin}^*|X_{in}^*; \eta_{ki}, \sigma_{ki})$ is the PDF of the kth indicator, $\mathscr{L}_\ell$ is the likelihood per adjective (see equation (6)).

The estimation of equation (10) is computationally intensive, since it involves the evaluation of several integrals.

To ease this computational burden, it is possible to consider a sequential estimation. The drawback of this approach is a loss of efficiency of the estimator. The first step of this approach consists of estimating $\mathscr{L}_\ell$ for all adjectives ℓ. The second step involves the estimation of the following partial likelihood function.

$$\mathscr{L}_{\text{partial}} = \prod_{n=1}^{N} \int_{X_{in}^*} \prod_{i=1}^{J_n} P(i|X_{in}, X_{in}^*; \beta)^{y_{in}} \cdot f(X_{in}^*|X_n; \lambda, \sigma_\omega) \cdot \prod_{k=1}^{K_i} f(\hat{I_{kin}}|X_{in}^*; \eta_{ki}, \sigma_{ki}) dX_{in}^*, \tag{11}$$

where $\hat{I_{kin}}$ results from the simple application of equation (12) below.

$$\hat{I_{kin}} = \sum_{\ell=1}^{L} \hat{c_\ell} \cdot H_{ki\ell n}, \tag{12}$$

where $\hat{c_\ell}$ is the fitted value of c_ℓ which is obtained from the maximization of $\mathscr{L}_\ell$, and L is the total number of adjectives (related to latent variable X_{in}^*) which were reported by the sample of respondents. To be more precise, L is the result of the count of all adjectives related to X_{in}^* which were reported at least by one of the decision-makers. The expression $H_{ki\ell n}$ is defined as follows.

$$H_{ki\ell n} = \begin{cases} 1 & \text{if } \ell \text{ is selected by } n \text{ for indicator } k \text{ of the perceptional variable} \\ & \text{related to alternative } i \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

3 A case study: mode choice in Switzerland

The framework presented in section 2 is applied on a case study, which we describe in this section. We first present the two different surveys which were conducted to collect the necessary data. We subsequently present the quantification model, a comparative analysis between the use of discrete and continuous scales for the measurement of adjectives and an application of the integrated model framework of Figure 2.

3.1 Revealed preferences survey

A revealed preferences (RP) survey was conducted in the framework of a joint project between PostBus, an important bus company in Switzerland, and EPFL's Transportation Center. Information on all trips performed in one day by inhabitants of suburban regions of Switzerland was collected. Respondents were surveyed in German or French, depending on the language region they were living in.

Based on the assumption that qualitative factors such as perceptions influence the choice, the survey also included a question about respondents' perceptions of transportation modes. Table 1 reports this particular survey question. For each transportation mode a respondent was asked to provide three adjectives that characterize it best in his opinion.

With the help of social scientists, the adjectives were then classified into eleven themes: comfort, perception of time, perception of cost, difficulty of access, flexibility, efficiency, reliability, environmental impact, appreciation, feeling and look (Bierlaire et al., 2011).

To illustrate our methodology we focus only on the perception of comfort in public transportation (PT). We hence consider adjectives that are (i) related to comfort only and (ii) reported in rows 2, 3 and 4 of Table 1. The specfic theme of comfort was selected due to an important range of relevant adjectives (22 in total).

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: The survey question designed to evaluate individuals' perceptions of the different transportation modes.

In order to use these data in a model, it is necessary to assign a numerical value to each adjective. For that purpose, an additional survey is performed.

3.2 Adjective quantification survey

A second survey was conducted to quantify the adjectives reported in the RP survey (Glerum et al., 2014). Our approach consists of asking additional subjects (called the 'evaluators') to rate the strength of each adjective on a scale of comfort. A positive (resp. negative) rating implies that the evaluator thinks that the adjective characterizes a positive (resp. negative) perception of comfort.

We moreover consider two types of scale. Precisely, some of the evaluators are asked to give a rating on a scale ranging from -2 to 2, while the remaining evaluators are asked to give a rating on a scale ranging from -1000 to 1000. A snapshot of part of the survey is shown in Figure 3, for an example of continuous scale.

The survey was conducted online and 277 questionnaires were eventually collected. The sample mostly consisted of students and employees of Ecole Polytechnique Fédérale de Lausanne (EPFL). Although the adjectives were originally collected in German or French, they were translated into English, since many individuals in the sample did not speak the former languages. Summary statistics on the respondents' socio-economic characteristics are shown in Tables 2 and 3.

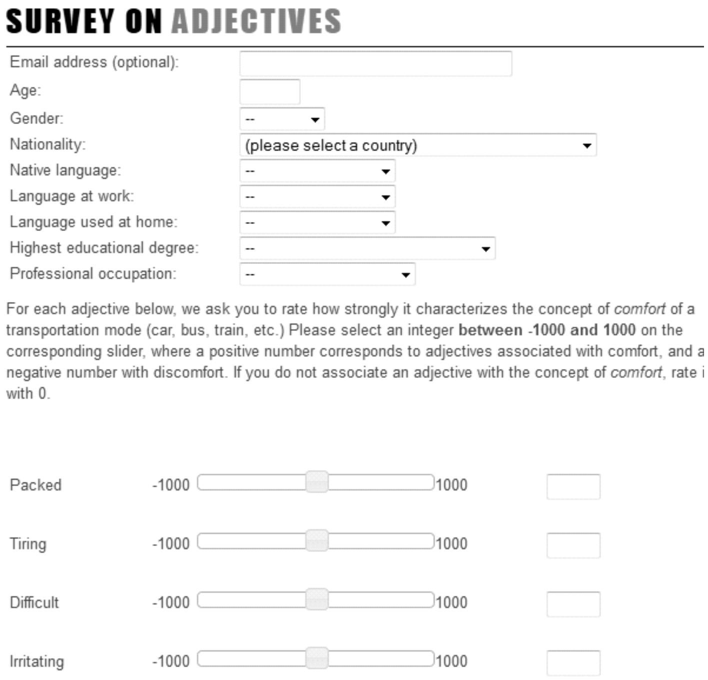

Figure 3: Snapshot of part of the online questionnaire, in the case of the continuous scale.

Education		**Gender**		**Nationality (continued)**	
Professional school	3.6%	Male	60.6%	Iran	1.8%
High school	17.7%	Female	39.4%	Italy	4.3%
University Bachelor Degree	30.7%	**Nationality**		Luxembourg	1.4%
University Master Degree	35.7%	Armenia	0.4%	Madagascar	0.4%
University PhD Degree	12.3%	Austria	0.4%	Mexico	0.4%
Profession		Belgium	1.1%	Morocco	0.4%
Employee	33.9%	Brazil	1.4%	Netherlands	1.1%
Independent worker	1.1%	Canada	0.7%	Poland	0.7%
Liberal profession	1.1%	Chile	0.7%	Portugal	1.1%
Middle management	2.2%	China	10.1%	Romania	0.7%
Retired	0.4%	Croatia	0.4%	Russian Federation	0.4%
Student	57.8%	Czech Republic	0.4%	Spain	1.1%
Top management	0.4%	El Salvador	0.4%	Sweden	1.8%
Unemployed	0.7%	Finland	1.1%	Switzerland	45.8%
Worker	1.1%	France	6.1%	Taiwan	1.8%
Other	1.4%	Germany	2.9%	Turkey	3.2%
Age		Greece	1.4%	United Kingdom	0.4%
Average	28.71	Hungary	0.4%	United States	1.4%
Standard deviation	10.42	India	2.2%	Other	1.8%

Table 2: Summary statistics of the socio-economic characteristics of the respondents of the adjective quantification survey. (sample size = 277).

	Native language	Home language	Work language
Arabic	0.4%	1.4%	-
Cantonese	0.7%	0.7%	0.4%
Danish	0.4%	0.4%	-
Dutch	1.4%	0.4%	-
English	3.2%	7.9%	35.7%
Finnish	1.1%	0.7%	-
French	45.8%	51.6%	53.8%
German	9.0%	5.8%	2.2%
Greek	1.8%	1.4%	-
Italian	5.1%	4.0%	0.7%
Mandarin chinese	11.2%	11.2%	5.8%
Persian	1.8%	1.4%	-
Portuguese	2.5%	0.4%	-
Russian	0.7%	0.4%	-
Spanish	2.5%	2.2%	0.4%
Swedish	1.8%	1.1%	-
Turkish	3.2%	2.9%	-
Other	7.2%	6.1%	1.1%

Table 3: Summary statistics of the languages of the respondents of the adjective quantification survey. (sample size = 277).

Figures 4 and 5 show histograms of the ratings of the evaluators for all adjectives[1] associated with comfort, for the discrete and continuous scales, respectively. Summary statistics for both scales are provided in Table 7 in the appendix. For many adjectives, there is a general agreement about the positive (e.g. ‘comfortable’, ‘restful’) or negative (e.g. ‘uncomfortable’, ‘stressful’) connotation of the adjective. For some of the adjectives, the consensus is particularly strong (e.g. ‘comfortable’, ‘suffocating’), while for others, more neutral responses may occur. Such responses may be due to some ambiguity in the meaning of the adjective (e.g. ‘hardly full’) or lack of concern (e.g. ‘unsuitable with strollers’ may only affect individuals with children).

For positive, negative or neutral adjectives, heterogeneity in the responses is clearly present. For some adjectives, the effect is more visible when the measurement is per-

[1]We note that some of the expresssions (such as ‘bad air’) reported by the respondents of the RP survey are not adjectives. Since they are freely reported by the respondents, we decided to keep them in the analysis.

formed on the continuous scale rather than on the discrete scale (e.g. 'restful'). This motivate the formulation of the quantification model presented in section 2.3.

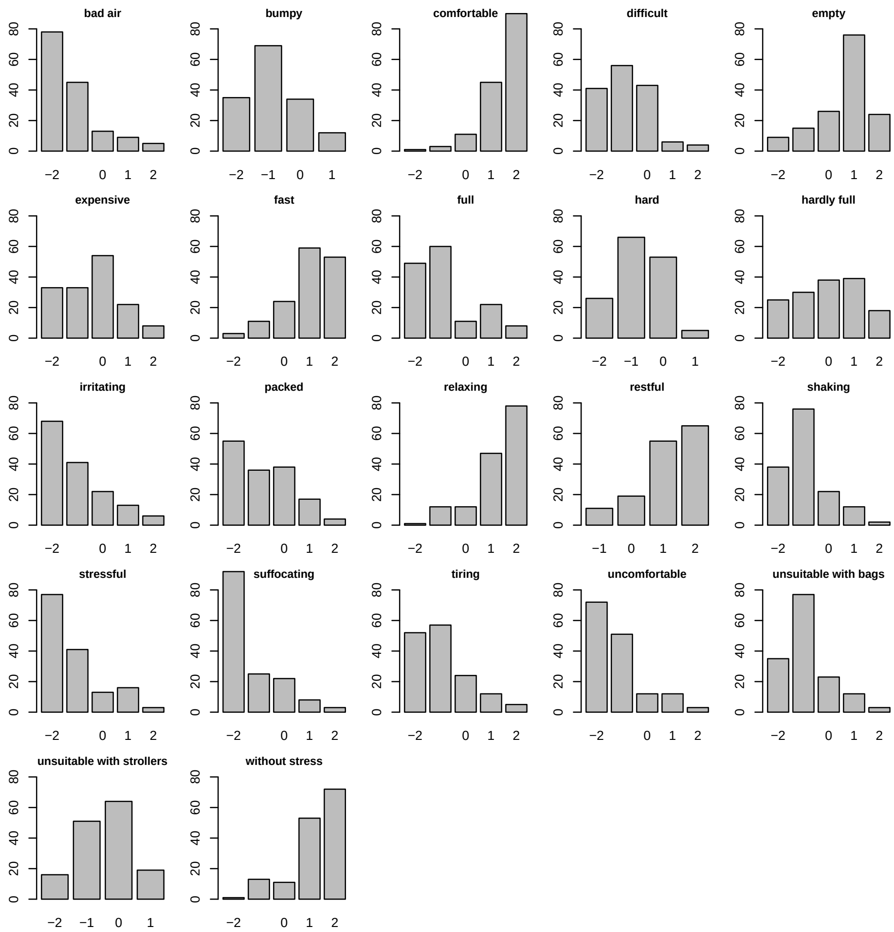

Figure 4: Histograms of the discrete ratings of the evaluators for each adjective (sample size = 150).

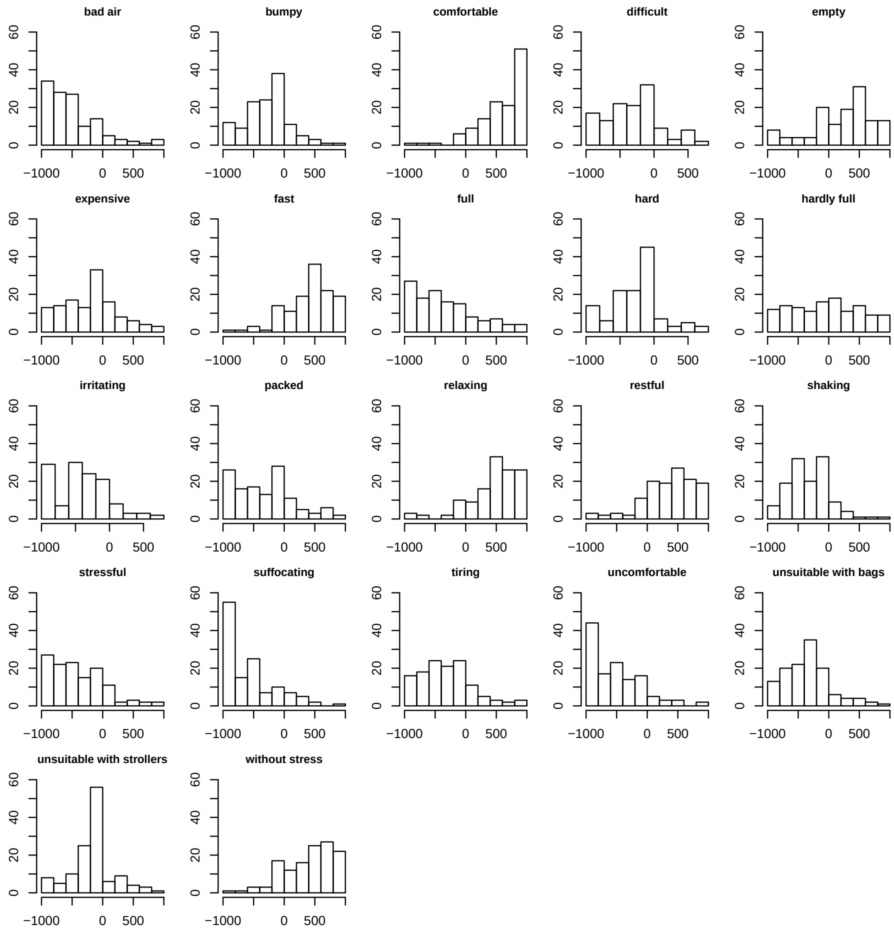

Figure 5: Histograms of the continuous ratings of the evaluators for each adjective (sample size = 127).

3.3 Quantification of adjectives characterizing comfort

For each adjective, we hence developed a LVM as described in section 2.3. This section describes the specification of the models and their estimation results.

3.3.1 Specification

Exploratory analysis on the data showed that individuals provide different ratings depending on their age. In addition, Chinese- or Mandarin Chinese-speaking individuals tend not to provide responses in the extreme boundaries of the scale and for some adjectives (i.e. 'bad air', 'expensive', 'hardly full', 'restful'). Finally, male respondents show slightly different response patterns from female respondents.

Based on these preliminary results and extended specification testing, we obtained the below specification of quantification model. In particular in this specification we have imposed the following constraints on the parameters: $\sigma_\gamma = 0$ and $\sigma_D = \sigma_C$.

The structural equation (1) of the LVM for each adjective is given as follows.

$$J^*_{\ell m} = c_\ell + \beta_{\text{LangCN}}\text{LangCN} + \beta_{\text{Female}}\text{Female}, \tag{14}$$

where *LangCN* is an indicator of whether the respondent is Chinese- or Mandarin Chinese-speaking and *Female* is an indicator of whether the respondent is a woman. The measurement equations are specified as in equations (2) and (4), with the following expression of $\sigma := \sigma_D = \sigma_C$:

$$\sigma = \bar{\sigma} + \beta_{\text{Age}<27} \cdot \text{Age} \cdot (\text{Age} < 27) + \beta_{\text{Age}\geq 27} \cdot \text{Age} \cdot (\text{Age} \geq 27), \tag{15}$$

where *Age* denotes the respondent's age, expression $(\text{Age} < 27)$ is an indicator of whether the respondent is younger than 27 years, expression $(\text{Age} \geq 27)$ indicates whether the respondent is 27 years old or older, and $\bar{\sigma}$, $\beta_{\text{Age}<27}$ and $\beta_{\text{Age}\geq 27}$ are parameters to estimate.

Since the discrete ratings are symmetrically labeled, we constrain the threshold parameters in equation (4) to be symmetric and centered around 0 (see also Johnson, 2003). For that reason, we estimate parameters $\delta_{1\ell}$ and $\delta_{2\ell}$ instead of $\tau_{1\ell}, \ldots, \tau_{4\ell}$. The relations between these parameters and the thresholds are reported below:

$$\begin{aligned}
\tau_{1\ell} &= -\delta_{1\ell} - \delta_{2\ell} \\
\tau_{2\ell} &= -\delta_{1\ell} \\
\tau_{3\ell} &= \delta_{1\ell} \\
\tau_{4\ell} &= \delta_{1\ell} + \delta_{2\ell}
\end{aligned} \tag{16}$$

A natural way of capturing extreme versus moderate response behavior would be to include all variables (i.e. *LangCN*, *Female* and *Age*) in the expression of the scale σ. However the above specification of the LVM (which includes home language and gender in the expression of the shift) happened to better fit the data. In fact, the scale's symmetry imposes some restrictivity. For example, it assumes that effects such as extreme response behavior impact the rating on both sides of the mean rating.

3.3.2 Estimation results

The parameters of the quantification model of each adjective can be estimated by maximizing the likelihood in equation (6). The estimation results are shown in Table 4.

We can draw the following conclusions from the estimation results:

	c_l		β_{LangCN}		β_{Female}		$\bar{\sigma}$		$\beta_{\text{Age}<27}$		$\beta_{\text{Age}\geq 27}$		δ_{1l}		δ_{2l}	
Adjective name	Value	SD	Value	SD	Value	SD	Value	SD	Value	SD	Value	SD	Value	SD	Value	SD
bad air	-0.934	0.0773	0.378	0.18	-0.266	0.114	-0.692	0.231	0.0291	0.0108	0.0107	0.00552	0.186	0.0479	0.727	0.0845
bumpy	-0.602	0.0643	0.217*	0.142	0.00408*	0.0903	-0.713	0.252	0.0189	0.0112	0.00774*	0.00616	0.3	0.0475	0.852	0.0899
comfortable	1.29	0.0739	-0.36	0.151	0.0229*	0.102	-0.859	0.275	0.027	0.013	0.0148	0.00649	0.272	0.0719	0.784	0.0918
difficult	-0.667	0.0684	0.323	0.151	0.0322*	0.0995	-0.796	0.245	0.0256	0.0111	0.0134	0.00605	0.421	0.0585	0.672	0.0813
empty	0.466	0.0795	-0.253*	0.188	0.223	0.122	-0.766	0.259	0.0351	0.0116	0.017	0.00626	0.265	0.0493	1.11	0.115
expensive	-0.417	0.0782	0.0807*	0.179	0.127*	0.12	-0.23*	0.241	0.00809*	0.011	0.00303*	0.00573	0.462	0.0607	0.625	0.0843
fast	0.912	0.0704	-0.103*	0.148	-0.131*	0.0992	-0.857	0.255	0.0271	0.0116	0.0156	0.00628	0.272	0.0504	0.804	0.0847
full	-0.693	0.0858	0.479	0.199	-0.0644*	0.128	-0.339*	0.268	0.0175*	0.0121	0.00665*	0.00658	0.122	0.0355	0.97	0.1
hard	-0.601	0.0612	0.438	0.129	-0.0203*	0.0864	-1.06	0.255	0.03	0.0112	0.0171	0.0063	0.46	0.0579	0.787	0.0901
hardly full	-0.178	0.0874	-0.113*	0.21	0.411	0.137	-0.277*	0.245	0.018	0.011	0.00741*	0.006	0.352	0.0556	0.816	0.0996
irritating	-0.85	0.0718	0.631	0.158	0.0541*	0.104	-0.521	0.225	0.0152*	0.0104	0.00624*	0.00548	0.234	0.0464	0.59	0.073
packed	-0.675	0.0796	0.538	0.185	-0.112*	0.119	-0.604	0.256	0.0269	0.0114	0.0114	0.00637	0.381	0.0577	0.628	0.0823
relaxing	0.956	0.0752	-0.22*	0.165	0.102*	0.107	-0.596	0.282	0.0198*	0.0131	0.00839*	0.0067	0.164	0.0442	0.776	0.085
restful	0.826	0.0732	-0.343	0.154	0.194	0.105	-0.718	0.272	0.0226	0.0123	0.0126	0.00667	0.232	0.0488	0.809	0.0871
shaking	-0.618	0.0596	0.112*	0.128	-0.0101*	0.0838	-1.15	0.254	0.034	0.0114	0.0187	0.0064	0.199	0.0391	0.853	0.0834
stressful	-0.888	0.0777	0.687	0.168	-0.117*	0.113	-0.497	0.229	0.0168*	0.0108	0.00764*	0.00534	0.154	0.0405	0.7	0.0815
suffocating	-1.13	0.0799	0.509	0.176	-0.104*	0.118	-0.653	0.227	0.0252	0.0105	0.0115	0.00541	0.332	0.0631	0.483	0.0752
tiring	-0.718	0.0722	0.478	0.164	-0.0065*	0.105	-0.797	0.242	0.0302	0.0111	0.0132	0.00583	0.249	0.0474	0.748	0.0833
uncomfortable	-1.12	0.0791	0.649	0.172	0.0802*	0.112	-0.712	0.221	0.028	0.0104	0.0123	0.00528	0.174	0.0469	0.862	0.0912
unsuitable with bags	-0.683	0.0671	0.451	0.143	-0.116*	0.0977	-0.462	0.212	0.0077*	0.00977	0.00501*	0.0051	0.234	0.0451	0.974	0.0954
unsuitable with strollers	-0.311	0.0587	0.162*	0.13	-0.0538*	0.0885	-0.923	0.258	0.0244	0.0112	0.015	0.0065	0.431	0.0518	0.813	0.0999
without stress	0.941	0.0719	-0.468	0.149	0.00696*	0.102	-0.665	0.262	0.0182*	0.0122	0.011	0.00647	0.138	0.0392	0.811	0.0846

Table 4: Parameter estimates of the quantification models, with standard deviations (SD) (* Statistical significance $<$ 90%).

- All constants c_ℓ are significant and have the expected sign, i.e. the coefficient of an adjective intuitively related to discomfort has a negative sign, while the coefficient of an adjective related to comfort has a positive sign. For example, adjective 'empty' has the expected positive sign, since travelers have more space in an empty transportation mode and hence feel more comfortable in it.
- Among the coefficients interacting the intercept, it can be noticed that Chinese- or Mandarin Chinese-speaking individuals give less extreme answers than individuals with other home languages. For some adjectives (i.e. 'bad air', 'empty', 'hardly full', 'restful'), female respondents tend to assign values which are closer to the boundaries than male respondents. For both variables, the related coefficient have indeed a sign which is the opposite from the sign of c_ℓ.
- If we now consider the coefficients interacting the scale, it can be seen that the older individuals become the more extreme they are in their responses. However this effect is stronger for individuals below 27 years and weaker above that age.

Alhough the significance of the parameters may vary from one adjective to another, we decided to keep a generic specification across adjectives, to show that some factors affect the rating behavior in a consistent way.

3.4 Comparison between discrete and continuous scales

One of the main goals of this paper is to provide a comparative analysis of the use of different scale types (discrete versus continuous). For that purpose, we estimated two other models: a *continuous-scale model*, which is a restricted version of the quantification model of section 3.3, which only includes the continuous measurement equation, and a *discrete-scale model*, which a restriction of the same model but that only includes the discrete measurement equation.

Consequently, several indices were defined to assess the fit of each of the three models. They include (1) the number of unsignificant parameters for each model, (2) the number of observations for which the probability of choosing a discrete rating is above the chance level ($= 1/5$) for the discrete-scale model and the quantification model (on a forecasting sample) and (3) the loglikelihood of the continuous part of the continuous-scale model and the quantification model.

We first investigate the number of unsignificant parameters for each model. In the computation, we only account for the constant c_l and parameters relative to socio-economic characteristics of the evaluator, i.e. β_{LangCN}, β_{Female}, $\beta_{\text{Age}<27}$, $\beta_{\text{Age}\geq 27}$. The other parameters are not counted since they are not present in all three models. Figure 6 provides a graphical representation of the index. A first observation is that the discrete-scale model performs very poorly, since at least 3 out of the 5 parameters are unsignificant for all adjectives. A second observation is that for some adjectives, the continuous-scale model performs better than the quantification model and for the rest of the adjectives, the reverse occurs. On average, they perform similarly, since the average number of unsignificant parameters is 1.81 for the continuous-scale model and 1.86 for the quantification model.

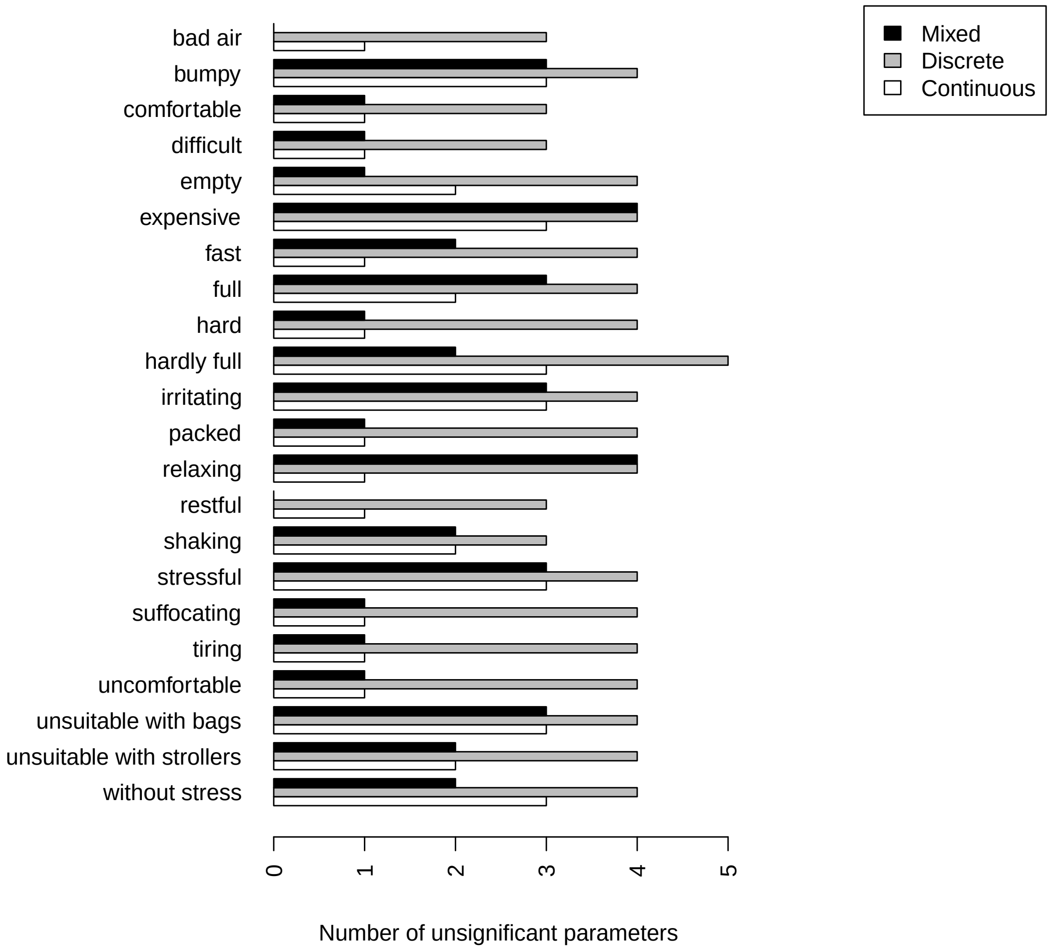

Figure 6: Number of unsignificant parameters in the quantification model, the continuous-scale model and the discrete-scale model (at a 90% confidence level).

The second and third indices aim at assessing the forecasting capability of the models. To compute them, we have first estimated the model of interest (i.e. the discrete-scale model or the quantification model) on a random sample consisting of 80% of the observations of the whole data set. Consequently we have applied the model on the remaining 20% of the observations. We describe their computation in details hereafter.

The second index is the number of observations (in the sample containing 20% of the data) where the predicted probability of selecting the discrete rating which was given by the evaluator is above the chance level. It is computed for both the quantification model and the discrete-scale model (see the second and third columns of Table 5). The average number of observations with probabilities above the chance level is slightly higher for the quantification model than for the discrete-scale model, showing a slightly better prediction power. Moreover, since the standard deviation of this index is smaller for the quantification model, this shows an increased consistency of the model specification across adjectives.

The third index is the log-likelihood of the continuous component of the quantification model and the continuous-scale model (see the fourth and fifth columns of Table 5). We observe a slightly better fit for the quantification model, since the log-likelihood of the quantification model is slightly higher. In addition, the standard deviation of this index is smaller for the quantification model, showing a more consistent prediction capability across adjectives.

	Number of observations with probabilities > chance level		Log-likelihood for the continuous part	
Adjective	Quantification model	Discrete-scale model	Quantification model	Continuous-scale model
bad air	25	24	-24.86	-25.72
bumpy	22	17	-21.72	-21.66
comfortable	29	29	-21.21	-20.22
difficult	30	27	-28.57	-29.75
empty	20	21	-30.02	-29.59
expensive	25	22	-26.51	-26.26
fast	24	24	-22.31	-21.98
full	22	21	-31.53	-32.36
hard	24	24	-26.92	-26.72
hardly full	18	18	-31.28	-32.37
irritating	23	22	-26.32	-26.60
packed	26	27	-35.66	-36.30
relaxing	25	25	-21.52	-21.92
restful	25	24	-27.08	-27.60
shaking	25	21	-18.56	-18.15
stressful	23	23	-24.63	-25.21
suffocating	17	16	-24.22	-23.71
tiring	24	24	-26.97	-28.30
uncomfortable	26	26	-24.33	-23.42
unsuitable with bags	24	20	-25.55	-27.18
unsuitable with strollers	24	24	-25.30	-25.76
without stress	28	28	-25.69	-26.55
Average	24.05	23.05	-25.94	-26.24
Standard deviation	3.09	3.42	3.89	4.27

Table 5: Fit indices resulting from the estimation of the model on 80% of the data and from its application on 20% of the data (sample size = 55).

This comparative analysis shows that both the quantification model and the continuous-scale model perform better than the discrete-scale model. The quantification model is slightly better due to an increased accuracy in its prediction capability.

3.5 Application to a transportation mode choice case study

In this section we provide an illustration of the use of the outcomes of the models of section 3.3. In the following example, we use the values c_ℓ of the adjectives as indicators of the perception of comfort of PT (see section 3.1).

We formulate a HCM as described in section 2.4. We assume that the respondents to the RP survey have the choice among three transportation modes, i.e. private motorized modes (PMM), public transportation (PT) and soft modes (SM). Moreover we consider nine indicators of the perception of comfort of PT, i.e. 3 different PT modes $\times$ 3 reported adjectives. The purpose is to assess the impact of the perception of comfort of PT on mode choice.

The HCM specification is the same as in Glerum et al. (2014). The difference with this previous research is that we are now using the estimates of c_ℓ as numerical values for the adjectives, instead of using directly the ratings of the evaluators.

The utility of equation (7) is a linear expression that includes the following variables: the cost and duration of the trip, the trip distance (for alternative SM only), an indicator that states whether the trip is work-related only, a indicator of living in a French-speaking region of Switzerland and a variable capturing the perception of comfort in PT, which is interacted with time.

The structural equation (8) of the LVM is a linear function of a domicile in a French-speaking region, an age below 50 years, an indicator of a full- or part-time job and the number of cars in the respondent's household.

Name	Value	SD	Name	Value	SD
DCM parameters			$\beta_{\text{French}_{\text{PT}}}$	-0.0525*	0.306
ASC_{PT}	-0.161*	0.201	$\beta_{\text{French}_{\text{PMM}}}$	0.964	0.271
ASC_{PMM}	0.42	0.184	β_{comfort}	1.33	0.3
β_{cost}	-0.0654	0.00806	LVM parameters		
$\beta_{\text{time}_{\text{PT}}}$	-0.0203	0.0028	λ_{mean}	7.2	0.688
$\beta_{\text{time}_{\text{PMM}}}$	-0.0323	0.00342	λ_{French}	-0.706	0.288
β_{distance}	-0.235	0.0205	$\lambda_{\text{age}_{50}}$	-1.12	0.225
$\beta_{\text{work}_{\text{PT}}}$	-0.044*	0.234	λ_{active}	-1.15	0.242
$\beta_{\text{work}_{\text{PMM}}}$	-0.575	0.221	λ_{cars}	-0.71	0.226

Table 6: Parameter estimates of the choice model and structural model components of the HCM, with standard deviations (SD) (* Statistical significance $< 90\%$).

The model is estimated using the sequential approach described in section 2.4.3. The parameter estimates of the choice model component and of the structural equation for the perception of comfort in PT are displayed in Table 6. The estimates of the measurement model are shown in Table 8 in the appendix. As expected, the positive sign of parameter β_{comfort} and its significance show that a change in individuals' perception of comfort in PT can modify significantly their mode choices. Moreover due to the interaction of the perceptional variable with time, we can conclude that an increase of travel time has a more important (negative) effect on the choice of PT for individuals with a poor perception of comfort in PT.

4 Conclusion

In this paper, we presented a model which assigns an objective numerical value to an adjective measuring a perception. The model has the following important contributions. First, it accounts for the heterogeneity of response behavior in the rating process of adjectives. Second, it combines the use of different scale types (discrete and continuous) and happens to be more robust across adjectives than methods using a single scale type.

Besides the above methodological implications, we show that the method can be useful to obtain reliable indicators of a latent perceptional variable, where biases induced by socio-demographic differences are removed. This result is in particular useful to measure the impact of a perception in a hybrid choice model.

This research highlights two important considerations for the research on psychometrics. First, we have proposed a methodology that will allow open (or semi-open) questions to be used more often in the measurement of psychological constructs. In particular, Krosnick et al. (2005) states that ‘Because open-ended questions do not present answer choices to participants, theses sources of researcher-induced measurement error do not distort responses in principle.’ Second, heterogeneity of response behavior should systematically be accounted for in the specification of latent variable models capturing constructs such as attitudes, perceptions, etc. Furthermore, appropriate scales should be used, in order to capture correctly the construct of interest.

References

Arentze, T. A., Dellaert, B. G. C. and Chorus, C. G. (forthcoming). Incorporating mental representations in discrete choice models of travel behaviour: Modelling approach and empirical application, *Transportation Science* . Accepted for publication.

Arnold, H. J. (1981). A test of the validity of the multiplicative hypothesis of expectancy-valence theories of work motivation, *The Academy of Management Journal* **24**(1): 128–141.

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

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

Bierlaire, M., Curchod, A., Danalet, A., Doyen, E., Faure, P., Glerum, A., Kaufmann, V., Tabaka, K. and Schuler, M. (2011). Projet de recherche sur la mobilité combinée, rapport définitif de l’enquête de préférences révélées, *Technical Report TRANSP-OR 110704*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Chun, K.-T., Campbell, J. B. and Yoo, J. H. (1974). Extreme response style in cross-cultural research a reminder, *Journal of Cross-Cultural Psychology* **5**(4): 465–480.

Churchill, G. (1991). *Marketing research: methodological foundations*, Dryden Press.

Clarke III, I. (2001). Extreme response style in cross-cultural research, *International Marketing Review* **18**(3): 301–324.

Cummins, R. A. and Gullone, E. (2000). Why we should not use 5-point likert scales: The case for subjective quality of life measurement, *Proceedings of the second international conference on quality of life in cities*, pp. 74–93.

Dolnicar, S. and Grün, B. (2007). Cross-cultural differences in survey response patterns, *International Marketing Review* **24**(2): 127–143.

Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales, *Journal of Usability Studies* **5**(3): 104–110.

Glerum, A., Atasoy, B. and Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models, *Journal of Choice Modelling* **10**: 11–33.

Glerum, A. and Bierlaire, M. (2012). Accounting for response behavior heterogeneity in the measurement of attitudes: an application to demand for electric vehicles, *Proceedings of the 12th Swiss Transport Research Conference*, Ascona, Switzerland.

Holbrook, A. L., Krosnick, J. A., Visser, P. S., Gardner, W. L. and Cacioppo, J. T. (2001). Attitudes toward presidential candidates and political parties: Initial optimism, inertial first impressions, and a focus on flaws, *American Journal of Political Science* **45**(4): 930–950.

Horeni, O., Arentze, T. A., Dellaert, B. G. C. and Timmermans, H. J. P. (2010). Revealed or hidden? insights into ways of measuring mental representations online: a comparative study of apt and cnet applied to an online agent, *Proceedings of the Bi-Annual Design and Decision Support Systems Conference*, Eindhoven, The Netherlands, pp. 1–28.

Javaras, K. N. (2004). *Statistical analysis of Likert data on attitudes*, PhD thesis, University of Oxford.

Javaras, K. N. and Ripley, B. D. (2007). An ‘unfolding’ latent variable model for likert attitude data: Drawing inferences adjusted for response style, *Journal of the American Statistical Association* **102**(478): 454–463.

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style, *Psychometrika* **68**(4): 563–583.

Johnson, T. R. and Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style, *Journal of Educational and Behavioral Statistics* **35**(1): 92–114.

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

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

Krosnick, J., Judd, C. and Wittenbrink, B. (2005). The measurement of attitudes, *in* D. Albarracn, B. Johnson and M. Zanna (eds), *The Handbook of Attitudes*, Lawrence Erlbaum, Mahwah, NJ, pp. 21–76.

Leung, K. and Bond, M. H. (1989). On the empirical identification of dimensions for cross-cultural comparisons, *Journal of Cross-Cultural Psychology* **20**(2): 133–151.

Likert, R. (1932). A technique for the measurement of attitudes, *Archives of Psychology* **22**(140).

Mossholder, K. W., Settoon, R. P., Harris, S. G. and Armenakis, A. A. (1995). Measuring emotion in open-ended survey responses: An application of textual data analysis, *Journal of Management* **21**(2): 335–355.

Potkay, C. R. and Allen, B. P. (1973). The adjective generation technique: an alternative to adjective check lists, *Psychological Reports* **32**: 457–458.

Rossi, P. E., Gilula, Z. and Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach, *Journal of the American Statistical Association* **96**(453): 20–31.

Russell, C. J. and Bobko, P. (1992). Moderated regression analysis and likert scales: too coarse for comfort., *Journal of Applied Psychology* **77**(3): 336.

Sarman, I., Scagnolari, S. and Maggi, R. (2013). Acceptance of life-threatening hazards among young tourists: a stated choice experiment, *Proceedings of the 3rd International Choice Modelling Conference*, Sydney, Australia.

Schuman, H. and Presser, S. (1996). *Questions and answers in attitude surveys: experiments on question form, wording, and context*, Quantitative Studies in Social Relations, Sage Publications.

Shaftel, J., Nash, B. L. and Gillmor, S. C. (2012). Effects of the number of response categories on rating scales, *Proceedings of the annual conference of the American Educational Research Association*, Vancouver, Canada.

Ter Hofstede, F., Audenaert, A., Steenkamp, J.-B. E. M. and Wedel, M. (1998). An investigation into the association pattern technique as a quantitative approach to measuring means-end chains, *International Journal of Research in Marketing* **15**(1): 37–50.

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

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

Zax, M. and Takahashi, S. (1967). Cultural influences on response style: Comparisons of japanese and american college students, *The Journal of Social Psychology* **71**(1): 3–10.

Appendix

	Discrete ratings		Continuous ratings	
Adjective	Mean	SD	Mean	SD
bad air	-1.21	1.05	-498.42	446.77
bumpy	-0.85	0.87	-249.92	378.71
comfortable	1.47	0.77	628.40	391.59
difficult	-0.83	0.97	-283.72	414.83
empty	0.61	1.06	235.99	508.11
expensive	-0.41	1.14	-188.22	468.11
fast	0.99	1.00	439.12	383.65
full	-0.80	1.20	-317.10	516.68
hard	-0.75	0.78	-227.70	381.66
hardly full	-0.03	1.27	-28.43	556.44
irritating	-1.01	1.15	-392.19	409.18
packed	-0.81	1.13	-303.38	491.58
relaxing	1.26	0.96	476.70	418.67
restful	1.16	0.91	384.13	432.51
shaking	-0.91	0.91	-304.04	346.92
stressful	-1.15	1.09	-404.65	455.01
suffocating	-1.30	1.03	-583.78	432.53
tiring	-0.93	1.06	-304.28	441.29
uncomfortable	-1.18	1.02	-516.06	447.93
unsuitable with bags	-0.86	0.93	-328.65	391.46
unsuitable with strollers	-0.43	0.85	-128.90	363.26
without stress	1.21	0.96	436.15	408.85

Table 7: Means and standard deviations (SD) of the discrete ratings (sample size = 150) and continuous ratings (sample size = 127).

Name	Value	SD	Name	Value	SD	Name	Value	SD
α_1	0.0	fixed	η_1	0.149	0.0161	σ_1	-0.407	0.044
α_2	-0.0327*	0.141	η_2	0.0971	0.0233	σ_2	-0.109	0.0397
α_3	-0.336	0.191	η_3	0.108	0.029	σ_3	-0.104	0.0497
α_4	-1.4	0.208	η_4	0.278	0.0206	σ_4	-0.517	0.0916
α_5	-1.31	0.193	η_5	0.194	0.0263	σ_5	-0.35	0.0702
α_6	-1.69	0.226	η_6	0.272	0.0246	σ_6	-0.763	0.163
α_7	-0.454	0.159	η_7	0.204	0.0202	σ_7	-0.427	0.0652
α_8	-0.768	0.197	η_8	0.211	0.0244	σ_8	-0.199	0.0613
α_9	-1.41	0.248	η_9	0.264	0.027	σ_9	-0.408	0.102

Table 8: Parameter estimates of the measurement model of the HCM, with standard deviations (SD) (* Statistical significance $< 90\%$).

Note: The measurement equations (9) of the LVM are linear functions of the perception of comfort in PT:

$$I^*_{kin} = \alpha_{ki} + \eta_{ki} X^*_{in} + \sigma_{ki} \nu_{kin}, \qquad \text{with } \nu_{kin} \sim \mathcal{N}(0,1), \tag{17}$$

where α_{ki}, η_{ki} and σ_{ki} are parameters to estimate.