ESTIMATION OF MODE CHOICE MODELS WITH LATENT CLASSES AND PSYCHOMETRIC INDICATORS

Ricardo Hurtubia, Universidad de Chile, rhurtubia@fau.uchile.cl
My Hang Nguyen, CITEC Ingénieurs Conseils S.A.
Aurélie Glerum, Ecole Polytechnique Fédérale de Lausanne
Michel Bierlaire, Ecole Polytechnique Fédérale de Lausanne

RESUMEN

Este artículo propone un método para introducir indicadores psicométricos en la especificación de modelos de elección discreta con clases latentes, a través de ecuaciones de tipo ordinal que relacionan las respuestas a los indicadores con atributos del tomador de decisiones. El método es implementado para un caso de estudio de elección de modo de transporte y comparado con métodos alternativos para la estimación de modelos de clases latentes incorporando indicadores. El método propuesto genera valores significativamente diferentes para los parámetros estimados y permite una compresión más detallada del comportamiento de cada clase latente.

Palabras clave: clases latentes, indicadores psicométricos, elección modal

ABSTRACT

This paper proposes a method to introduce psychometric indicators in the specification of integrated latent class and discrete choice models, through the definition of ordinal measurement equations that relate the indicators to attributes of the decision maker. The method is implemented for a mode-choice case study and compared with alternative methods to introduce indicators. Results show that the proposed method generates significantly different estimates for the class and choice models and provide additional insight into the behavior of each class.

Keywords: latent class, psychometrics, mode choice

1 INTRODUCTION

Discrete choice models usually consider quantitative attributes as the principal variables that explain choice behavior (Ben-Akiva and Lerman; 1985). However, there are more complex, unobserved factors that may have a relevant effect in the way choices in general are made. Some of these latent factors are the decision maker's lifestyle, personal attitudes or perceptions (McFadden; 1986). For example, in the particular case of travel demand modeling, a user's lifestyle can make him follow an activity pattern that triggers particular travel needs that affect mode choice. Likewise, users may have a bias against a particular mode that is explained by his own perception of the safety or flexibility of this mode. In a similar way, a user might prefer to use the car because he perceives it as an indicator of social status, or avoid the use of public transportation because he associates it to a lower socioeconomic status. Though these subjective features have an obvious effect on choice, it is not straightforward to characterize and integrate them into choice models. We address that aspect in this article.

The introduction of latent factors into discrete choice models has been treated under two main approaches: latent variable models (LVM) and latent class models (LCM). The latent variable approach deals with the explicit modeling of unobserved psychological characteristics of the decision maker, such as attitudes and perceptions. The latent class approach assumes that the population can be segmented into discrete groups that have different preferences or perceptions and, therefore, have different choice behaviors.

Psychometric indicators are additional information that can be used to specify and estimate latent constructs. They usually reflect the preferences of decision makers on topics that are (closely or not so) related to the choice that is being analyzed/modeled. Examples of psychometric indicators range from the answers to questions about the level of agreement with a statement or the "grade" that is given to the quality of a service or object (Likert; 1932), to the set of adjectives that individuals use to characterize the topic in question (Glerum and Bierlaire; 2012).

Although the use of indicators should clearly help to estimate better latent class models, its use has been mostly developed and applied in the latent variable approach (Hess et al.; 2011). However, the LCM approach has characteristics that make it, in some cases, preferable over other methods to capture heterogeneity (Greene and Hensher; 2003; Shen; 2009) and, if possible, should be improved with the integration of psychometric data. The use of latent classes in discrete choice models also allows for the identification of market segments that can be used to design more effective (segment-specific) policies or marketing strategies.

This paper proposes a method to introduce psychometric indicators in the specification of discrete choice models with latent classes. The method uses ordinal logit models as measurement relationships between the observed answers and the "utility" a respondent of a particular class will perceive for providing each of these answers. The novel feature of this method consists of specifying the measurement relationships as class-specific structural relations between the aforementioned utility and the attributes of the decision maker/respondent. The structure of the proposed model is inspired by the Generalized Random Utility Model (Walker and Ben-Akiva; 2002). The method is applied on two datasets for transport mode choice but should be easily implemented in other choice contexts.

The paper is organized as follows. Section 2 reviews the use of latent class models in discrete choice models and the importance of psychometric indicators to characterize the classes. Section 3 presents the modeling approach adopted in this research and designed to provide a better specification of such models. Section 4 provides an application of the methodology on a mode choice case study for low-density areas in Switzerland. Section 5 concludes on the advantages of the proposed modeling approach.

2 LATENT CLASS MODELS IN DISCRETE CHOICE ANALYSIS

Widely used in social sciences for quantitative analysis (Lazarsfeld and Henry; 1968), latent class models were not proposed in the form of class-membership probabilities within choice models until the work of Kamakura and Russell (1989). Class-membership models explain the probability of an individual belonging to a consumer segment as a function of the consumer's characteristics; they are a powerful instrument because they allow to relate attributes of the decision maker with unobserved behavioral classes and, therefore, simplify the market segmentation process.

There is evidence in the literature suggesting that latent class models are a very convenient, flexible and intuitive way to introduce taste heterogeneity in discrete choice models. For example, Bhat (1997) applied the latent class approach to the transport mode choice problem finding that the endogenous segmentation into classes allows for better data fit and more intuitive results compared to other approaches used to capture heterogeneity. Greene and Hensher (2003), Shen (2009) and Hess et al. (2011) analyzed the LCM approach, comparing it with alternative methodologies like the Mixed Logit Model (McFadden and Train; 2000) concluding that, while both offer a good way to capture unobserved heterogeneity, experimental results suggest that the latent class approach allows for a better behavioral interpretation of the results. Keane and Wasi (2012) compared the latent class approach with several other models that account for taste heterogeneity. Although not being the best in terms of model fit, the LCM approach was identified as the one allowing the most intuitive understanding of the patterns of heterogeneity.

Several application of integrated choice and latent class models can be found in the transport and land use-related literature. For example, the aforementioned works by Bhat (1997) and Shen (2009), applied the LCM approach to the choice of transport mode while Greene and Hensher (2003) did it for route choice. In the area of land use, Walker and Li (2007) identified different lifestyle classes among the population of a city, showing that the segments are key determinants in the choice of residential location. Zhang et al. (2009) used a latent class structure to model different intra-household choice mechanisms regarding car ownership, finding different behaviors between classes of households that would have not been identified without the segmentation. Wen and Lai (2010) used the latent class approach in the airline choice problem, identifying significantly different willingness to pay across consumer segments. Similar result were obtained by Wen et al. (2012) but in the context of the choice of mode to access stations of a high-speed train. Koutsopoulos and Farah (2012) used latent classes to identify and model different patterns (or regimes) of driving behavior for a microscopic traffic simulator.

2.1 Psychometric indicators

Psychometric indicators can be used improve the specification and estimation process of latent constructs (like classes) because they are a measurable manifestation of the effect of unobserved attributes in the preferences of individuals (Walker and Ben-Akiva; 2002). Moreover, the use of indicators in discrete choice models may help to identify latent classes that are not captured or described by the choice data alone (Ben-Akiva, McFadden, Train, Walker, Bhat, Bierlaire, Bolduc, Börsch-Supan, Brownstone, Bunch et al.; 2002). Despite this, most methodological developments are focused on the use of indicators to estimate choice models using a LVM approach (Ben-Akiva, Walker, Bernardino, Gopinath, Morikawa and Polydoropoulou; 2002), with few examples applied under the LCM approach.

Ben-Akiva and Boccara (1995) introduced the use of indicators to the estimation of models with a latent choice set by measuring the user's perceived availability of an alternative and modeling a linear relationship between this indicator, the modeled availability and the "desirability" (a proxy of the utility) of each alternative. They find that using this approach generates better predictive results than a standard logit model and that the use of indicators allows for more robust estimates.

Gopinath (1995) postulated the existence of two classes of shippers in the maritime freight choice context and used indicators to measure the latent attitude of the shippers towards different freight services attribute. The latent attitudinal variables was then used as an explanatory variable of the class-membership model. Similarly, Hosoda (1999) estimated mode choice models for shopping trips with latent classes that are functions of continuous latent variables like the "level of consciousness" of the traveler. In these last two cases, indicators are indirectly related to the class-membership model because they are first used to measure attitudinal latent variables which are then used as explanatory variables in class-membership models.

In the context of tourism destination choice, Boxall and Adamowicz (2002) modeled natural park choice in Central Canada and used psychometric indicators related to motivations for taking a trip, identifying four groups of travelers. The likelihood of membership in these classes was influenced by socioeconomic characteristics but also latent motivations which were related to the indicators. Thus, the group membership is a direct function of how the decision-makers respond to the questions and the model could not be used for predicting demand.

In a similar context, Morey et al. (2006) developed a fishing-location choice model based on three classes of fishermen that were identified using attitudinal data about the characteristics of a particular fishing location. However, the response probabilities to the psychometric indicators are estimated as single, class-specific parameters and are not structurally related to attributes of the decision makers. A similar approach is proposed by Collins and Lanza (2010) in the context of social and health sciences and by Atasoy et al. (2013) in the context of transport mode choice.

In the surveyed literature, the class-membership probabilities are not directly related to indicators through measurement relationships that take into account the attributes of the decision makers. This paper proposes a method to do so, through the use of ordinal models. The specification of the class-specific measurement relationships leads to a better characterization of the classes since it integrate psychometric information. Moreover it allows to interpret the responses to psychometric indicators behaviorally.

3 METHODOLOGY

In this section we first present the general framework of latent class models. In a second stage, we introduce the use of psychometric indicators to help identify the classes.

3.1 Latent class model

Latent class models assume that discrete segments of the population have different choice behaviors, explained by different perceptions of the attributes of the alternatives, different taste parameters or different decision protocols. These differences can often be linked to the lifestyle, attitudes and even political or ideological views of the decision maker. In the context of discrete choice analysis, this translates into a class-specific utility function of choosing alternative i by decision maker n:

$$U_{in}^{s} = V(X_{in}, X_n, \beta^s) + \varepsilon_{in}^{s} \tag{1}$$

where $V(X_{in}, X_n, \beta^s)$ is the deterministic part of the utility function, X_{in} is a vector of attributes of the alternative i for decision maker n, X_n is a vector of characteristics of n and β^s is a vector of parameters (to be estimated) that is specific to class s. The term ε_{in}^{s} is a random component accounting for unobserved attributes and characteristics that can also be class-specific. Assuming an i.i.d. Extreme Value distribution for the random component, we can write the probability of an individual n choosing alternative i, conditional on the class s to which he belongs, as a logit:

$$P_n(i|s) = \frac{e^{V(X_{in}, X_n, \beta^s)}}{\sum_{j \in C_s} e^{V(X_{jn}, X_n, \beta^s)}} \tag{2}$$

where C_s is the set of alternatives considered by individuals belonging to class s.

Since classes are latent or unobserved, it is not possible to deterministically relate an individual to a class. It is possible however to assume that the membership to a class depends on the characteristics of the decision maker and that this relation is described by a *class-membership functionf*, such that

$$F_{ns} = f(X_n, \gamma^s) + \xi_{ns}, \tag{3}$$

where F_{ns} is a latent continuous variable that is related to the probability of belonging to class s and can be perceived as the "utility" to belong to one class, and γ^s is a vector of parameters to be estimated. Assuming that ξ_{ns} are i.i.d. $EV(0,1)$, the probability for an individual n to belong to a particular class s is given by:

$$P_n(s) = \frac{e^{f(X_n, \gamma^s)}}{\sum_{r \in S} e^{f(X_n, \gamma^r)}} \tag{4}$$

where S is the set of classes.

Using (2) and (4) we can write the complete probability of individual n choosing an alternative i as the following expression:

$$P_n(i) = \sum_{s \in S} P_n(i|s) P_n(s). \tag{5}$$

The vector of parameters β^s and γ^s of the utility functions in equations (2) and (4), with $s \in S$, can be estimated by maximum likelihood.

3.2 Latent class model with psychometric indicators

Psychometric indicators can be introduced by assuming that the probability of giving a agreement level I_k to the kth question/indicator, with $k = 1, \ldots, K$ will depend on the class of the respondent. This allows to write the joint probability of choosing i and answering I_k for individual n as:

$$P_n(i, I_k) = \sum_{s \in S} P_n(i|s) P_n(s) \prod_{k=1}^{K} P_n(I_k|s) \tag{6}$$

where $P_n(I_k|s)$ is the probability of answering I_k to the kth indicator if the respondent n belongs to class s. As mentioned in Section 2.1, this probability is usually estimated directly as a single parameter or a constant (Morey et al.; 2006; Collins and Lanza; 2010; Atasoy et al.; 2013).

We propose to model the response probability $P_n(I_k|s)$ as a function of the attributes of the respondent (or decision maker), conditional on the class. For this we consider a continuous latent construct that varies with both the characteristics and the class of the respondent, and we derive an ordered logit model from it. Our hypothesis is that, by doing so, we enhance the characterization of the class-membership model.

We focus on the case where indicators take the form of questions where an ordered response is provided. A typical example of this is when the respondent is asked about his level of agreement to a certain statement, where such level of agreement is classified in a Likert scale (Likert; 1932). The response probability must be modeled as a function of the characteristics of the decision maker only. It is convenient to do so using an ordinal logit approach, since the responses to the indicators consist of a few integer values. We define the *item response function* g relative to the answer of individual n to indicator I_k as:

$$G^s_{I_k,n} = g(X_{kn}; \alpha^s_k) + \nu^s_{kn} \tag{7}$$

where α^s_k is a indicator- and class-specific vector of parameters to be estimated, $\nu^s_{kn} \sim \text{Logistic}(0,1)$ is a disturbance term and $G^s_{I_k,n}$ is a latent continuous variable that can be seen as the “utility” of giving a particular level of agreement to indicator I_k. We assume that the value of the unobservable variable $G^s_{I_k,n}$ increases with the level of agreement ℓ to indicator k and that the probability of answering ℓ comes defined by:

$$P_n(I_k = \ell|s) = P(\tau^s_{\ell-1} < G^s_{I_k=\ell,n} \leq \tau^s_\ell) \tag{8}$$

where $\ell = 1, \ldots, L$ is the level of agreement to indicator I_k and τ^s_ℓ are strictly increasing class-specific thresholds defining an ordinal relation between the utility $G^s_{I_k,n}$ and the answers to I_k. The probability of an individual n providing an answer ℓ to indicator I_k is:

$$P_n(I_k = \ell|s) = \frac{1}{1+\exp(-(\tau^s_\ell - g(X_{kn}; \alpha^s_k)))} - \frac{1}{1+\exp(-(\tau^s_{\ell-1} - g(X_{kn}; \alpha^s_k)))} \tag{9}$$

Because a complete set of thresholds τ_ℓ, for $\ell = 1, \ldots, L-1$, cannot be fully identified, it is necessary to manually fix one of them (Greene and Hensher; 2009). For example, if the first threshold is fixed to zero ($\tau_1^s = 0$), then only the difference between thresholds (δ_ℓ) has to be estimated if the following definitions are considered:

$$\begin{aligned} \tau_2^s &= 0 + \delta_1^s \\ \tau_3^s &= \tau_2^s + \delta_2^s \\ &\dots \\ \tau_{L-1}^s &= \tau_{L-2}^s + \delta_{L-2}^s \end{aligned} \tag{10}$$

The parameters of the joint model of choice, class-membership and response to psychometric indicators can be simultaneously estimated by maximizing the following likelihood function:

$$\mathscr{L} = \prod_n \left\{ \sum_s \left\{ P_n(i|s) \prod_k P_n(I_k|s) \right\} P_n(s) \right\}, \tag{11}$$

where we adopt the following simplified notations:

$$P_n(i|s) := \prod_i P_n(i|s)^{y_{in}} \tag{12}$$

$$P_n(I_k|s) := \prod_\ell P_n(I_k = \ell|s)^{y_{k\ell n}} \tag{13}$$

where y_{in} is a variable that assumes the value of 1 if individual n chose alternative i and 0 otherwise, and $y_{k\ell n}$ assumes the value of 1 if individual n chose answer ℓ to the indicator (or question) I_k. The proposed approach is applied to a mode choice case study, presented next.

4 CASE STUDY: MODE CHOICE IN RURAL SWITZERLAND

Data from a revealed preferences travel survey conducted in 2009 in rural areas of Switzerland was collected (EPFL; 2011). The travel survey describes socioeconomics and the complete tour of trips of the respondent for a given weekday including mode, purpose, departure and arrival times. Additionally, as psychometric indicators, the survey collected responses in terms of level of agreement to a series of statements about the environment, the transport system, lifestyle preferences and mobility habits (for more details see Hurtubia et al.; 2010). The answers were collected using a five point Likert scale ranging from strong disagreement (level 1) to a strong agreement (level 5). After data cleaning and processing, the observations of trips and set of answers to the psychometric indicators of 1763 respondents were considered for estimation. In total, 2265 trips with an associated choice of transport mode were recorded, given that a respondent could report several trips per day. Each trip was considered as an independent decision and the potential panel effect was ignored. From the surveyed trips, 28% were performed by Public Transport (PT), including bus, metro and train, 66% by Private Motorized Modes (PMM), including car as driver, car as passenger, motorcycle and taxi and 6% by Soft Model (SM), including bicycle and walking. After weighting each observation in order to match the aggregate population distributions from the Census in the area of study (EPFL; 2011), the modal market shares turned out to be 34% for PT, 61% for PMM and 5% for SM.

4.1 Specification

This case study is an extension of the model and results presented by Atasoy et al. (2013). For comparison purposes, the specification of utility functions, definition of latent classes and selection of psychometric indicators are the same as those proposed in the aforementioned article.

Atasoy et al. (2013) identified class 1 as individuals living with their families (children and/or spouse) who have high income while class 2 corresponds to single individuals of lower income.

The class-membership functions are the following

$$\begin{aligned} f(X_n, \gamma^1) &= \mathrm{ASC}^1 + \gamma^1_{child} child_n + \gamma^1_{inc} high\_inc_n \\ f(X_n, \gamma^2) &= \gamma^2_{single} single_n \end{aligned} \tag{14}$$

The class-membership model depends on three main socioeconomic attributes of the decision maker: a dummy variable indicating if the traveler n belongs to a household with children ($child_n$), a dummy indicating if the income in the household is above CHF 8000 per month ($high\_inc_n$) and a dummy indicating if individual n lives alone or with his parents ($single_n$).

The mode choice model considers three alternatives: Private Motorized Modes (PMM), including car as driver, car as passenger, motorcycle and taxi, Public Transport (PT), including bus, metro and train, and Soft Modes (SM) including bicycle and walking. The class-specific utilities for mode choice are described in each column of Table 1. Because there was no observations of soft modes chosen by individuals falling in the "single" category, this alternative was made unavailable for class 2. This can be interpreted as a deterministic membership to class 1 for those individuals who choose to walk or bike.

In Table 1, TT_{PMM} and TT_{PT} are the travel times for private modes and public transport respectively, $cars$ is the number of cars in the household, $children$ is the number of children under age 15 in the household and $bikes$ is the number of bicycles available to the members of the household. $French$ is a dummy variable indicating if the respondent lives in the French part of Switzerland, $WorkTrip$ is a dummy indicating that the purpose of the trip was work, $Urban$ is a dummy indicating the origin or destination of the trips is in an urban area and $Student$ is a dummy indicating if the respondent is a student (up to the university or trainee level).

After a factor analysis process, Atasoy et al. (2013) selected the following statement of the survey to be used as indicators:

- I1 (PT and children): *It is hard to take public transport when I travel with my children.*
- I2 (Flexibility of car): *With my car, I can go where I want whenever I want.*
- I3 (Family oriented): *I would like to spend more time with my family and friends.*

The item response functions of each indicator are the following.

$$g(X_{1n}; \alpha^s_1) = ASC^s_{I1} + \alpha^s_{Children} HasChildren_n \tag{15}$$

Table 1: Mode choice model specification

	Variable (Class 1)			Variable (Class 2)	
Parameter	V_{PMM}	V_{PT}	V_{SM}	V_{PMM}	V_{PT}
ASC^1_{PMM}	1	-	-	-	-
ASC^2_{PMM}	-	-	-	1	-
ASC^1_{SM}	-	-	1	-	-
β^1_{cost}	$Cost_{PMM}$	$Cost_{PT}$	-	-	-
β^2_{cost}	-	-	-	$Cost_{PMM}$	$Cost_{PT}$
$\beta^1_{TT,PMM}$	TT_{PMM}	-	-	-	-
$\beta^2_{TT,PMM}$	-	-	-	TT_{PMM}	-
$\beta^1_{TT,PT}$	-	TT_{PT}	-	-	-
$\beta^2_{TT,PT}$	-	-	-	-	TT_{PT}
$\beta^1_{distance}$	-	-	$Dist_{SM}$	-	-
β_{cars}	$cars$	-	-	$cars$	-
$\beta^1_{children}$	$children$	-	-	-	-
$\beta^2_{children}$	-	-	-	$children$	-
$\beta_{language}$	$French$	-	-	$French$	-
β^1_{work}	$WorkTrip$	-	-	-	-
β^2_{work}	-	-	-	$WorkTrip$	-
β_{urban}	-	$Urban$	-	-	$Urban$
$\beta_{student}$	-	$Student$	-	-	$Student$
β^1_{bikes}	-	-	$bikes$	-	-

$$g(X_{2n};\alpha_2^s) = ASC_{I2}^s + \alpha_{cars}^s cars_n \tag{16}$$

$$g(X_{3n};\alpha_3^s) = ASC_{I3}^s + \alpha_{family}^s HasChildren_n working_n \tag{17}$$

The answer to indicator I1 will be affected by a dummy indicating the presence of children in the household; the number of cars in the household affects the answer to question I2 and the answer to indicator I3 depends on the interaction of two dummy variables indicating that the person has children and a full time job.

4.2 Estimation results

For comparison purposes, three models are estimated. The first one is simply an integrated choice and latent class model without indicators (LCM1), with likelihood

$$\mathscr{L}_{\text{LCM1}} = \prod_n \{P_n(i|\text{class }1)\cdot P_n(\text{class }1) + P_n(i|\text{class }2)\cdot P_n(\text{class }2)\} \tag{18}$$

The second one incorporates indicators and estimates the item response probabilities directly as parameters (LCM2). It has the following likelihood:

$$\begin{aligned}\mathscr{L}_{\text{LCM2}} = \prod_n \{&P_n(i|\text{class }1)\cdot \pi_{11}\cdot\pi_{21}\cdot\pi_{31}\cdot P_n(\text{class }1)\\ &+P_n(i|\text{class }2)\cdot \pi_{12}\cdot\pi_{22}\cdot\pi_{32}\cdot P_n(\text{class }2)\},\end{aligned} \tag{19}$$

where $P_n(I_k|s) := \pi_{ks}$. The third one uses the methodology proposed in this paper (LCM3) and has likelihood

$$\begin{aligned}\mathscr{L}_{\text{LCM3}} = \prod_n \{&P_n(i|\text{class }1)\cdot P_n(I1|\text{class }1)\cdot P_n(I2|\text{class }1)\cdot P_n(I3|\text{class }1)\cdot P_n(\text{class }1)\\ &+P_n(i|\text{class }2)\cdot P_n(I1|\text{class }2)\cdot P_n(I2|\text{class }2)\cdot P_n(I3|\text{class }2)\cdot P_n(\text{class }2)\}\end{aligned} \tag{20}$$

All models have the same specification for the utility functions relative to the choice model and the class-membership function. Results for the choice model and the class-membership model are shown in Table 2. The estimated item response probabilities for LCM2 and the parameters for the indicator measurements of LCM3 are shown in Tables 3 and 4 respectively.

The choice model parameters for cost time and show the expected sign for all classes in the three models, with a significant variation in magnitude between models. Class 2 shows a systematic higher sensitivity to cost and travel time for all modes in the three models. Most of the remaining parameters show intuitive values and no change of sign across models, with some exceptions. For example $\beta^2_{children}$ is negative and significant for LCM1 and it becomes positive and not significant for LCM3. However, the result for LCM3 seems to be more intuitive given the fact that individuals

Table 2: Estimation results

		LCM1		LCM2		LCM3	
	Parameters	estimate	t-test	estimate	t-test	estimate	t-test
Mode choice	ASC^1_{PMM}	-0.417	-0.417*	-0.945	-3.83	-1.25	-4.30
	ASC^2_{PMM}	-0.571	-1.49*	-0.936	-3.37	-0.731	-2.54
	ASC^1_{SM}	0.587	1.67*	0.512	1.70*	0.642	2.07
	β^1_{cost}	-4.15	-2.12	-2.70	-3.14	-1.23	-1.53*
	β^2_{cost}	-30.5	-4.83	-30.2	2.82	-39.1	-6.98
	$\beta^1_{TT,PMM}$	-0.211	-0.42*	-1.61	-4.77	-1.30	-3.80
	$\beta^2_{TT,PMM}$	-26.8	-4.96	-11.1	-6.83	-10.6	-6.46
	$\beta^1_{TT,PT}$	-0.257	-0.98*	-0.692	-3.62	-0.701	-3.55
	$\beta^2_{TT,PT}$	-8.91	-4.85	-4.45	-5.90	-3.91	-5.35
	$\beta^1_{distance}$	-18.4	-8.42	-19.9	-9.54	-19.8	-9.10
	β_{cars}	1.24	10.18	1.23	11.34	1.29	11.18
	$\beta^1_{children}$	0.403	2.76	0.404	4.83	0.346	3.47
	$\beta^2_{children}$	-0.434	-1.89	-1.03	-1.72*	0.211	0.97*
	$\beta_{language}$	1.20	0 5.71	1.20	6.79	1.20	6.22
	β^1_{work}	-0.990	-3.98	-0.785	-4.85	-0.623	-3.37
	β^2_{work}	0.0881	0.22*	-0.130	-0.43*	-0.396	-1.34*
	β_{urban}	0.528	3.20	0.390	2.82	0.459	3.23
	$\beta_{student}$	3.73	8.37	3.70	8.45	3.95	8.86
	β^1_{bikes}	0.400	4.96	0.205	3.21	0.214	3.26
Class	ASC_{class}	-0.215	-0.86*	-0.629	-3.25	-0.589	-3.39
	γ^1_{child}	0.136	0.51*	3.92	4.84	0.967	5.41
	γ^1_{inc}	0.693	2.76	0.460	2.22	0.684	4.50
	γ^2_{single}	0.408	1.34*	0.704	3.57	0.743	3.33
Log-like for choices		-994.7		-1032.5		-1006.7	
Log-likelihood for I1		-		-2068.4		-2033.1	
Log-likelihood for I2		-		-2202.6		-2151.5	
Log-likelihood for I3		-		-2160.6		-2153.5	

*Parameter not significant at the 95% confidence level

Table 3: Item response probabilities for LCM2

	$s=1$		$s=2$	
Probability	estimate	t-test	estimate	t-test
$P(I1=1\|s)$	0.166	13.00	0.002	0.78*
$P(I1=2\|s)$	0.246	16.14	0.008	0.67*
$P(I1=3\|s)$	0.306	14.11	0.958	34.60
$P(I1=4\|s)$	0.176	13.45	0.029	2.36
$P(I1=5\|s)$	0.106	**	0.003	**
$P(I2=1\|s)$	0.031	5.60	0.020	3.31
$P(I2=2\|s)$	0.033	5.73	0.027	3.94
$P(I2=3\|s)$	0.121	11.10	0.169	10.80
$P(I2=4\|s)$	0.371	23.87	0.364	18.03
$P(I2=5\|s)$	0.444	**	0.420	**
$P(I3=1\|s)$	0.013	3.63	0.004	1.35*
$P(I3=2\|s)$	0.047	6.84	0.040	4.80
$P(I3=3\|s)$	0.254	17.08	0.414	19.78
$P(I3=4\|s)$	0.491	29.91	0.430	20.46
$P(I3=5\|s)$	0.195	**	0.112	**

*Parameter not significant at the 95% confidence level

** The probability for $I=5$ is computed directly as $1-\sum_{k=1}^{4} P(I=k|s)$, $\forall s$ and, therefore, does not have an associated t-test

Table 4: Indicator measurement parameters for LCM3

	$s=1$		$s=2$	
parameter	estimate	t-test	estimate	t-test
ASC^s_{I1}	2.04	12.97	5.18	3.13
$\alpha^s_{Children}$	-1.28	-6.85	3.87	10.50
$\delta^s_{1,I1}$	1.57	15.02	0.461	0.35*
$\delta^s_{2,I1}$	1.96	10.77	7.40	4.08
$\delta^s_{3,I1}$	1.18	6.66	1.94	9.12
ASC^s_{I2}	2.26	8.60	3.31	9.81
α^s_{cars}	5.11	7.03	2.84	4.29
$\delta^s_{1,I2}$	0.845	5.15	0.781	3.51
$\delta^s_{2,I2}$	1.32	9.82	1.32	9.82
$\delta^s_{3,I2}$	1.79	17.44	1.74	17.06
ASC^s_{I3}	3.86	12.88	6.26	3.20
α^s_{family}	0.309	2.05	0.987	5.76
$\delta^s_{1,I3}$	1.31	5.26	3.33	1.76*
$\delta^s_{2,I3}$	2.07	13.51	2.69	13.96
$\delta^s_{3,I3}$	2.39	19.99	2.08	18.84

*Parameter not significant at the 95% confidence level in class 2 are likely to be single and without children. Another case of change of sign is that of β^2_{work} which is positive and not significant for LCM1 and becomes negative with a higher significance in LCM3.

The estimates for the class-membership model confirm that class 1 corresponds to high income individuals living with their family while class 2 corresponds to single individuals with lower income. An important difference in magnitude is observed in these parameters among the three modeling approaches. LCM2 assigns a higher relative weight to the presence of children in the household as an explanatory variable while LCM3 assigns a relatively higher weight to the income level. In general the inclusion of indicators (in both LCM2 and LCM3) allows for the estimation of more significant parameters in the class-membership model.

The Log-likelihood for choice in Table 2 can be interpreted as a measurement of the fit of each model within the sample. While the best fit is obtained by the model without indicators (LCM1) the model proposed in this paper (LCM3) shows a better fit than the model estimating the item response probabilities directly as parameters (LCM2). It is important to mention here that the objective of the introduction of psychometric indicators in the model is not to improve the fit but to gain additional insight of user behavior, hence the lower fit for both model including indicators.

Regarding the measurement of indicators, both LCM2 and LCM3 generate response probabilities (see Table 3) that are consistent with observed response rates. Some additional behavioral interpretation is possible when looking at the indicator measurement parameters of LCM3 (see Table 4). For example, for indicator I1 (difficulty of using public transport with children), it is possible to see that class 2 has a strong inertial tendency to be indifferent, confirming that individuals in class 2 are likely to have no children. On the other hand, individuals in class 1 show a more heterogeneous behavior in their responses, which tends to be of disagreement when the household has children.

The models forecast market shares with some differences. In terms of value of time LCM1 predicts a counter-intuitive higher value of time for class 2. The models including indicators (LCM2 and LCM3) produce a more intuitive VOT for each class, although LCM3 predicts a much higher VOT for private motorized modes (PMM) in class 1. The reference VOT for Switzerland is 27.66 CHF/hour for business travels by car (Axhausen et al.; 2008). However, estimation data was obtained from a survey that was conducted in rural areas of Switzerland, where income tends to be higher, while the reference VOT considers both rural and urban areas. This, besides the fact that many individuals in class 1 have at least a wage of 50 CHF/hour[1], justifies considering the results provided by LCM3 as reasonable since, under some circumstances, the value of travel time savings should be close to the wage level (Jara-Diaz; 2007). This, however, requires further research to reach a conclusion.

5 CONCLUSIONS

We propose a new type of model specification that incorporates psychometric indicators into integrated choice and latent class models through an ordinal logit model. Moreover the ordinal logit model relates the answers to the indicators with socioeconomic characteristics of the respondents, hence allowing for a better characterization of the latent classes.

[1]computed as CHF 8000 divided by 160 hours of work per month

Table 5: Market shares and value of time

Models		PMM [%]	PT [%]	SM [%]	VOT PMM [CHF/h]	VOT PT [CHF/h]
LCM1	Class 1	60.97	28.73	10.30	3.06	3.72
	Class 2	60.41	39.59	-	52.63	17.53
	Overall	61.23	33.81	4.96	28.97	10.94
LCM2	Class 1	54.91	36.13	8.96	35.78	15.38
	Class 2	65.73	34.27	-	22.05	8.84
	Overall	62.7	32.35	4.94	29.53	12.40
LCM3	Class 1	51.79	38.01	10.2	63.27	16.21
	Class 2	70.98	29.02	-	34.16	5.99
	Overall	61.74	33.69	4.57	36.94	18.40

The method is tested for a mode choice case study in rural areas of Switzerland. Results show that the inclusion of the ordinal measurement of psychometric indicators generates significantly different estimates for the class-membership model. The additional behavioral insights provided by the parameters of the indicator-measurement equations allows for a richer analysis of the latent classes, giving the analyst more tools to identify different market segments.

The proposed method forecasts values of time of different magnitude when compared with latent class models that estimate the item response probabilities of the indicators as single parameters.

One of the advantages of the proposed methodology is the closed form of the ordinal logit used for measurement of the indicators. This allows for a simpler estimation procedure, without the need of integration techniques as it is in most cases when latent variables are included in choice models.

Some of the estimates in the models presented here have a low significance level and some of the utility functions for classes and indicators have considerably simple specifications. We believe that using a larger set of observations should allow to incorporate more explanatory variables in the class-membership and indicator measurement utilities, therefore expanding the possibilities of behavioral analysis and market segmentation.

ACKNOWLEDGMENTS

Research in this article has been partially funded by the Complex Engineering Systems Institute (ICM: P-05-004-F, CONICYT: FBO16) and by PostBus Switzerland.

References

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

Axhausen, K. W., Hess, S., Koenig, A., Abay, G., Bates, J. J. and Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New Swiss results, *Transport Policy* **15**(3): 173–185.

Ben-Akiva, M. E. and Boccara, B. (1995). Discrete choice models with latent choice sets, *International Journal of Research in Marketing* **12**: 9–24.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Börsch-Supan, A., Brownstone, D., Bunch, D. et al. (2002). Hybrid choice models: Progress and challenges, *Marketing Letters* **13**(3): 163–175.

Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models, *In perpetual motion: Travel behaviour research opportunities and application challenges, Elsevier, Amsterdam* pp. 431–470.

Bhat, C. (1997). An endogenous segmentation mode choice model with an application to intercity travel, *Transportation Science* **31**(1): 34–48.

Boxall, P. and Adamowicz, W. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach, *Environmental & Resource Economics* **23**(4): 421–446.

Collins, L. M. and Lanza, S. T. (2010). *Latent class and latent transition analysis: with application in the social, behavioral, and health sciences*, Wiley Series in Probability and Statistics.

EPFL (2011). Projet de recherche sur la mobilité combinée, OPTIMA: Rapport définitif de l'enquête de préférences révélées, *Technical report*, CarPostal Suisse SA and Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.
URL: *http://infoscience.epfl.ch/record/167785*

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

Gopinath, D. A. (1995). *Modeling heterogeneity in discrete choice processes: application to travel demand*, PhD thesis, Massachusetts Institute of Technology.

Greene, W. H. and Hensher, D. A. (2009). *Modeling ordered choices*, Cambridge University Press.

Greene, W. and Hensher, D. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit, *Transportation Research Part B: Methodological* **37**(8): 681–698.

Hess, S., Ben-Akiva, M., Gopinath, D. and Walker, J. (2011). Advantages of latent class over continuous mixture of logit models, *Technical report*, Working paper, University Press, Harrisburg.

Hosoda, T. (1999). *Incorporating unobservable heterogeneity in discrete choice model: mode choice model for shopping trips*, PhD thesis, Massachusetts Institute of Technology.

Hurtubia, R., Atasoy, B., Glerum, A., Curchod, A. and Bierlaire, M. (2010). Considering latent attitudes in mode choice: The case of Switzerland, *Proceedings of the World Conference on Transport Research*, Lisbon, Portugal.

Jara-Diaz, S. (2007). *Transport economic theory*, Elsevier Science.

Kamakura, W. A. and Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research* **26**(4): pp. 379–390.

Keane, M. and Wasi, N. (2012). Comparing alternative models of heterogeneity in consumer choice behavior, *Journal of Applied Econometrics* .

Koutsopoulos, H. and Farah, H. (2012). Latent class model for car following behavior, *Transportation Research Part B: Methodological* **46**(5): 563–578.

Lazarsfeld, P. and Henry, N. (1968). *Latent structure analysis*, Houghton, Mifflin.

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

McFadden, D. (1986). The choice theory approach to market research, *Marketing Science* **5**(4): 275–297.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response, *Journal of applied Econometrics* **15**(5): 447–470.

Morey, E., Thacher, J. and Breffle, W. (2006). Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model, *Environmental and Resource Economics* **34**: 91–115.

Shen, J. (2009). Latent class model or mixed logit model? a comparison by transport mode choice data, *Applied Economics* **41**(22): 2915–2924.

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

Walker, J. L. and Li, J. (2007). Latent Lifestyle Preferences and Household Location Decisions, *Journal of Geographical Sytems* **9**: 77–101.

Wen, C.-H., Wang, W.-C. and Fu, C. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice, *Transportation Research Part E: Logistics and Transportation Review* **48**(2): 545 – 554.

Wen, C. and Lai, S. (2010). Latent class models of international air carrier choice, *Transportation Research Part E: Logistics and Transportation Review* **46**(2): 211–221.

Zhang, J., Kuwano, M., Lee, B. and Fujiwara, A. (2009). Modeling household discrete choice behavior incorporating heterogeneous group decision-making mechanisms, *Transportation Research Part B: Methodological* **43**(2): 230–250.