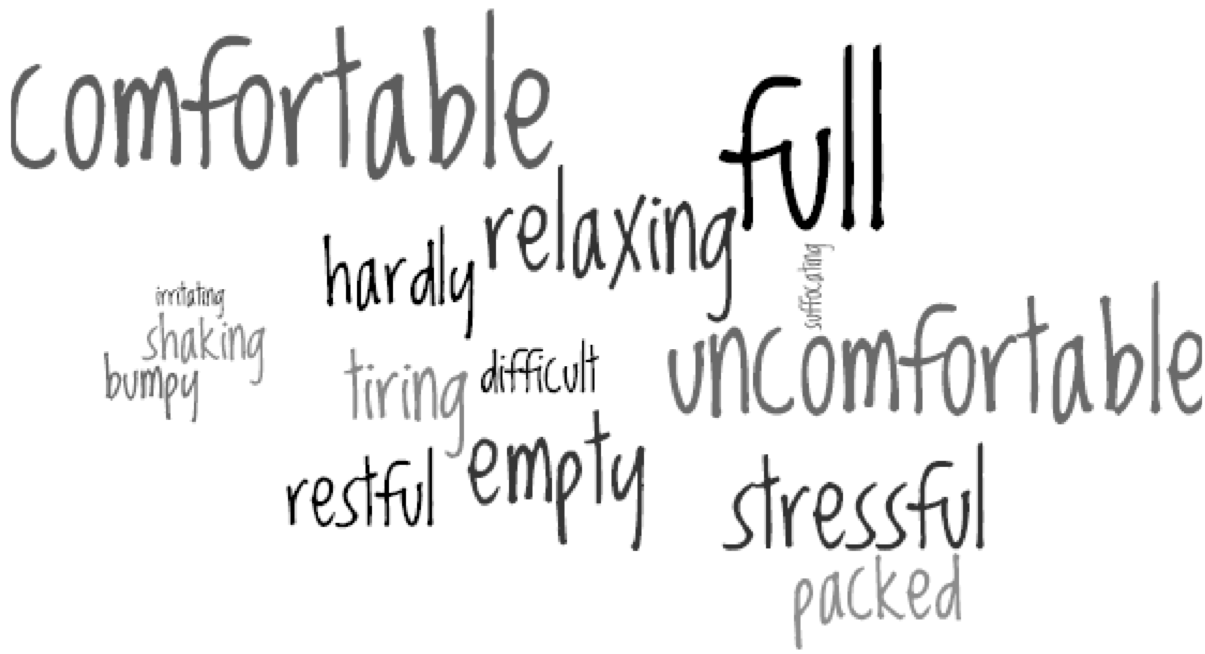

The use of word data to measure perception in hybrid choice models

Aurélie Glerum
Michel Bierlaire

Travel Behaviour Research: Current Foundations, Future Prospects

13th International Conference on Travel Behaviour Research
Toronto 15-20, July 2012

The use of word data to measure perception in hybrid choice models

Aurélie Glerum
ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)
School of Architecture, Civil and Environmental Engineering (ENAC)
Transport and Mobility Laboratory (TRANSP-OR)
Phone: + 41 21 693 24 35
E-mail: aurelie.glerum@epfl.ch

Prof. Michel Bierlaire
ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)
School of Architecture, Civil and Environmental Engineering (ENAC)
Transport and Mobility Laboratory (TRANSP-OR)
Phone: + 41 21 693 25 37
E-mail: michel.bierlaire@epfl.ch

Abstract
This research aims at developing a hybrid choice model (HCM) where a perceptional variable is measured by means of adjectives reported by individuals. Due to the qualitative nature of adjectives, the main challenges of the study involve their quantification and their integration into HCMs. In order to address these issues, we first obtain measures of the strength of the adjectives on the scale of the perceptional variable by using ratings from external evaluators. Second, an advanced measurement model of the perceptional variable is specified, in order to account for variations occurring in the answers from the evaluators.

Keywords
Discrete choice models, latent variables, perceptions, transportation mode choice, word data, hybrid choice models, structural equation models, measurement indicators.

Preferred Citation
Glerum, A. and M. Bierlaire (2012). The use of word data to measure perception in hybrid choice models, *Proceedings of the 13th International Conference on Travel Behavior Research (IATBR)*, July 15-20, 2012, Toronto, Canada.

1. Introduction

Hybrid choice models (HCM) (Walker, 2001; Ben-Akiva et al., 2002; Walker and Ben-Akiva, 2002) have been increasingly used in the recent years, due to their capability to integrate latent factors as explanatory variables of choice. These models have proved to be particularly relevant to analyze transportation decisions, such as the choice of a transportation mode. Some important psychological factors such as perceptions are indeed assumed to underlie transportation decisions and this effect can be assessed by the use of HCMs.

Due to their qualitative nature, perceptions are difficult to capture. So far, a common way to collect information on a perception was to design a list of related survey statements and ask respondents to rate them on a five-point Likert scale, ranging from a total disagreement to a total agreement (Likert, 1932; Thorndike, 1920; Bearden and Netemeyer, 1999). However this approach shows an important drawback: the statements designed to capture individuals' opinions reflect the survey designer's conception of an attitude and not the respondents' representation of it.

Social scientists recently developed powerful data collection techniques to gather information on individuals' perceptions (Kaufmann et al., 2001; Kaufmann et al., 2010). Indicators of such latent constructs consist of adjectives freely reported by respondents and describing, e.g. the perception of transportation mode alternatives.

In this paper we aim at developing an HCM that uses adjectives as measurements of a perceptional variable. The latter is assumed to impact on transportation mode choice.

The use of this new type of measurement indicators to capture a perceptional variable raises a number of issues. Precisely, adequate methods must be developed regarding (i) the quantification of such measurements and (ii) their integration into the HCM framework.

Computers have proved to be poor evaluators of qualitative concepts. Recent research in information technology emphasizes on the fact that individuals are more effective than computers for some specific tasks, such as evaluating a language (Franklin et al, 2011; Venetis et al., 2012). Since our aim is to quantify the adjectives measuring a perceptional construct, we ask external individuals (which we denote by 'evaluators') to rate the adjectives on a scale of the targeted perceptional variable. The aggregation of the evaluators' ratings thus allows to obtain a measure of the strength of each adjective.

The assessment of the impact of a latent variable on choice is possible through the integration of *structural equation models (SEM)* (Bollen, 1989) into a *discrete choice model (DCM)*, leading to the HCM framework[1]. Many studies have demonstrated the validity of such a framework (Espino et al., 2006, Abou-Zeid et al., 2010, Van Acker et al., 2011, Daziano and Bolduc, 2011, Atasoy et al., forthcoming). With qualitative indicators, the measurement component of the *latent variable model (LVM)* of an HCM deserves a particular attention. Precisely, the subjectivity due to ratings of an adjective by different evaluators must be handled correctly in order to obtain a reliable measure of the perception. A model which

[1] The *latent variable model (LVM)* component of an HCM consists of a SEM.

quantifies the adjectives is specified as a part of the measurement component of the LVM. It precisely consists of an additional LVM.

The methodology developed in this research is applied on a case study. The data we use result from a travel diary survey performed in low-density areas of Switzerland, which aimed at understanding the inhabitants' transportation mode choices. In particular, respondents are asked to report adjectives characterizing the transportation modes. From the adjectives, perceptional variables are identified.

The paper is structured as follows. Section 2 presents the two data sets used in this study, that is, the data from the travel diary survey and the data resulting from the evaluation of the adjectives. Section 3 presents the modeling framework integrating (1) a choice model, (2) a LVM characterizing the perceptional variable and (3) a LVM quantifying the adjectives. It finally shows how the components are integrated. Section 4 presents an example of specification and estimation of the LVM quantifying the adjectives. Section 5 concludes and presents the future steps of this research.

2. The data

The model presented in this paper is based on data from two different surveys. The latter are described in this section.

2.1. Revealed preferences survey

This research is based on a case study which aims at analyzing the transportation mode preferences of individuals living in low-density areas of Switzerland. 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 by inhabitants of suburban regions in one day was collected, including characteristics such as chosen transportation mode, price and duration of the trip.

Based on the assumption that individuals also consider more qualitative factors such as *perceptions* in their choice of daily transportation mode, the survey also included a question designed to collect information on respondents' perceptions of different transportation modes.

Table 1 reports the survey question described above. For each transportation mode a respondent had to give three adjectives that characterize it best in his opinion.

With the help of social scientists, the collected adjectives were then classified into eleven themes, that is, comfort, perception of time, perception of cost, difficulty of access, flexibility, efficiency, reliability, environmental impact, appreciation, feeling and look. Each of these themes reflects a different perception which can potentially affect the choice of transportation mode of an individual (Bierlaire et al., 2011).

For the purpose of this study we focus on the *perception of comfort in public transportation (PT)* only. We hence consider adjectives that are (i) related to comfort only and (ii) reported in rows 2, 3 and 4 of Table 1.

Table 1: The survey question designed to evaluate individuals' perceptions of the different transportation modes.

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:			

The theme of comfort was selected due to an important range of related adjectives. In this paper we moreover assess the effect on choice of the perception of comfort in PT only for simplification. But it is planned to investigate the other themes in future research, as well as their effect on other modes.

In order to evaluate the effect of the perception of comfort in PT on transportation mode choice, it is necessary to assign a value to each adjective on a scale of comfort. To do so, an additional survey is performed.

2.2. Adjective quantification survey

A second survey was conducted to quantify the adjectives reported in the RP survey. 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, part of the evaluators are asked to give a rating on a *discrete scale*, ranging from -2 to 2, while the remaining evaluators are required to give a rating on a *continuous scale*, ranging from -1000 to 1000.

Figure 1 and Figure 2 show histograms of the ratings of the evaluators for all adjectives associated to comfort, for the discrete and continuous scales, respectively.

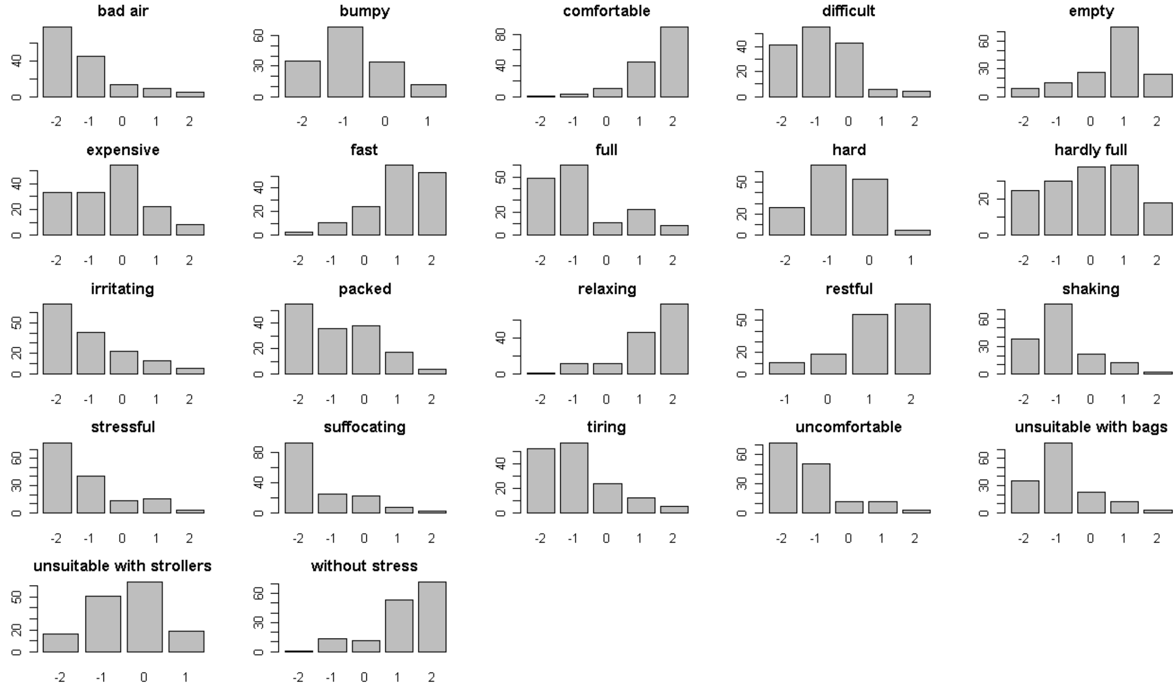

Figure 1: Histograms of the discrete ratings of the evaluators for each adjective.

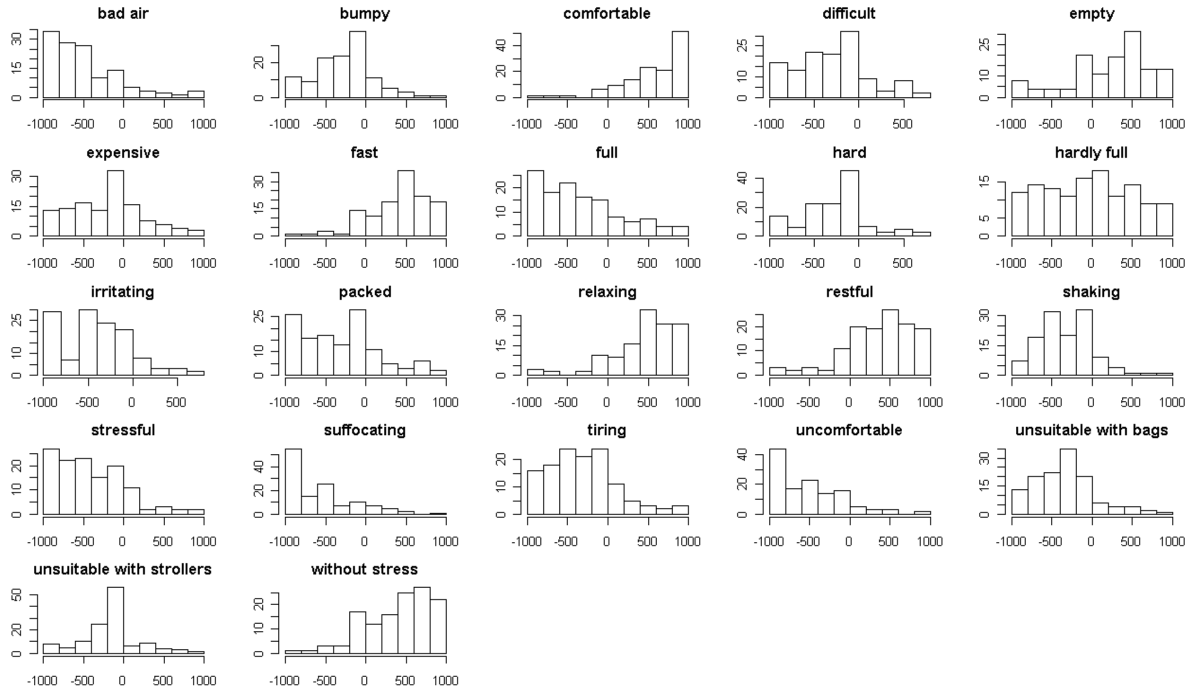

Figure 2: Histograms of the continuous ratings of the evaluators for each adjective.

3. The integrated model framework

The aim of this research is to develop an HCM which can integrate latent variables measured by qualitative indicators. Precisely the indicators are adjectives, whose numerical values are inferred by a quantification model.

The HCM framework involves three components:

- A discrete choice model
- A latent variable model
- A quantification model of the measurements of the latent variable

In this section we first present the individual components of the HCM and then their integration.

3.1. 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 characteristics X_{in} of i and n and of a latent attribute X_n^*:

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

where β is a list of parameters to estimate.

3.2. Latent variable model

The latent variable X_n^* cannot be directly observed and must be indirectly measured by means of indicators. Therefore a *latent variable model (LVM)* relates X_n^* to a list of K measurement indicators I_{kn}^*, with $k = 1, \dots, K$, and expresses the X_n^* as a function of observed attributes X_n of individual n.

In this research we make use of qualitative indicators of the perceptional variable X_n^*. These indicators consist of adjectives characterizing the perceptional variable, which are reported freely by survey respondents (see Section 2.2 for more details). We assume that the kth adjective reported by individual n has a unobservable score I_{kn}^*, which represents an *indirect measurement* of the perceptional variable X_n^*.

The measurement component of the LVM is described by a set of functions r_k, relating the latent variable X_n^* to its indicators I_{kn}^*:

$$I_{kn}^* = r_k(X_n^*; \eta_k) + \upsilon_{kn}, \text{ with } \upsilon_{kn} \sim \mathcal{N}(0, \sigma_k), \tag{2}$$

where η_k and σ_k are parameters to estimate, for $k = 1, \dots, K$.

The structural component of the LVM is defined by a function h, relating socio-economic information X_n of respondent n to the latent variable X_n^*:

$$X_n^* = h(X_n; \mu) + \omega_n, \text{ with } \omega_n \sim \mathcal{N}(0, \sigma_\omega), \tag{3}$$

where μ and σ_ω are parameters to estimate.

3.3. Quantification model

In this section we show how the data from the adjective quantification survey (see Section 2.2) can be used to obtain a measure of I_{kn}^*, by developing a model which can be applied to assign a numerical score to each adjective.

First, let us define J_{ln}^* as the score of adjective l indirectly assigned by respondent n. The kth indirect measurement I_{kn}^* of latent variable X_n^* is related to J_{ln}^* by the following expression:

$$I_{kn}^* = \sum_{l=1}^{L} J_{ln}^* \cdot H_{kln}, \tag{4}$$

where L is the number of adjectives and

$$H_{kln} = \begin{cases} 1, \text{ if } l \text{ is selected by } n \text{ for indicator } k \\ 0, \text{ otherwise} \end{cases}. \tag{5}$$

Now, the *quantification model* of the adjectives consists of a second LVM. It comprises a structural equation and two types of measurement equation, that is, a measurement equation based on a discrete scale and a measurement equation based on the continuous scale (see Section 2.2).

The structural equation relates the score J_{lm}^* given to adjective l by evaluator m to his socio-economic information X_m as follows:

$$J_{lm}^* = c_l + \delta_\gamma, \text{ with } \delta_\gamma \sim \mathcal{N}(g(X_m; \gamma), \sigma_\gamma), \tag{6}$$

where γ and σ_γ are parameters to estimate and g is a function of X_m and γ. Score J_{lm}^* is the sum of a constant c_l to estimate and an error term δ_γ, capturing the evaluator-specific bias. Socio-economic information X_m of an evaluator m, such as the education level are indeed assumed to have an effect on the indirect measure J_{lm}^* of an adjective j.

Let us define J_{lm}^{D} as the discrete measurement of score J_{lm}^{*}. The discrete measurement equation is given as follows:

$$\tilde{J}_{lm}^{D} = \lambda_D \cdot J_{lm}^{*} + \delta_D, \text{ with } \delta_D \sim Logistic(0,1), \quad (7)$$

where λ_D is a parameter to estimate. In Equation (7), $\tilde{J}_{lm}^{D}$ represents the latent continuous response variable underlying J_{lm}^{D} (Agresti, 2002, Abou Zeid, 2009). The observed discrete measurement J_{lm}^{D} is related to the latent continuous variable $\tilde{J}_{lm}^{D}$ by the following formula:

$$J_{lm}^{D} = \begin{cases} \tilde{J}_{lm}^{D} \leq \tau_1 \\ \tau_1 < \tilde{J}_{lm}^{D} \leq \tau_2 \\ \tau_2 < \tilde{J}_{lm}^{D} \leq \tau_3, \\ \tau_3 < \tilde{J}_{lm}^{D} \leq \tau_4 \\ \tau_4 < \tilde{J}_{lm}^{D} \end{cases} \quad (8)$$

where τ_1, τ_2, τ_3 and τ_4 are thresholds to estimate.

The continuous measurement J_{lm}^{C} of score J_{lm}^{*} is expressed by the following measurement equation:

$$J_{lm}^{C} = \alpha_C + \lambda_C \cdot J_{lm}^{*} + \delta_C, \text{ with } \delta_C \sim \mathcal{N}(0, \sigma_C), \quad (9)$$

where α_C, λ_C and σ_C are parameters to estimate.

3.3.1. Estimation

The parameters of the quantification model are estimated using maximum likelihood techniques. The likelihood function $\mathcal{L}_l$ for adjective l is given by the following formula:

$$\mathcal{L}_l = \prod_{m=1}^{M} \int_{J_{lm}^{*}} f(J_{lm}^{C} | J_{lm}^{*}; \alpha_C, \lambda_C, \sigma_C) f(J_{lm}^{D} | J_{lm}^{*}; \lambda_D, \tau_1, \tau_2, \tau_3, \tau_4) f(J_{lm}^{*} | X_m; \gamma, \sigma_\gamma) \, dJ_{lm}^{*}, \quad (10)$$

where M is the total number of evaluators, $f(J_{lm}^{C} | J_{lm}^{*}; \alpha_C, \lambda_C, \sigma_C)$, $f(J_{lm}^{D} | J_{lm}^{*}; \lambda_D, \tau_1, \tau_2, \tau_3, \tau_4)$ and $f(J_{lm}^{*} | X_m; \gamma, \sigma_\gamma)$ are the density functions of J_{lm}^{C}, J_{lm}^{D} and J_{lm}^{*}, respectively.

3.3.2. Application of the model

Using the parameter values resulting from the estimation described in Section 3.3.1, the quantification model can then be applied to infer a value to J_{lm}^* for each adjective l and each evaluator m. Let us denote this fitted value by $\widehat{J_{lm}^*}$. It is obtained by taking the deterministic part of Equation (6):

$$\widehat{J_{lm}^*} = c_l, \ \forall m \tag{11}$$

In addition, the same model can be applied to infer the value of each adjective l reported by survey respondent n for indicator k. Indeed we have:

$$\widehat{I_{kn}^*} = \sum_{l=1}^{L} \widehat{J_{ln}^*} \cdot H_{kln}, \tag{12}$$

3.4. Integrated model

Using the inferred scores $\widehat{I_{kn}^*}$ for the K measurement indicators and the N respondents of the RP survey, the HCM described in Sections 3.1 and 3.2 can be estimated by maximum likelihood techniques. The following likelihood function is considered:

$$\mathcal{L} = \prod_{n=1}^{N} \int_{X_n^*} \prod_{i=1}^{I} P(y_{in}|X_{in}, X_n^*; \beta)^{y_{in}} \cdot f(X_n^*|X_n; \mu, \sigma_\omega) \cdot \prod_{k=1}^{K} f(\widehat{I_{kn}^*}|X_n^*; \eta_k, \sigma_k)\, dX_n^* \quad , \tag{13}$$

where I is the number of choice alternatives, $P(y_{in}|X_{in}, X_n^*; \beta)$ is the probability that respondent n chooses alternative i, y_{in} is an indicator of the actual choice of n, $f(X_n^*|X_n; \mu, \sigma_\omega)$ is the density function of X_n^* and $f(\widehat{I_{kn}^*}|X_n^*; \eta_k, \sigma_k)$ is the density function of the kth indicators, for $k = 1, \dots, K$.

Figure 3 shows a diagram of the integrated model, 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. 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.

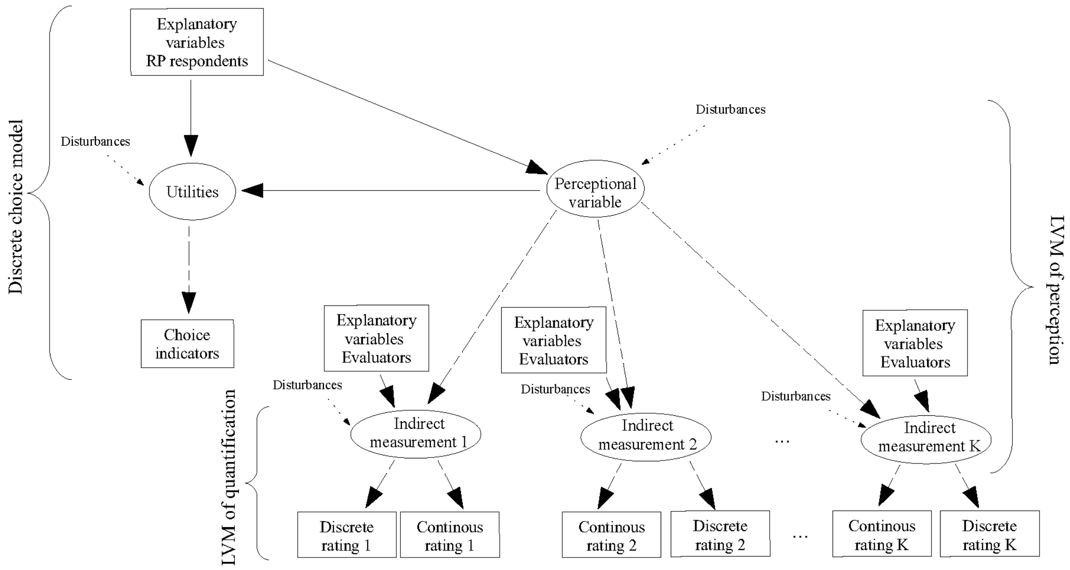

Figure 3: Diagram of the integrated model framework, based on Walker and Ben-Akiva (2002).

4. Application example of the quantification model

In this section, we present an example of specification and estimation of the quantification model for five adjectives out of the twenty-two adjectives appearing in Figure 1 and Figure 2, that is, ‘bad air’, ‘comfortable’, ‘difficult’, ‘empty’ and ‘full’.

4.1. Specification

Exploratory analysis on the data from the adjective quantification survey showed that the higher the level of education the evaluator has, the higher its rating of an adjective will be in absolute value. We hence introduce the education level $Educ_m$ an explanatory variable of the unobserved score J^*_{lm} of evaluator m for adjective l. The structural equations relative to quantification model are specified as follows:

$$\begin{aligned} J^*_{lm} &= c_l + \delta_\gamma \\ &= c_l + \gamma_{Educ,l} \cdot Educ_m + \sigma_\gamma \cdot \delta_\gamma', \text{ with } \delta_\gamma' \sim \mathcal{N}(0,1) \end{aligned} \tag{14}$$

where $\gamma_{Educ,l}$ is an adjective-specific parameter capturing the bias occurring on the score of an adjective due to the different education levels of the evaluators.

The discrete measurement equations of the quantification model are specified as follows:

$$\tilde{J}^D_{lm} = \lambda_{Dl} \cdot J^*_{lm} + \delta_{Dl}, \text{ with } \delta_{Dl} \sim Logistic(0,1), \tag{15}$$

where λ_{Dl} are adjective-specific parameters and $\tilde{J}^D_{lm}$ is the latent continuous variable related to the observed discrete measurement J^D_{lm} by the following relation:

$$J^D_{lm} = \begin{cases} \tilde{J}^D_{lm} \leq \tau_{1l} \\ \tau_{1l} < \tilde{J}^D_{lm} \leq \tau_{2l} \\ \tau_{2l} < \tilde{J}^D_{lm} \leq \tau_{3l}, \\ \tau_{3l} < \tilde{J}^D_{lm} \leq \tau_{4l} \\ \tau_{4l} < \tilde{J}^D_{lm} \end{cases} \tag{16}$$

where τ_{1l}, τ_{2l}, τ_{3l} and τ_{4l} are adjective-specific thresholds to estimate.

The continuous measurement equations are specified as follows:

$$J^C_{lm} = \alpha_{Cl} + \lambda_{Cl} \cdot J^*_{lm} + \delta_{Cl}, \text{ with } \delta_{Cl} \sim \mathcal{N}(0, \sigma_{Cl}), \tag{17}$$

where α_{Cl}, λ_{Cl} and σ_{Cl} are adjective-specific parameters to estimate.

4.2. Estimation results

The parameters of the quantification model can be estimated by maximizing Formula (10). The estimation results are shown in Table 2.

Table 2: Parameter estimates of the quantification model, with values of *t*-test.

Name	Value	*t*-test	Name	Value	*t*-test
$c_{bad\ air}$	-1.79	-6.32	$w_{1,comfortable}$	0.753	3.27
$c_{comfortable}$	1.84	4.93	$w_{2,comfortable}$	2.2	2.42
$c_{difficult}$	-0.513	-2.11	$w_{3,comfortable}$	2.85	8.02
c_{empty}	0.53	2.33	$\alpha_{C,difficult}$	1.46	6.22
c_{full}	-0.0199	-0.09	$\sigma_{C,difficult}$	-2.66	-26.08
$\gamma_{Educ,bad\ air}$	-0.0679	-3.35	$w_{1,difficult}$	1.17	6.89
$\gamma_{Educ,comfortable}$	0.233	15.98	$w_{2,difficult}$	3.08	11.71
$\gamma_{Educ,difficult}$	-0.215	-20.69	$w_{3,difficult}$	2.19	4.97
$\gamma_{Educ,empty}$	0.136	9.31	$\alpha_{C,empty}$	-0.955	-4.6
$\gamma_{Educ,full}$	-0.22	-21.42	$\sigma_{C,empty}$	-2.69	-23.76
$\alpha_{C,bad\ air}$	1.31	5.14	$w_{1,empty}$	0.568	5.49
$\sigma_{C,bad\ air}$	-2.67	-25.93	$w_{2,empty}$	1.78	5.69
$w_{1,bad\ air}$	0.426	3.75	$w_{3,empty}$	3.38	12.02
$w_{2,bad\ air}$	2.14	8.73	$\alpha_{C,full}$	0.742	3.43
$w_{3,bad\ air}$	1.6	4	$\sigma_{C,full}$	-2.67	-23.47
$\alpha_{C,comfortable}$	-2.03	-5.6	$w_{1,full}$	0.243	3.45
$\sigma_{C,comfortable}$	-2.82	-36.25	$w_{2,full}$	2.32	9.94
			$w_{3,full}$	1.91	5.47

We can draw the following conclusions from the estimation results:

- The estimation results show that all parameters are significant, except constant c_{full}.
- All constants c_l 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.
- The level of education significantly affects the score reported for each adjective. Moreover the parameters $\gamma_{Educ,l}$ have the expected sign. Exploratory analysis had indeed shown that the higher the level of education one evaluator has, the higher the ratings he assigns to the adjectives are, in absolute value. For example, $\gamma_{Educ,difficult}$ is significantly different from 0 and has a negative sign, which is consistent with the fact that adjective ‘difficult’ is intuitively associated with discomfort.

We remark that some parameters of the model are normalized for identification purposes. First, parameters λ_{Cl} and σ_{Cl} are fixed to 1. Second, variable $\tilde{J}^{D}_{lm}$ is centered to 0. Instead of estimating parameters τ_{1l}, τ_{2l}, τ_{3l} and τ_{4l}, we define and estimate variables w_{1l}, w_{2l} and w_{3l}, such that

$$\tau_{1l} = -w_{2l} \qquad (18)$$

$$\tau_{2l} = -w_{1l} \tag{19}$$
$$\tau_{3l} = w_{1l} \tag{20}$$
$$\tau_{4l} = w_{3l}. \tag{21}$$

The final value of the loglikelihood of the quantification model is $\mathcal{L} = -1747$. Considering a null model where only parameters σ_{Cl}, τ_{1l}, τ_{2l}, τ_{3l} and τ_{4l} are estimated, we can compute the index of fit $\bar{\rho}^2$ as follows:

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}-Q}{\mathcal{L}_0}, \tag{22}$$

where $\mathcal{L}_0 = -2170$ is the loglikelihood for the null model and $Q = 35$ is the number of parameters in the quantification model. We obtain an index of fit of $\bar{\rho}^2 = 0.18$.

5. Conclusion

In this paper we presented the framework of an integrated model which allows for capturing the impact of perceptions on choice, when qualitative indicators in the form of adjectives are used as measurements. In particular, a model to quantify the adjectives is developed and tested on a real case study. The results are consistent with expectations, which demonstrates the validity of a data collection procedure and modeling approach based on evaluators.

This research aims at highlighting the fact that a different type of indicators can be used as measurements of a perceptional variable. Though classical indicators resulting from agreements to opinion statements are easy to code and integrate into an LVM, qualitative indicators might reflect perceptions in a more realistic way, since the adjectives are freely reported by the respondents.

The next steps in the research involve the estimation of the quantification model for all adjectives related to comfort and its integration into the HCM framework, in order to analyze the effect of the perception of comfort in public transportation on mode choice.

Future works also include the development of LVMs for the other perceptional constructs which were identified from the reported adjectives and their introduction into the HCM framework.

Acknowledgements

The authors would like to thank PostBus for their advice and feedback on the collaborative project which was set up and for funding it. They also would like to acknowledge the Urban Sociology Laboratory (LaSUR) and the Urban and Regional Planning Laboratory (CEAT) from EPFL who co-conducted the RP survey used as a case study for this research. They are also grateful to Alberto Monticone, who established the classification of the adjectives, and Peng Cui, who performed preliminary work on the data from the adjective quantification survey.

References

Abou Zeid, M. (2009). Measuring and modeling activity and travel well-being, Ph.D. Thesis, Massachusetts Institute of Technology.

Abou-Zeid, M., M. Ben-Akiva, M. Bierlaire, C. Choudhury and S. Hess (2010). Attitudes and value of time heterogeneity, paper presented at *Applied Transport Economics: A Management and Policy Perspective*, 523–545.

Agresti, A. (2002). *Categorical Data Analysis*, Wiley Series in Probability and Statistics,Wiley-Interscience.

Atasoy, B., A. Glerum and M. Bierlaire (forthcoming). Attitudes towards mode choice in Switzerland, *disP - The Planning Review*. Accepted for publication.

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

Bearden, W. and R. Netemeyer (1999). *Handbook of marketing scales: multi-item measures for marketing and consumer behavior research*, Association for Consumer Research, Sage Publications.

Bierlaire, M., A. Curchod, A. Danalet, E. Doyen, P. Faure, A. Glerum, V. Kaufmann, K. Tabaka and M. Schuler (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, ENAC, EPFL.

Bollen, K. A. (1989). *Structural equations with latent variables*, Wiley.

Daziano, R. A. and D. Bolduc (2011). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model, *Transportmetrica*, 1–33.

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

Franklin, M. J., D. Kossmann, T. Kraska, S. Ramesh and R. Xin (2011), CrowdDB: Answering Queries with Crowdsourcing, *Proceedings of ACM SIGMOD Conference.*, June 12-16, 2011, Athens, Greece.

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

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

Likert, R. (1932). A technique for the measurement of attitudes, *Archives of Psychology* **22**(140):1–55.

Thorndike, E. L. (1920). A constant error in psychological ratings, *Journal of Applied Psychology* **4**(1): 25–29.

Van Acker, V., P. L. Mokhtarian and F. Witlox (2011). Going soft: on how subjective variables explain modal choices for leisure travel, *European Journal of Transport and Infrastructure Research*, **11**, 115–146.

Venetis, P., H. Garcia-Molina, K. Huang and N. Polyzotis (2012). Max Algorithms in Crowdsourcing Environments, *Proceedings of WWW*, April 16-20, 2012, Lyon, France.

Walker, J. and M. Ben-Akiva (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.