Investigating the subjective and objective factors influencing teenagers' school travel mode choice – An integrated choice and latent variable model

Maria Kamargianni [a,*], Subodh Dubey [b,1], Amalia Polydoropoulou [c,2], Chandra Bhat [b,d,1]

[a] UCL Energy Institute, University College London, Central House, 14 Upper Woburn Place, WC1H 0NN London, UK
[b] The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, United States
[c] Department of Shipping, Trade and Transport, University of the Aegean, Korai 2a, Chios 82100, Greece
[d] King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

ABSTRACT

In this paper, we apply Bhat and Dubey's (2014) new probit-kernel based Integrated Choice and Latent Variable (ICLV) model formulation to analyze children's travel mode choice to school. The new approach offered significant advantages, as it allowed us to incorporate three latent variables with a large data sample and with 10 ordinal indicators of the latent variables, and still estimate the model without any convergence problems. The data used in the empirical analysis originates from a survey undertaken in Cyprus in 2012. The results underscore the importance of incorporating subjective attitudinal variables in school mode choice modeling. The results also emphasize the need to improve bus and walking safety, and communicate such improvements to the public, especially to girls and women and high income households. The model application also provides important information regarding the value of investing in bicycling and walking infrastructure.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Discrete Choice Models (DCMs) consider aggregate consumer demand to be the result of a combination of several decisions made by each individual of a population under consideration, where each decision of each individual consists of a choice made among a finite set of available alternatives (Ben-Akiva and Lerman, 1985). DCMs explain individual choice behavior as the consequence of preferences that an individual ascribes to her or his available set of alternatives, with the assumption that the consumer then chooses the most preferred available outcome. Under certain assumptions, consumer preferences can be represented by a utility function such that the choice is the utility maximizing outcome. These utility maximizing models have traditionally presented an individual's choice process as somewhat of a "black box", in which the inputs are the attributes of available alternatives and the individual's characteristics, and the output is the observed choice (Ben-Akiva et al., 2002). Behavioral researchers have stressed the importance of the cognitive workings inside the

black box in determining choice behavior (Olson and Zanna, 1993; Gärling et al., 1998), and a substantial amount of research now has been conducted to uncover cognitive decision-making strategies that appear to violate the basic axioms of utility theory (Morikawa, 1989; Gopinath, 1994; Bhat, 1997; Rabin, 1998; Walker, 2001; Johansson et al., 2006; Kamargianni et al., 2014).

Over the last few decades, numerous improvements have been made that aim to better unravel the underlying process leading up to observed choice outcomes, while also better predicting the outcomes of choice behavior. These methods are integrated in Hybrid Choice Models (HCMs). HCMs, by combining "hard information" (such as socioeconomic characteristics) with "soft information" on population heterogeneity (such as psychological characteristics), attempt to more realistically explain individual choice behavior and in doing so a substantial part of the population heterogeneity (Ben-Akiva et al., 2002).

Among the numerous versions of HCMs is the explicit modeling of latent psychological factors such as attitudes and perceptions (latent variables). The Integrated Choice and Latent Variable (ICLV) model inside the HCM conceptual framework permits the inclusion of attitudes, opinions and perceptions as psychometric latent variables in such a way that consumer behavior is better understood, while the model also gains in predictive power (Ashok et al., 2002; Ben-Akiva et al., 2002; Bolduc et al., 2005; Bhat and Dubey, 2014).[3]

Although the number of applications of ICLV models has been on the rise in the last decade (see, for example, Bolduc et al., 2005; Johansson et al., 2006; Temme et al., 2008; Abou-Zeid et al., 2011; Daly et al., 2012; Polydoropoulou et al., 2014; Kamargianni and Polydoropoulou, 2013; Alvarez-Daziano and Bolduc, 2013), Bhat and Dubey (2014) indicate that the conceptual value of ICLV models has not been adequately translated to benefits in practice because of the difficulties in model convergence and full likelihood estimation, and the very lengthy estimation times of these models even when convergence is achieved. These issues are particularly the case when more than one or two latent variables are considered within the traditional logit kernel-based ICLV model formulation, since the number of latent variables has a direct impact on the dimensionality of the integral that needs to be estimated in the log-likelihood function. The consequence has been that most ICLV models in the literature have gravitated toward the use of a very limited number of latent constructs (typically a single latent variable), rather than exploring a fuller set of possible latent variables.[4] In addition, in the frequentist full likelihood estimation method for the traditional logit kernel-based ICLV, the use of ordinal indicators creates substantial problems because of the increase in the number of multiplicative mixing components in the integrand of the resulting likelihood function. As detailed by Bhat and Dubey (2014), convergence in likelihood estimation becomes challenging as the number of mixing components in the integrand of a logit based-kernel ICLV model increases. Thus, it is not unusual to use only continuous indicators in such frequentist-based ICLV estimations. Also, while Alvarez-Daziano and Bolduc (2013) present a Bayesian Markov Chain Monte Carlo (MCMC) simulation approach to estimating the logit kernel-based ICLV model, the approach requires extensive simulation and can become cumbersome when estimating highly non-linear models such as the ICLV model (see Franzese et al., 2010 for a discussion of this issue). The Bayesian approach also poses convergence assessment problems as the number of latent variables or the number of ordinal indicator variables increases in the logit kernel-based ICLV model.

In the context of the above application difficulties with the logit-based ICLV model, Bhat and Dubey (2014) proposed an MNP kernel-based ICLV formulation that allows the incorporation of a large number of latent variables in the choice model without convergence difficulties or estimation time problems. There are three key reasons behind smooth convergence and reasonable estimation time in their proposed approach: (1) The dimensionality of integration is independent of the number of latent variables (this is not the case with previous logit kernel based ICLV models) and is dependent only on the number of ordinal variables and number of alternatives; this allows the analyst to incorporate as many latent variables as required without worrying about estimation time, (2) The use of a Composite Marginal Likelihood (CML) approach as opposed to a full-likelihood approach simplifies the high dimensional integral in the estimation function into a number of manageable lower dimensional integrals; the net result is that the dimensionality of integration in the estimation function is now independent of the number of latent variables and the number of ordinal indicator variables, and is only of the order of the number of alternatives in the choice model,[5] and (3) an analytic approximation is employed to evaluate the multivariate normal probabilities instead of using a simulation based approach such as a GHK simulator, resulting in smoothness of the

---

[3] A precursor to the ICLV model is structural equations modeling (SEM), originating in the early works of Jöreskog (1977). However, the SEM field has focused almost exclusively on non-nominal outcome analysis (see Gates et al., 2011 and Hoshino and Bentler, 2013). Indeed, traditional SEM software (such as LISREL, MPLUS, and EQS) is either not capable of handling nominal indicators or at least are not readily suited to handle nominal indicators (see Temme et al., 2008). Thus, ICLV models and SEM models, while having some common traits, are often not referenced within a single paper. In fact, none of the earlier trasnportation-based ICLV model applications in the past that we are aware of bring up the topic of SEMs, with the exception of Temme et al., 2008. In addition, many SEMs in the past (see, for example, Golob, 2003) are estimated using three-stage or two-stage sequential estimation methods (see Temme et al., 2008 and Katsikatsou, 2013 for discussions of these sequential methods). The problem with such sequential methods is that they do not account for sampling variability induced in earlier steps in the later steps, leading to inefficient estimation. In addition, the use of such sequential methods will, in general, also lead to inconsistent estimation (see Walker, 2001 and Alvarez-Daziano and Bolduc, 2013 for discussions of the reasons). As a result, almost all ICLV applications have used a full information likelihood estimation procedure, using simulation when the dimensionality of integration increases.

[4] While it is true that the power of the ICLV models arises in part through the use of a small set of latent variables to generate a parsimonious factor-analytic error dependency structure across a large number of alternatives in a choice model, and across these alternatives with other continuous/ordinal outcomes, the point is that almost all earlier studies have placed very restrictive factor-analytic structures by specifying a single latent variable (due to computational problems otherwise) rather than attempting to test (and capture) a richer factor analytic structure by specifying more latent variables.

[5] A complete description of Composite Marginal Likelihood (CML) approach is beyond the scope of this paper; readers are referred to Bhat (2014) for a comprehensive discussion of the CML approach in the context of choice modeling.

analytically approximated log-likelihood surface and leading to well-behaved surfaces for the gradient and hessian functions; this is a substantial advantage over the non-smooth surfaces in simulation-based approaches that frequently causes convergence problems. In addition, the analytic approximation implies that the analyst needs to maximize a function that has no more than bivariate normal cumulative distribution functions to be evaluated, regardless of the number of latent variables, the number of ordinal indicators, or the number of alternatives in the choice model.[6]

The aim of this paper is to empirically apply Bhat and Dubey's (2014) formulation to develop a mode choice ICLV model that incorporates three latent psychological factors associated with safety consciousness, environmental consciousness and physical activity propensity. The data used in this research originates from the first wave of the survey undertaken in the Republic of Cyprus in 2012 that collected travel mode choice data from individuals close to their teenage years (11–18 years old; for ease in presentation, we will refer to these individuals as teenagers).[7] The sample is drawn from the same survey as that used in Kamargianni and Polydoropoulou (2013), but we use three fundamental latent constructs that represent a combination of underlying affective value norms/beliefs, lifestyle orientations, and personality traits rather than a more generic single "willingness to walk/cycle" attitude as the latent construct (as emphasized by Temme et al., 2008, recent research has highlighted the importance of considering basic underlying constructs in mode choice modeling, rather than using relatively superficial constructs such as the willingness or propensities to use specific modes that anyway are already considered in the form of the underlying modal utility in choice models). Specifically, we use three latent constructs to explain the school mode choice of teenagers ("safety consciousness", "green lifestyle", and "physical activity propensity"). In doing so, we obtain a much richer interpretation of the individual factors affecting mode choice, which we discuss in substantial detail. The current paper also proposes new measures of fit for comparing the ICLV model with a model without latent constructs, an issue that has not received the attention it deserves. In addition to these important empirical differences, we, of course, use a very different ICLV formulation and estimation method in this paper relative to Kamargianni and Polydoropoulou (2013).

To be sure, many recent empirical investigations of travel mode choice have adopted a single psychological construct associated with either safety consciousness, green lifestyle, or physical activity propensity in mode choice modeling. For example, some studies have used a safety consciousness latent construct to examine safety issues related to the transport network/built-environment characteristics and their impact on active transport behavior (for example, see Chataway et al., 2014 and Heinen and Handy, 2012) and the choice of public transport (see Johansson et al., 2006; Daly et al., 2012; Tyrinopoulos and Antoniou, 2013). Similarly, studies have examined the impact of environmental consciousness or protection tendency (based on diverse indicators collected from attitudinal surveys), revealing that a green lifestyle positively affects the probability of choice of active transport (walking and cycling), and reduces the probability of choosing private motorized vehicles (for example, see Outwater et al., 2003; Anable, 2005; Hunecke et al., 2007; Shiftan et al., 2008; Atasoy et al., 2010; Daly et al., 2012; Rieser-Schüssler and Axhausen, 2012; Tyrinopoulos and Antoniou, 2013; Hess et al., 2013). Finally, while the notion that physical activity propensity has a positive impact on active transport mode choice is intuitive, no study in the transport sector that we are aware of has examined this effect (though studies in the public health field have revealed a positive impact of physical activity propensity on recreational walking and bicycling; see, for example, Weikert et al., 2010).

Unlike the studies mentioned above that have focused on a single psychological construct in explaining mode choice, we consider all the three constructs; safety consciousness, green lifestyle, and physical activity propensity; simultaneously. We are able to do so because of the probit kernel-based approach that easily and practically accommodates a multitude of latent variables. Indeed, to our knowledge, this is the first time that an ICLV model has been estimated simultaneously using more than two latent variables on a panel dataset using a simulation free approach. Finally, the majority of the existing studies that use latent variables in travel models have focused on adults' unobserved factors that affect travel behavior; in contrast, the emphasis here is on understanding how teenagers' own attitudes affect their mode choice patterns.

The rest of the paper is organized as follows. Section 2 presents the model formulation. Section 3 presents the data and sample characteristics. The estimation results of the models are presented and discussed in Section 4. Section 5 concludes the paper by summarizing the key findings and providing directions for further research.

2. Model formulation and estimation

There are three components to the model: (1) the latent variable structural equation model, (2) the latent variable measurement equation model, and (3) the choice model. These components are discussed in turn below. In the following presentation, we will use the index l for latent variables ($l = 1, 2, \ldots, L$), and the index i for alternatives ($i = 1, 2, \ldots, I$) and t for choice occasion ($t = 1, 2, \ldots, T$). As appropriate and convenient, we will suppress the index q for individuals ($q = 1, 2, \ldots, Q$) in parts of the presentation.

---

[6] Recently, Daziano (2015) has suggested an MCMC-based estimation approach for Bhat and Dubey's (2014) probit kernel-based ICLV model formulation. We are currently collaborating with Daziano to investigate the challenges and advantages of the Bhat and Dubey (2014) and MCMC-based estimation approaches.

[7] Teenagers have been known to be mature enough to make their own mode choices when they travel to school (see Clifton, 2003; Clifton et al., 2010, and Babey et al., 2009).

2.1. Latent variable structural equation model

For the latent variable structural equation model, we will assume that the latent variable z_l^* is a linear function of covariates as follows:

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

where $\boldsymbol{w}$ is a $(\widetilde{D} \times 1)$ vector of observed covariates (not including a constant), $\boldsymbol{a}_l$ is a corresponding $(\widetilde{D} \times 1)$ vector of coefficients, and η_l is a random error term assumed to be normally distributed. In our notation, the same exogenous vector $\boldsymbol{w}$ is used for all latent variables; however, this is in no way restrictive, since one may place the value of zero in the appropriate row of $\boldsymbol{\alpha}_l$ if a specific variable does not impact z_l^*. Also, since z_l^* is latent, it will be convenient to impose the normalization discussed in Stapleton (1978) and used by Bolduc et al. (2005) by assuming that η_l is standard normally distributed. Next, define the $(L \times \widetilde{D})$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_L)'$, and the $(L \times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, \ldots, z_L^*)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$. To allow correlation among the latent variables, $\boldsymbol{\eta}$ is assumed to be standard multivariate normally distributed: $\boldsymbol{\eta} \sim N[\boldsymbol{0_L}, \boldsymbol{\Gamma}]$, where $\boldsymbol{\Gamma}$ is a correlation matrix (as indicated earlier in Section 1, it is typical to impose the assumption that $\boldsymbol{\eta}$ is diagonal, but we do not do so to keep the specification general). In matrix form, Eq. (1) may be written as:

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

2.2. Latent variable measurement equation model

All the indicator variables (that provide information on the latent variables) are ordinal in nature in our empirical context. In the general case, let there be G ordinal indicator variables, and let g be the index for the ordinal variables ($g = 1, 2, \ldots, G$). Let the index for the ordinal outcome category for the gth ordinal variable be represented by j_g. For notational ease only, assume that the number of ordinal categories is the same across the ordinal indicator variables, so that $j_g \in \{1, 2, \ldots, J\}$. Let y_g^* be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the gth ordinal indicator variable, and let the individual under consideration choose the n_gth ordinal outcome category for the gth ordinal indicator variable. Then, in the usual ordered response formulation, we may write:

$$y_g^* = \delta_g + \boldsymbol{d}_g' \boldsymbol{z}^* + \xi_g, \quad \text{and} \quad \psi_{g,n_g-1} < y_g^* < \psi_{g,n_g},$$

where δ_g is a scalar constant, $\boldsymbol{d_g}$ is an $(L \times 1)$ vector of latent variable loadings on the underlying variable for the gth indicator variable, and ξ_g is a standard normally distributed measurement error term (the normalization on the error term is needed for identification, as in the usual ordered-response model; see McKelvey and Zavoina, 1975). Note also that, for each ordinal indicator variable, $\psi_{g,0} < \psi_{g,1} < \psi_{g,2} \cdots < \psi_{gJ_g-1} < \psi_{gJ_g}$; $\psi_{g,0} = -\infty$, $\psi_{g,1} = 0$, and $\psi_{gJ_g} = +\infty$. For later use, let $\boldsymbol{\psi_g} = (\psi_{g,2}, \psi_{g,3} \ldots, \psi_{gJ_g-1})'$ and $\boldsymbol{\psi} = (\boldsymbol{\psi_1'}, \boldsymbol{\psi_2'}, \ldots, \boldsymbol{\psi_N'})'$. Stack the G underlying continuous variables y_g^* into a $(G \times 1)$ vector y^* and the G constants δ_g into a $(G \times 1)$ vector $\boldsymbol{\delta}$. Also, define the $(G \times L)$ matrix of latent variable loadings $\boldsymbol{d} = (\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_G})'$, and let $\boldsymbol{\Sigma}$ be the correlation matrix of $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_G)$. Stack the lower thresholds $\psi_{g,n_g-1}(g = 1, 2 \ldots, G)$ into a $(G \times 1)$ vector $\boldsymbol{\psi_{low}}$ and the upper thresholds $\psi_{g,n_g-1}(g = 1, 2 \ldots, G)$ into another vector $\boldsymbol{\psi_{up}}$ Then, in matrix form, the measurement equation for the ordinal indicators may be written as:

$$\boldsymbol{y}^* = \boldsymbol{\delta} + \boldsymbol{d}\boldsymbol{z}^* + \xi, \quad \boldsymbol{\psi_{low}} < \boldsymbol{y}^* < \boldsymbol{\psi_{up}} \tag{3}$$

2.3. Choice model

Assume a typical random utility-maximizing model, and let i be the index for alternatives ($i = 1, 2, 3, \ldots, I$). Note that some alternatives may not be available to some individuals during some choice instances, but the modification to allow this is quite trivial. Hence, for ease in presentation, we assume that all alternatives are available to all individuals at each of their choice instances. The utility for alternative i at time period t ($t = 1, 2, \ldots, T$) for individual q is then written as[8] (suppressing the index q):

$$U_{ti} = \boldsymbol{\beta}' \boldsymbol{x}_{ti} + \boldsymbol{\gamma_i'}(\boldsymbol{\varphi}_{ti}\boldsymbol{z}^*) + \varepsilon_{ti}, \tag{4}$$

where $\boldsymbol{x_{ti}}$ is a $(D \times 1)$-column vector of exogenous attributes. $\boldsymbol{\beta}$ is a $(D \times 1)$-column vector of corresponding coefficients, $\boldsymbol{\varphi}_{ti}$ is an $(N_i \times L)$-matrix of exogenous variables interacting with latent variables to influence the utility of alternative i, $\boldsymbol{\gamma_i}$ is an $(N_i \times 1)$-column vector of coefficients capturing the effects of latent variables and its interaction effects with other exogenous variables, and ε_{ti} is a normal error term that is independent and identically normally distributed across *individuals and choice occasions*. The notation above is very general. Thus, if each of the latent variables impacts the utility of alternative i purely through a constant shift in the utility function, $\boldsymbol{\varphi}_{ti}$ will be an identity matrix of size L, and each element of $\boldsymbol{\gamma}_i$ will

---

[8] The notations used here are drawn from Bhat (2015).

capture the effect of a latent variable on the constant specific to alternative i. Alternatively, if the first latent variable is the only one relevant for the utility of alternative i, and it affects the utility of alternative i through both a constant shift as well as an exogenous variable, then $N_i = 2$, and φ_{ti} will be a $(2 \times L)$-matrix, with the first row having a '1' in the first column and '0' entries elsewhere, and the second row having the exogenous variable value in the first column and '0' entries elsewhere.

Next, let the variance–covariance matrix of the vertically stacked vector of errors $\boldsymbol{\varepsilon}_t [= (\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tI})']$ be $\boldsymbol{\Lambda}$ and let $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \ldots, \boldsymbol{\varepsilon}_T')'$ ($TI \times 1$ vector). The covariance of $\boldsymbol{\varepsilon}$ is $\mathbf{IDEN}_T \otimes \boldsymbol{\Lambda}$, where $\mathbf{IDEN_T}$ is the identity matrix of size T. Define the following vectors and matrices:

$$\boldsymbol{x}_t = (\boldsymbol{x}_{t1}, \boldsymbol{x}_{t2}, \ldots, \boldsymbol{x}_{tI})' \ (I \times D \text{ matrix}), \quad \boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T')' \ (TI \times D \text{ matrix}),$$

$$\boldsymbol{U}_t = (U_{t1}, U_{t2}, \ldots, U_{tI})' \ (I \times 1 \text{ vector}), \quad \boldsymbol{U} = (\boldsymbol{U}_1', \boldsymbol{U}_2', \ldots, \boldsymbol{U}_T')' \ (TI \times 1 \text{ vector}),$$

$\boldsymbol{\varphi}_t = (\boldsymbol{\varphi}_{t1}', \boldsymbol{\varphi}_{t2}', \ldots, \boldsymbol{\varphi}_{tI}')' \ \left(\sum_{i=1}^{I} N_i \times L \right)$ matrix, $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots, \boldsymbol{\varphi}_T')' \ \left(T \sum_{i=1}^{I} N_i \times L \right)$. Also, define the $(I \times \sum_{i=1}^{I} N_i)$ matrix $\boldsymbol{\gamma}$, which is initially filled with all zero values. Then, position the $(1 \times N_1)$ row vector $\boldsymbol{\gamma}_1'$ in the first row to occupy columns 1 to N_1, position the $(1 \times N_2)$ row vector $\boldsymbol{\gamma}_2'$ in the second row to occupy columns $N_1 + 1$ to $N_1 + N_2$, and so on until the $(1 \times N_I)$ row vector $\boldsymbol{\gamma}_I'$ is appropriately positioned. Then, in matrix form, we may write the following equation for the vector of utilities across all choice instances of the individual:

$$\boldsymbol{U} = \boldsymbol{x}\boldsymbol{\beta} + (\mathbf{IDEN}_T \otimes \boldsymbol{\gamma})\boldsymbol{\varphi}\boldsymbol{z}^* + \boldsymbol{\varepsilon} = \boldsymbol{x}\boldsymbol{\beta} + \lambda\boldsymbol{z}^* + \boldsymbol{\varepsilon}, \quad \text{where } \lambda = (\mathbf{IDEN}_T \otimes \boldsymbol{\gamma})\boldsymbol{\varphi}(TI \times L \text{ matrix}) \tag{5}$$

In the above vector equation, the inclusion of the individual-specific $\boldsymbol{z}^*$ vector (that does not vary across the two choice occasions of the same individual) engenders a correlation across the multiple choice occasions of the same individual (because $\boldsymbol{z}^*$ is a stochastic vector).

As in the case of any choice model, one of the alternatives has to be used as the base when introducing alternative-specific constants and variables that do not vary across the I alternatives. Also, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\tilde{\boldsymbol{\Lambda}}$ of $\boldsymbol{\varsigma} = (\varsigma_2, \varsigma_3, \ldots, \varsigma_I)$, where $\varsigma_i = \varepsilon_i - \varepsilon_1$ $(i \neq 1)$, are estimable. $\boldsymbol{\Lambda}$ is constructed from $\tilde{\boldsymbol{\Lambda}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. In addition, an additional scale normalization needs to be imposed on $\tilde{\boldsymbol{\Lambda}}$, which may be accomplished normalizing the first element of $\tilde{\boldsymbol{\Lambda}}$ to the value of one.

General and necessary identification conditions for ICLV models have yet to be developed, but good discussions of sufficiency conditions may be found in Stapleton (1978), Vij and Walker (2014), Alvarez-Daziano and Bolduc (2013) and Bhat and Dubey (2014).

Let $\boldsymbol{\omega}$ be the collection of parameters to be estimated: $\boldsymbol{\omega} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vech}(\boldsymbol{\Gamma}), \boldsymbol{\delta}, \text{Vech}(\boldsymbol{d}), \boldsymbol{\psi}, \text{Vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}, \text{Vech}(\boldsymbol{\gamma}), \text{Vech}(\tilde{\boldsymbol{\Lambda}})]$, where $\text{Vech}(\boldsymbol{\alpha})$, $\text{Vech}(\boldsymbol{d})$, and $\text{Vech}(\boldsymbol{\gamma})$ represent vectors of the elements of the $\boldsymbol{\alpha}$, $\boldsymbol{d}$, and $\boldsymbol{\gamma}$, respectively, and $\text{Vech}(\boldsymbol{\Gamma})$ represents the vector of the non-zero upper triangle elements of $\boldsymbol{\Gamma}$ (and similarly for other covariance matrices).

To proceed, we first develop reduced form equations by replacing the right side of Eq. (1) for $\boldsymbol{z}^*$ in Eqs. (3) and (5) to obtain the following system:

$$\boldsymbol{y}^* = \boldsymbol{\delta} + \boldsymbol{d}(\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\eta}) + \xi = \boldsymbol{\delta} + \boldsymbol{d}\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{d}\boldsymbol{\eta} + \xi \tag{6}$$

$$\boldsymbol{U} = \boldsymbol{x}\boldsymbol{\beta} + \lambda(\boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\eta}) + \boldsymbol{\varepsilon} = \boldsymbol{x}\boldsymbol{\beta} + \lambda\boldsymbol{\alpha}\boldsymbol{w} + \lambda\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{7}$$

The estimation approach follows Bhat and Dubey (2014), and is not replicated here to conserve on space.

3. Data and latent constructs

The data used in this paper originates in the first wave of a survey undertaken in the Republic of Cyprus in February 2012 that aimed at capturing teenagers' school travel mode choice behavior. The survey was launched by the TransDem Lab (University of the Aegean) in co-operation with the Ministry of Education of Cyprus (MOEC), collecting stated preference (SP) data (for more details about the data collection and questionnaire design, please see Kamargianni, 2014). The survey was administered only in public schools in Cyprus.

The SP scenarios were designed in a way that was comprehensible to teenagers, based on pilot designs and testing. In our pilot survey, we tested various ways (i.e. text, pictures) to represent the levels of the attributes with a view to identify the one that is most clear and understandable to participants. Sidewalks and bike path availability were quite clear to participants when pictures were used. The pictures represented actual, real life constructs of the study area; this was to ensure that participants do not visualize other non-relevant built environments; something that could easily happen with verbal descriptions of these attributes. For weather, it refers to the weather condition during the time window when a participant generally leaves for school. Further, we tested various weather condition levels (i.e. a summer sunny day at around 20 °C). However, based on the results of the pilot survey, there were insignificant differences between detailed and simple weather descriptions, and thus the latter set of weather descriptions was employed.

The attributes and their levels used to develop the scenarios are presented in Table 1. A structured experimental design was undertaken to generate two scenarios for each individual (the details of this experimental design are available in Kamargianni and Polydoropoulou, 2013). The scenarios have five alternative transport modes for the trip to school: (1) Car – Escorted by an adult, (2) PTW (Powered Two Wheelers/motorcycles), (3) Bus, (4) Walk and (5) Bicycle. The attributes of the modes are travel time (specific to all alternatives), travel cost (specific to car, PTW and bus), parking place availability (specific to car, PTW and bike), walking time from home to the bus stop (specific to bus), existence of bicycle lanes around school area (specific to bike), walkability/condition of sidewalks (specific to walk) and weather conditions. Thus, each scenario presented has five alternatives and seven attributes (though not all attributes are relevant to all alternatives, as just described).

The sample characteristics are presented in Table 2. 50% of participants are aged between 14 and 15 years old, while 53% of the participants are girls. Regarding parental educational level, the majority of both parents in each teenager's household have received secondary education. A large fraction of households (about 30%) did not respond to the household income question, and so we imputed income for these households using the procedure in Bhat (1997). We do not provide the model description and the model results for this income imputation model to conserve on space. Readers are referred to Bhat (1997) for the methodology, and the detailed imputation procedure is available on request from the authors. The income distribution in Table 2 was obtained after supplementing the missing income values with the predicted values. Unfortunately, the dis-aggregate (actual) distribution of income is not available for Cyprus. Thus, it is not possible to make a comparison of the income distribution (after imputation) with official statistics. On the other hand, the distribution of parents' education levels (a reasonable proxy for income levels) is available from the Statistical Service of Cyprus (2014), and is presented in Table 2. The actual and sample distributions for the education levels are not too far apart, though there is a bias toward higher education levels for both parents in the sample. Finally, 45% of the participants have as a hobby a sports activity, and 51% participate in sports.

The sample used for the modeling process consisted of 4342 SP responses, corresponding to 2171 individual high school students (that is, each teenager was asked two SP questions regarding mode choice). We carefully constructed the feasible choice set of alternatives for the SP experiments for each individual as follows (only the feasible choice set of alternatives, based on the respondent's specific current situation, were presented in the SP experiment for each respondent). The alternative "PTW (motorcycle)" is considered available to a respondent only if the respondent has a PTW in her or his household, or if the respondent's close friend has a motorcycle available. The alternative "Walk" is assumed to be available to participants who live within a distance of 2.1 km from school, as this is the maximum walking distance found in the sample. The alternative "Car" is available to all students, as all the participants' households have at least one car available in their respective households and all the households have at least one driver (an adult with driving license). The alternative "Bus" is considered available to all participants (because bus stops are close to the homes of the participants and close to schools where the participants study), as is the alternative "Bicycle" (all households have at least one bicycle available).

The availability percentages of each mode and the mode shares from the stated preference responses are provided below (note that the mode shares correspond to the stated responses, not to the actual current mode choices of the teenagers):

(1) Car (availability: 100%, observed mode share: 39%).[9]
(2) PTW (availability: 25%, observed mode share: 5%).
(3) Bus (availability: 100%, observed mode share: 19%).
(4) Walk (availability: 87%, observed mode share: 16%).
(5) Bicycle (availability: 100%, observed mode share: 21%).

In this study, we consider three latent constructs: safety-consciousness, green-lifestyle and physical activity propensity. The first latent variable *Safety consciousness* reflects an individual's concern toward his/her safety. However, in the context of this paper, this latent variable refers exclusively to safety from traffic crashes. The second latent variable, *Green lifestyle*, reflects an individual's concern toward the planet Earth, the only habitat currently available to humans. The final latent variable *Physical activity propensity* reflects an individual's desire to be physically fit or the happiness they derive by playing their favorite sports.

The attitudinal and the perceptual indicators of the participants that were used for the construction of the three latent variables just discussed are presented in Table 3 (the seven point Likert scale used for the indicators is provided at the bottom of Table 3). Teenagers, as a group, appear to be neutral to the safety-related questions. They also, again as a group, appear to be appreciative of the importance of a "green lifestyle" and being physically active.

Our expectation, based on the conditions in Cyprus, is that individuals who are safety conscious will avoid cycling, walking, and using the motorcycle, because of the high rates of crashes involving non-motorists and motorcyclists in Cyprus (see Cyprus Police, 2013). Also, a higher sensitivity to safety issues increases the perceived threats of the transport network and the built environment (see Zhu and Lee, 2008 and Kamargianni and Polydoropoulou, 2014), and thus should affect the use of modes that do not have a protective protection barrier (as does a car). We also associate safety consciousness with avoidance

---

[9] Figures released by the Department of Transport show that there were 769,249 passenger vehicles registered in 2012, while the population of the country was 838,897 (Statistical Service of Cyprus, 2014). The International Road Federation (IRF) mentions that Cyprus holds the highest car ownership rate in the world (Jin, 2007). In personal conversations with the Ministry of Education of Cyprus before the launch of our survey, it was pointed out to us that the households of all students enrolled in schools in Cyprus owned at least one vehicle. This information is collected by the Ministry as part of a suite of demographic characteristics elicited when a student enrolls in school. However, these statistics are not publicly available.

Table 1
Stated preference attributes and attribute levels.

Attribute	Car	PTW	Bus	Walk	Bicycle
Travel time (in mins)	5, 6, 7, 8, 10, 12, 14, 15, 17, 25	5, 6, 7, 8, 10, 12, 14, 15, 17	8, 10, 12, 14, 15, 17, 20, 25, 30	6, 8, 10, 12, 15, 17	5, 6, 7, 8, 10, 12, 15, 17
Travel cost (in Euros)	1, 1.5, 2, 2.5, 3	1, 1.5, 2, 2.5	0, 0.5, 1.0, 1.5	–	–
Parking place	–	Available, Not available	–	–	Available, Not available
Walking time to the bus stop (in minutes)	–	–	2, 3, 5, 7, 10, 15	–	–
Bike lane (specific to bicycle)					

Bike Paths   No Bike Paths

Walkability/sidewalks

Wide Sidewalks   Narrow Sidewalks   Narrow Sidewalks with safety bars   Too Narrow Sidewalks with obstacles

Weather conditions

Table 2
Socio-economic characteristics of the sample.

Socio-economic characteristics	Categories	Percentage	Census distribution
Age	11–13	24	NA
	14–15	50	NA
	16–18	26	NA
Gender	Male	47	48%
	Female	53	52%
Household income	Less than 2000€	58	NA
	2001–3999€	27	NA
	4000–5999€	8	NA
	More than 6000€	7	NA
Parents' education status	Father – Secondary Education	68	79%
	Father – Bachelor	21	15%
	Father – Master or PhD	11	6%
	Mother – Secondary Education	59	79%
	Mother – Bachelor	29	15%
	Mother – Master or PhD	12	5%
Sports as a hobby	Positive response	45	NA
Participate in sports activity	Positive response	51	NA

NA: Not Available.

Table 3
Indicators of the latent variables.

Indicators	Mean	Std. dev.
"Safety Consciousness"		
Willing to cycle to school, but afraid of being hit by a car	3.76	2.186
Parents do not allow walking to school because of safety reasons	3.42	2.195
Feel safe when use the bus	4.05	1.859
"Green Lifestyle"		
Recycle daily	4.42	2.015
Switch off appliances in order to reduce electricity consumption	5.09	1.914
Concerned about environmental protection issues	4.60	1.875
Willing to switch to active mode of transportation in order to protect environment	4.17	2.070
Prefer bus, as it "greener" than private transport mode	4.07	2.040
"Physical Propensity"		
Willing to go to school by bike or on foot to exercise myself	4.18	2.264
Willing to substitute the motorized mode with active mode of transport in order to be fit	4.31	2.038
7-Likert scale: 1 = Completely disagree ,..., 7 = Completely agree		

of bus services, because of the relatively poor quality of bus facilities and equipment in Cyprus (European Commission – DG REGIO, 2006). In addition, this assumption is enhanced by findings from other surveys showing that one of the psycho-social benefits that individuals derive from cars is safety (Hiscock et al., 2002). Individuals feel safer when they use private motorized vehicles than when they use public or active transport (Ellaway et al., 2003).

We also postulate that the latent variable *Green Lifestyle*, which reflects a behavior that contributes to the preservation of the environment (Anable, 2005), has a significant positive impact on the propensity to use the bus mode, while negatively impacting the utilities of the car and PTW ("non-environment friendly") modes. This is because we expect pro-green individuals to engage in sustainable environmental practices. They would be more likely to avoid partaking in activities that may have a negative consequence on the environment, such as excess driving that contributes to smog and greenhouse gas emissions.

Finally, we hypothesize that the latent variable *Physical Activity Propensity* will positively affect the choice of walk and bicycle, and negatively impact the choice of motorized modes. Essentially, active transport (walk and bicycle) is a way to partake in physical activity (Bhat and Lockwood, 2004; Spissu et al., 2009).

A point to note here. In our analysis, we collected ten indicators as identified in Table 3. These very naturally fell into the three categories of safety consciousness, green lifestyle, and physical propensity. At the same time, previous studies of mode choice have identified these individual factors as potential influencers of mode choice, as just discussed. Thus, we did not undertake any exploratory factor analysis of the indicators to identify the three latent constructs, but rather based the identification of the latent constructs on intuitiveness and results from previous studies.[10] As stated by Golob, 2003: "*Theory and*

---

[10] Indeed, almost all applications of ICLV models in the transportation literature that collect just a handful of indicators use such a combination of intuitiveness, judgment, and earlier studies to identify the latent constructs (see, for example, Daly et al., 2012; Bolduc et al., 2008; de Abreu e Silva et al., 2014; La Paix et al., 2013; Temme et al., 2008). This is different from studies in psychology that collect a battery of tens (and sometimes hundreds) of indicators, and use exploratory factor analysis to identify a much fewer number of factors (or latent constructs) for use in structural equations modeling. In these studies, it is very difficult to move from the large number of indicators to the factors without an analytic variance-minimization approach.

good sense must guide model specification". However, once the three latent constructs were identified, we tested our hypotheses of the loadings of the latent constructs in the measurement equations for the indicators and for the choice model, while also ensuring sufficiency conditions for econometric identification. That is, once the three latent constructs were identified, the final specification in the measurement equations for the indicators and for the choice model was based also on statistical testing using nested predictive likelihood ratio tests and non-nested adjusted predictive likelihood ratio tests (of the type discussed in Section 4.3.).

4. Estimation results

This section presents and discusses the estimation results of the model, which was estimated using the Gauss software. The Gauss code is available at http://www.caee.utexas.edu/prof/bhat/FULL_CODES.htm.

In the structural model system that relates the latent variables to explanatory variables, we considered various age groups, gender, levels of household income, education levels of parents, and participation in sports activities as potential explanatory variables. In the school travel mode choice model, we consider additional mode-specific variables (travel times and costs), as well as bicycle and walking facility variables and the latent variables themselves. In our specifications, we tested alternative functional forms for continuous variables such as age and income, including a logarithmic effect, piecewise linear effects, and dummy variables for different ranges. Of these, the dummy variable specification for specific ranges of age and income came out to be the best specification, and is the one used. For other discrete explanatory variables, such as parental education level, we attempted the most general specification (by introducing as many dummy variables as the number of education categories minus 1) and then combined categories that were not statistically different to develop a parsimonious specification.

4.1. Latent variables – structural and measurement models

The results of the structural and measurement models are presented in Table 4. All the explanatory variables used in the structural equation models are statistically significant at the 95% confidence level.

The structural model related to the *Safety consciousness* latent variable indicates that there are no age-related differences in safety consciousness among boys, and that boys and girls above the age of 15 years have the same level of safety consciousness. However, girls at or below the age of 15 years (and especially girls at or below the age of 13 years) have a much higher safety consciousness than boys and girls above the age of 15 years. This result is in line with the results of other transport and psychological surveys showing that young girls are generally more sensitive to safety considerations, primarily because of what appears to be a mindset transfer of their parents' rather asymmetric concerns for a daughter's vulnerability to social dangers (including road safety issues) relative to that of a son (see, for example, Grow et al., 2008; Zhu and Lee, 2009, and Mitra and Buliung, 2012). Income also significantly impacts safety consciousness, with teenagers from higher income households being progressively more safety conscious than teenagers from lower income households. Although previous studies have shown that low-income children are exposed to disproportionately higher volumes of traffic than their peers from higher-income households (and so may be more worried about road safety; Green et al., 2004), the transport mode alternatives available to low income children are limited (see Green et al., 2004) and the parents of such children may also be less protective (Hiscock et al., 2002). These factors may explain the lower safety consciousness of children from low income households compared to high income households.

Green Lifestyle is affected by both the educational level of parents and the household's income level. A high educational level (Masters or PhD) both for the father and the mother affects positively a green lifestyle. Education makes individuals more concerned about environmental issues and more aware of potential threats due to climate change (Sundblad et al., 2007). What is particularly interesting in the education effects is that the mother's education level appears to matter more than that of the father's in terms of the green lifestyle perspective of children. This is not inconsistent with socio-psychological studies (for example, see Judkins and Presser, 2008) that indicate that (a) well-educated women/wives tend to be more ecologically conscious, have a more "eco-friendly" lifestyle, and act in a more sustainable way than their (even equally well-educated) husbands, and (b) mothers appear to have a more emotional connection with children in forming children's thoughts and perspectives all the way through to teenage years. The former is associated with an intensity for green lifestyle living, while the latter is associated with the intensity of perspective transfer, both of which combine to result in the higher effect of the mother's education level relative to that of the father's. The results in Table 4 also indicate that children from high income households are less likely to be "green" relative to their low income peers, though this may also be a reflection of a financially-triggered constraint in low income households of not being able to afford excesses in material consumption.

Although the third latent variable, *Physical Activity Propensity*, is widely investigated in health sciences, little work exists in the transport sector on how this variable affects mode choice. The results indicate that younger teenagers tend to have a greater propensity for physical activity, with this propensity reaching its minimum for children above the age of 15 years. This age-based reduction in physical activity propensity is consistent with earlier studies indicating that older teenagers tend to be more sedentary, and engage more in "hanging out" and "social" activities, than their younger peers (see, for example, Nelson et al., 2006). Not surprisingly, having sports as a hobby and participation in sports activities positively and

Table 4
Structural equation estimates.

Structural equation estimates		
Variable	Coefficient (α)	T-stat
Safety-conscious		
Age and gender (base is greater than 15 years old and boy)		
11–13 years old girl (Yes = 1, No = 0)	0.596	59.406
14–15 years old girl (Yes = 1, No = 0)	0.266	47.379
Household monthly income (base is less than 1000 euros)		
1000–2999 euros	−0.059	−12.042
3000–4999 euros	0.031	4.438
5000 and more euros	0.255	31.689
Green lifestyle		
Parents education status		
Father with a graduate degree	0.125	14.394
Mother with a graduate degree	0.220	25.083
Household monthly income (base is 6000 and more euros)		
Less than 2000 euros	0.014	2.454
2000–3999 euros	−0.030	−5.079
4000–5999 euros	−0.038	−5.142
Physical activity propensity		
Age (base is greater than 15 years old)		
11–13 years (Yes = 1, No = 0)	0.166	36.525
14–15 years (Yes = 1, No = 0)	0.018	4.630
Sports as a hobby (Yes = 1, No = 0)	0.178	50.030
Participates in sports activity (Yes = 1, No = 0)	0.086	24.618

Measurement equation estimates (*t*-statistics in parenthesis)		
	Constant (δ)	Factor loading (d)
Indicators of safety conscious		
Willing to cycle to school, but afraid of being hit by a car	0.426 (82.080)	0.631 (34.628)
Parents do not allow walking to school because of safety reasons	0.195 (31.160)	1.008 (33.543)
Feel safe when use the bus	0.984 (206.650)	−0.070 (−8.659)
Indicators of green lifestyle		
Recycle daily	0.239 (35.090)	0.906 (57.043)
Switch off appliances in order to reduce electricity consumption	1.007 (67.800)	1.287 (60.509)
Concerned about environmental protection issues	1.069 (64.430)	1.582 (59.002)
Willing to switch to active mode of transportation in order to protect environment	0.793 (166.360)	0.071 (5.588)
Prefer bus, as it "greener" than private transport mode	0.844 (180.000)	0.233 (19.181)
Indicators of physical activity propensity		
Willing to go to school by bike or on foot to exercise myself	0.682 (44.290)	1.330 (25.062)
Willing to substitute the motorized mode of transport with active mode of transport in order to be fit	1.174 (43.830)	1.309 (25.464)

significantly affects the propensity for physical activity, though the directionality of this relationship could, admittedly, as well be the other way around.

We also found a positive correlation between the safety-consciousness and physical activity propensity latent constructs (see Eq. (8); *t*-statistics in parenthesis). The correlation coefficient is 0.26 with a *t*-statistic of 15.51. While a clear explanation for this is elusive, one possibility is that those with a high physical activity propensity are also much more aware of potential physical safety issues, which then gets transferred to safety consciousness in the context of travel.

$$\mathbf{\Gamma} = \begin{bmatrix} 1.00\text{(fixed)} & 0.00\text{(fixed)} & \\ 0.00\text{(fixed)} & 1.00\text{(fixed)} & \\ 0.26\,(15.51) & 0.00\text{(fixed)} & 1.00\text{(fixed)} \end{bmatrix} \qquad (8)$$

The results of the measurement model are presented in the lower part of Table 4. The measurement model links the latent psychometric variables to the attitudinal and perceptual indicators (presented in Table 3), and captures the loading of each latent variable on each indicator. The parameter vectors δ and ψ (see Eq. (3)) map the scale of the underlying latent variable vector $\mathbf{z}^*$ to the observed ordinal indicators. In Table 4, the first numeric column, labeled as the "Constant", provides estimates of the elements of δ, while the second numeric column, labeled as the "Factor Loading", provides the loading of the latent variable on the indicator $\mathbf{d}$ (for instance, 0.631 is the loading of the safety consciousness latent variable on the "willingness to cycle to school, but afraid of being hit by a car"). In Table 4, we do not show the elements of the ψ vector to avoid clutter, but these are available on request from the authors (there are five thresholds for each of the ten indicators, given that the indicators are collected on a 7-point Likert scale; see Section 2.2).

All the loadings of the latent constructs on the indicators (the second numeric column of Table 4) have the expected signs. Thus, for example, the latent variable *Safety consciousness* has a positive effect on the indicator regarding parental concerns

about walking and safety, indicating the strong link between parents and their children's behavior (loading factor is 1.008, t-statistic is 33.543). The loading factor of the safety consciousness variable on the indicator "Feel safe when use the bus" is −0.070 (t-statistic = −8.659), indicating that safety conscious teenagers do not feel safe when they use the bus. This is not surprising, given the general state of bus safety and quality of the bus equipment in Cyprus, even though a modernization effort has been begun in the past few years. Other loadings in Table 4 are self-explanatory.

4.2. Mode choice model

The estimation results of the ICLV mode-to-school choice model are presented in Table 5. In arriving at the final specification, we tested a number of interaction terms between mode specific attributes and weather conditions (i.e. walking time to the bus station was interacted with sunny weather conditions); between mode specific attributes and latent constructs (i.e. walking and cycling travel times were interacted with *Physical activity propensity*); and between route specific attributes and latent constructs (i.e. separate bike paths dummy variable was interacted with *Safety consciousness*). However, none of the results turned out to be statistically significant.

Overall, the estimated values of the parameters are in agreement with prior expectations. The constants in the model capture intrinsic preferences for the modes in the population of teenagers under study, though they also serve to adjust for the presence of continuous variables (such as times and costs, and the latent variables themselves) in the modal utilities. Interestingly, in this case, the constants do seem to reflect the preference of teenagers in Cyprus for private motorized vehicles (the car mode and the PTW) and their reluctance to use the bus mode and the active transport modes (walk and bicycle). Indeed, Cyprus is a country heavily dependent on private motorized vehicles, while active transport is largely unused.

Table 5
Choice model estimation results.

Variables	Coefficient (β)	T-stat
Alternative specific constant (Car is the base category)		
Motorcycle	0.535	30.087
Public transit	−0.190	−18.457
Walk	−0.994	−94.986
Bicycle	−0.538	−54.031
Travel time (in minutes)		
Travel time by car	−0.014	−47.635
Travel time by PTW	−0.087	−45.494
Travel time by bus	−0.008	−26.939
Travel time – specific to walk	−0.009	−21.991
Travel time by bicycle	−0.032	−51.261
Travel cost (in Euro)		
Travel cost by car	−0.043	−19.243
Travel cost by PTW	−0.344	−31.474
Travel cost by bus	−0.208	−39.046
Weather conditions (*Rain is the base category*)		
Sunny weather – specific to walk	0.239	50.606
Sunny weather – specific to bicycle	0.193	41.835
Network characteristics		
Separate bicycle path – specific to bicycle (Yes = 1, No = 0)	0.155	37.353
Narrow sidewalks – specific to walk (base – "too narrow sidewalks")	0.183	33.407
Narrow sidewalks with safety bars – specific to walk (base – "too narrow sidewalks")	0.197	37.199
Wide sidewalks – specific to walk (base – "too narrow sidewalks")	0.440	87.290
Parking availability – specific to PTW (Yes = 1, No = 0)	0.079	9.519
Parking availability – specific to bicycle (Yes = 1, No = 0)	0.106	26.351
Latent variables (λ)		
Safety consciousness – specific to car	0.258	22.137
Safety consciousness – specific to bus	−0.242	−20.856
Safety consciousness – specific to walk	−0.138	−12.041
Green lifestyle – specific to car	−0.213	−10.589
Green lifestyle – specific to PTW	−0.313	−7.914
Green lifestyle – specific to bus	0.049	2.725
Physical activity propensity – specific to car	−0.230	−9.456
Physical activity propensity – specific to bus	−0.125	−5.176
Physical activity propensity – specific to walk	0.161	6.494
Physical activity propensity – specific to bicycle	0.177	7.206
Composite marginal log-likelihood value at convergence (overall ICLV model)	−450351.73	
Predictive log-likelihood value of choice model at converged values	−5943.74	
Predictive log-likelihood value of choice model at sample shares	−6383.37	
Predictive log-likelihood value of choice model at equal shares	−6988.18	

Travel times, travel costs and walking time from/to the bus station have the expected negative signs. We allowed mode-specific coefficients on times and costs because of the relatively substantial differences we noticed in some of these coefficients across modes. In terms of time, the disutility caused by an additional minute is in the same range for the car, bus, and walk modes, but higher for the PTW and bicycle modes. This may be tied to exposure issues, since the accident rates of PTW and bicycle modes per unit of exposure in Cyprus are much higher than the accident rates for other modes per unit of exposure (Cyprus Police, 2013). There are also variations in travel cost sensitivities across the modes, with the travel cost sensitivity for the car mode being the lowest. The higher travel cost sensitivity for the public transport mode relative to the car mode may be because, when the survey took place, high-school students of Cyprus had a student card that allowed them to use the bus without any charge. In our experimental design, however, we included scenarios where the students had to pay for using the bus. The substantial negative reaction of students to pay for using the bus is reflected in the high negative utility associated with cost for this mode. Weather conditions also significantly affect the mode choice behavior, with the results indicating that sunny weather positively affects walk and bicycle utilities.

In terms of networks characteristics, not surprisingly, the existence of separate bicycle paths positively and significantly affects the choice of the bicycle mode (relative to the base condition of too narrow sidewalks). Also, the provision of wider sidewalks appears to be a factor in choosing the walk mode. In particular, narrow sidewalks (with and without safety bars) increase the likelihood of walking (relative to the base condition of "too narrow sidewalks"), though the marginal increase in the likelihood because of safety bars appears to be negligible. On the other hand, the existence of wide sidewalks has a tangible and much higher impact on the choice of the walk mode relative to narrow sidewalks (with and without safety bars). Overall, the walking facility results indicate that providing for more of a "space cushion" for walking increases the probability of choosing to walk. Finally, in the category of explanatory variables, the availability of PTW and bicycle parking places at the school increases the probability of choosing these two alternatives. The latter result suggests that small investments in bicycling infrastructure can have a positive impact on the choice to use active means of transportation.

Table 5 also shows that the latent variables significantly affect mode choice. These effects are consistent with our expectations, as presented in Section 4.1.

In the choice model, we allowed for a general covariance matrix Λ across alternatives for the utilities at each choice occasion, but the estimable $\tilde{\Lambda}$ matrix that we obtained (for the difference of the error terms; see Section 3) did not reject the hypothesis of the original error terms in the choice model being independent and identically distributed (IID) with a scale of 0.5. Thus, we maintained the IID assumption for the choice error terms ε_{ti} at each choice occasion t. However, this does not mean that the overall utilities are IID across alternatives at each choice occasion, because of the presence of the stochastic latent construct vector z^*. For example, the physical activity propensity latent variable impacts both the utilities of walk and bike, engendering a positive correlation in the utilities of these two alternatives at each choice occasion.

4.3. Measurements of fit of ICLV and MNP choice models

This section presents the results of the multinomial probit (MNP) model that ignores the latent constructs, but considers observed and unobserved heterogeneity in the effects of mode choice. Unlike many earlier ICLV models that compare the ICLV model with a simple DCM without any accommodation of observed or unobserved heterogeneity, we consider both observed and unobserved heterogeneity in the MNP specification. In developing our MNP specification, we extensively tested for observed and unobserved heterogeneity. We could not find any statistically significant random (normal) distribution for any of the parameters in the MNP model; however, a few interactions (such as bicycle-path with gender and age) turned out to be significant in the MNP model.

The final MNP specification is available in an online supplement to this paper at http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/SchoolTravelICLV/MNP_Estimates.pdf, though we would like to state here that the effects of latent variables in the ICLV model did get manifested in the MNP model directly through socio-demographic variables. Thus, for example, in the MNP model, girls of age 11–15 years had a negative inclination toward walking, a reflection of safety consciousness as indicated in the structural equation component of the ICLV model. However, while some of the observed heterogeneity implicit in the ICLV model can be accommodated by including additional demographic variables in the MNP specification, the ICLV model also implicitly incorporates unobserved heterogeneity because the latent variables are stochastic. On the other hand, in our MNP specification, trying to explicitly capture unobserved heterogeneity through random intercepts and coefficients did not yield any statistically significant results.

The ICLV and MNP models may be compared based on a non-nested adjusted predictive likelihood ratio test. We are not aware of such a test being proposed and used earlier to test the ICLV model with a model that does not consider the latent constructs. To do so, we obtained the implied (predictive) log-likelihood values of the ICLV and MNP models at the convergent values (this is done because the model estimations are undertaken using a composite marginal likelihood procedure, but the implied log-likelihood values are the ones needed for the non-nested test). The implied log-likelihood values for the two models can be obtained immediately from the converged values of the parameters. The non-nested adjusted likelihood ratio test determines if the adjusted likelihood ratio indices of two non-nested models are significantly different. The adjusted likelihood ratio index can be calculated as follows:

$$\bar{\rho}^2 = 1 - \frac{\ell(\boldsymbol{\omega}) - M}{\ell(0)}$$

where $\ell(\boldsymbol{\omega})$ and $\ell(0)$ are the implied log-likelihood values at converged parameters and equal shares respectively and M is the number of parameters. To compute the $\bar{\rho}^2$ measure for the ICLV model, we go back to Eq. (7), which is the reduced form for the choice model. The number of parameters M in this reduced-form choice model corresponds to 45 (30 parameters in the choice model of Table 5, 14 parameters in the structural equations model of Table 4, and 1 parameter in the error covariance matrix of Eq. (8) in the structural equation system). The corresponding $\bar{\rho}^2$ value for the ICLV model is then 0.144. For the MNP model, the best specification includes 34 parameters in all, with the $\ell(\boldsymbol{\omega})$ value being −6070.59. Thus the $\bar{\rho}^2$ value for the MNP model is 0.127, with the difference between the $\bar{\rho}^2$ values for the ICLV and MNP models being $\tau = 0.017$. The probability that this difference could have occurred by chance is no larger than $\Phi\left\{-[-2\tau\ell(c) + (M_2 - M_1)]^{0.5}\right\}$ in the asymptotic limit. A small value of the probability of chance occurrence indicates that the difference is statistically significant and that the model with the higher value of adjusted likelihood ratio index is to be preferred. In our case, the statistic turned out to be $\Phi[-15.77]$, showing that the probability of the superior performance of the ICLV model being a chance occurrence is almost zero.

Table 6
Aggregate and disaggregate measures of fit in the estimation sample.

Alternatives	Age 11–13 years			Age 14–15 years			Age 16–19 years		
	Actual	ICLV	MNP	Actual	ICLV	MNP	Actual	ICLV	MNP
Car	42.64	33.38	21.78	37.16	34.04	20.02	40.64	32.98	22.38
Motorcycle	4.17	5.48	5.70	6.21	6.44	6.21	5.08	7.06	6.76
Bus	16.00	17.64	34.19	22.64	19.50	30.83	18.89	20.66	33.92
Walk	15.11	17.56	18.12	13.68	16.59	27.19	14.71	17.17	17.70
Bicycle	22.07	25.94	20.21	20.31	23.44	15.75	20.68	22.14	19.24
Observations	1034			2162			1146		
RMSE	–	4.72	12.50	–	2.76	10.62	–	3.84	10.71
MAPE	–	3.71	9.09	–	2.51	8.68	–	3.06	7.88
Log-likelihood at equal share	−1664.16			−3479.61			−1844.42		
Log-likelihood at convergence	−1389.57		−1434.78	−3013.54		−3038.94	−1523.33		−1598.87
# of parameters	45		34	45		34	45		34
$\bar{\rho}^2$ value	0.138		0.117	0.121		0.117	0.150		0.115
Predictive likelihood ratio	$\Phi[-8.91] \ll 0.0001$			$\Phi[-6.31] \ll 0.0001$			$\Phi[-11.84] \ll 0.0001$		

Alternatives	Income less than 2000			Income 2000–3999			Income 4000 and more		
	Actual	ICLV	MNP	Actual	ICLV	MNP	Actual	ICLV	MNP
Car	39.26	33.26	20.1	37.12	33.21	22.76	43.65	35.18	21.93
Motorcycle	5.43	6.36	6.02	5.38	6.45	6.08	5.88	6.52	7.81
Bus	21.99	19.68	35.82	19.88	19.61	33.62	12.69	17.62	16.37
Walk	14.26	17.01	21.6	13.57	17.29	21.64	15.94	16.53	28.22
Bicycle	19.06	23.69	16.47	24.06	23.43	15.90	21.83	24.16	25.67
Observations	2524								
RMSE	–	3.77	11.13	–	2.48	10.27	–	4.52	11.44
MAPE	–	3.32	8.70	–	1.92	9.00	–	3.39	8.69
Log-likelihood at equal share	−4062.22			−1866.26			−1039.70		
Log-likelihood at convergence	−3560.80		−3574.38	−1587.87		−1686.46	−797.06		−811.45
# of parameters	45		34	45		34	45		34
$\bar{\rho}^2$ value	0.112		0.111	0.125		0.078	0.190		0.187
Predictive likelihood ratio	$\Phi[-4.02] \ll 0.0001$			$\Phi[-13.65] \ll 0.0001$			$\Phi[-4.22] \ll 0.0001$		

Alternatives	Female			Male		
	Actual	ICLV	MNP	Actual	ICLV	MNP
Car	45.34	34.53	16.26	32.95	32.55	26.72
Motorcycle	3.78	6.09	5.22	6.60	6.60	7.30
Bus	19.26	18.27	35.64	21.20	20.57	28.63
Walk	12.93	16.51	25.88	15.80	17.60	18.94
Bicycle	18.69	24.60	17.00	23.45	22.68	18.41
Observations	2308			2034		
RMSE	–	5.85	16.04	–	0.94	5.10
MAPE	–	4.72	12.31	–	0.72	4.51
Log-likelihood at equal share	−3714.58			−3273.60		
Log-likelihood at convergence	−3151.74		−3231.88	−2794.60		−2840.61
# of parameters	45		34	45		34
$\bar{\rho}^2$ value	0.139		0.121	0.133		0.122
Predictive likelihood ratio	$\Phi[-12.22] \ll 0.0001$			$\Phi[-9.00] \ll 0.0001$		

To ensure that the superior data fit of the ICLV model in the full estimation sample is not simply an artifact of over-fitting, we also test the performance of the two models on different segments of the sample based on age, gender and income. Again, we are not aware of any earlier study that has used such a market segmentation approach to evaluate the effectiveness of an ICLV formulation. In particular, we compare the ICLV and MNP model fits on the segments based on (1) the disparity between the observed and predicted shares of each mode, as captured by the root mean square error (RMSE) and mean absolute percentage error (MAPE) statistics, and (2) a non-nested adjusted predictive likelihood ratio test.

To obtain the predicted modal shares for each segment from the ICLV model, the required parameters ($(\boldsymbol{\beta}, \lambda, \boldsymbol{\alpha}, \boldsymbol{\eta},$ and$\boldsymbol{\varepsilon})$) were drawn from their respective distributions and the $\boldsymbol{U}$ vector of size ($QT \times 1$) was formed using Eq. (7). Then, based on the highest value of the utility, the predicted mode choices were obtained for each choice instance of each individual belonging to the particular segment under consideration. These choices were aggregated across choice instances and individuals in the segment to obtain the segment-specific mode shares. This procedure was repeated 1000 times and the average across the 1000 times was taken as the segment-specific predicted mode shares.

Table 6 provides the actual aggregate mode shares (as a percentage) and the predicted mode shares (as a percentage) from the ICLV and MNP models for each of several segments. The corresponding number of observations in each segment, RMSE, MAPE, and non-nested likelihood predictive likelihood ratio test statistics are provided in the rows below the modal shares. The table clearly indicates the superior data fit of the ICLV model over the MNP model based on RMSE, MAPE and likelihood ratio statistics for all the segments. Indeed, the differences are remarkable. For example, for the "Age 11–13 years" segment, the RMSE (MAPE) value for the MNP model is 165% (145%) higher than that of the ICLV model. Similar substantial differences may be observed in all other segments. Finally, the non-nested adjusted predictive likelihood ratio test clearly favors the ICLV model over the MNP model for all the segments.

4.4. Analysis of the value of time

The model estimation results enable the calculation of the value of time (VOT) for the three motorized modes (car, PTW and bus). The VOT for the car mode is estimated to be 19.45 Euro per hour, while those for the PTW and bus modes are 15.17 Euro per hour and 7.50 Euro per hour, respectively. Unfortunately, we do not have a basis to assess these estimates because of the absence of earlier VOT values for teenagers. So, we compare our results with findings from surveys that focus on adults. The VOT for car is somewhat higher than (but in the same range as) the results of Wardman et al. (2012), who found out that the In-Vehicle-Time (IVT) in Greece was 15.0 Euro per hour for business purposes and 12.6 Euro per hour for commuting. It should also be pointed out that the VOT for car is not directly anticipated by teenagers, as their parents cover the car use costs, so it is comforting that Wardman's results and ours are in the same range. The VOT for PTW also appears rather high, though we do not have any basis to compare this VOT even for adults. The travel cost of PTW in Cyprus is usually covered by teenagers' pocket money and in doing so they anticipate better the travel costs of this mode. Thus, the high VOT indicates that teenagers are willing to pay a significant amount of money to drive a PTW, which reflects a desire for freedom and independent traveling. The VOT for bus is higher than that obtained in Polydoropoulou et al.'s (2014) survey of Greek adults, but lower than those found in studies in other EU countries (i.e. Switzerland: Glerum et al., 2011; UK: McNamara and Caulfield, 2013).

5. Conclusions

In the last decade, DCMs have evolved to include an explicit recognition of psychological factors to explain the decision making process of individuals. One such model formulation is the ICLV model, which has seen increasing use in the literature. But the conceptual value of ICLV models has not been adequately translated to benefits in practice because of the difficulties in model convergence and estimation, and the lengthy estimation times of these models even when convergence is achieved. Recently, Bhat and Dubey (2014) proposed an alternative formulation for these ICLV models based on a multinomial probit kernel rather than the multinomial logit kernel used in earlier ICLV studies, and combined this with a new MACML estimation technique, emphasizing the benefits of doing so.

The aim of this paper was to empirically apply and test the new MNP kernel-based ICLV formulation of Bhat and Dubey (2014) in the context of an analysis of children's mode choice to school. As expected, this new approach offered significant advantages, as the dimensionality of integration in the log-likelihood function is independent of the number of latent variables. Specifically, this approach allowed us to incorporate three latent variables with a large data sample and with 10 ordinal indicators of the latent variables, and still estimate the ICLV model without any convergence problems.

For the model estimation, a sample of 2171 teenagers (11–18 years old) was drawn from the first wave of a 2012 transport survey undertaken in Cyprus. In our analysis, we included three latent psychological factors (or constructs) to explain school mode choice: *Safety Consciousness*, *Green Lifestyle* and *Physical Activity Propensity*. The indicators for these constructs were collected in the survey on a 7-point Likert ordinal scale.

For comparison purposes, we also estimated a multinomial probit (MNP) model considering observed and unobserved heterogeneity in the effects of choice in order to have a 'fair' comparison between MNP and the ICLV model. The comparison of the two models was undertaken using the non-nested adjusted (predictive) likelihood ratio test on the full estimation

sample, as well as based on the predicted modal split and non-nested likelihood ratio on segments of the full sample. In all these comparisons, the ICLV model clearly and dominantly outperformed the MNP model.

As far as teenagers' mode choice behavior, the results indicate that transport network characteristics, such as the availability of a separate bicycle path, bicycle parking spaces, and the width of sidewalks significantly affect the choice of active transport. The latent variables entered very significantly in the mode choice model. As expected, *Safety consciousness* positively affects the choice of the car mode (escorted by an adult). *Green Lifestyle* favors the choice of bus, while *Physical activity propensity* increases the probability of choosing active transport (walk and bicycle).

The results of this paper are encouraging for the use of Bhat and Dubey's (2014) formulation of ICLV models, and it is hoped that it will promote the use of ICLV models in practice to formulate richer and more realistic behavioral representations of underlying decision processes. In addition the results are useful to researchers and authorities that deal with school transportation issues, as the model application provides important information regarding the value of investing in bicycling and walking infrastructure. It also suggests the need to improve bus and walking safety, and communicate such improvements to the public, especially to girls and women and high income households. Future research should include a more comprehensive application of the model results to promote green transport modes for school transportation and engender a more sustainable travel behavior perspective in the new generation of going-to-be adults.

Acknowledgements

References

Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C.F., Hess, S., 2011. Attitudes and value of time heterogeneity. In: Paper Presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C., January.

Alvarez-Daziano, R., Bolduc, D., 2013. Incorporating pro-environmental preferences towards green automobile technologies through a bayesian hybrid choice model. Transportmetrica A 9 (1), 74–106.

Anable, J., 2005. Complacent car addicts or aspiring environmentalists? Identifying travel behaviour segments using attitude theory. Transp. Policy 12 (1), 65–78.

Ashok, K., Dillon, W.R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. J. Market. Res. 39 (1), 31–46.

Atasoy, B., Glerum, A., Hurtubia, R., Bierlaire, M., 2010. Demand for public transport services: integrating qualitative and quantitative methods. In: Paper Presented at the 10th Swiss Transport Research Conference, Ascona, September.

Babey, S., Hastert, T., Huang, W., Brown, E.R., 2009. Sociodemographic, family, and environmental factors associated with active commuting to school among US adolescents. J. Public Health Policy 30, 203–220.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, vol. 9. MIT Press, Cambridge MA.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H.S. (Ed.), Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges. Elsevier, Amsterdam, pp. 431–470.

Bhat, C.R., 1997. Endogenous segmentation mode choice model with an application to intercity travel. Transp. Sci. 31 (1), 34–48.

Bhat, C.R., 2014. The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. Found. Trends Econom. 7 (1), 1–117.

Bhat, C.R., 2015. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Transp. Res. Part B 79, 50–77.

Bhat, C.R., Dubey, S.K., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. Transp. Res. Part B 67, 68–85.

Bhat, C.R., Lockwood, A., 2004. On distinguishing between physically active and physically passive episodes and between travel and activity episodes: an analysis of weekend recreational participation in the San Francisco bay area. Transp. Res. Part A 38 (8), 573–592.

Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid choice models with logit kernel: applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), Integrated Land-Use and Transportation Models: Behavioral Foundations, Elsevier, Oxford, pp. 275–302.

Bolduc, D., Boucher, N., Alvarez-Daziano, R., 2008. Hybrid choice modeling of new technologies for car choice in Canada. Transp. Res. Rec.: J. Transp. Res. Board 2082, 63–71.

Chataway, E.S., Kaplan, S., Nielsen, T.A., Prato, C.G., 2014. Safety perceptions and reported behavior related to cycling in mixed traffic: a comparison between Brisbane and Copenhagen. Transp. Res. Part F 23, 32–43.

Clifton, J., 2003. Independent mobility among teenagers: exploration of travel to after-school activities. Transp. Res. Rec.: J. Transp. Res. Board 1854, 74–80.

Clifton, J., Akar, G., Smith, A.L., Voorhees, C.C., 2010. Gender differences in adolescent travel to school. Exploring the links with physical activity and health. In: Women Issues in Transportation Conference Proceedings, vol. 46, issue 2, pp. 203–222.

Cyprus Police, 2013. Table of traffic accidents and victims by year from 2009–2013 <http://www.police.gov.cy/police/police.nsf/All/F0968D99D3A77811C2257CAE0040930B?OpenDocument>.

Daly, A., Hess, S., Patruni, B., Potoglou, D., Rohr, C., 2012. Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behavior. Transportation 39 (2), 267–297.

Daziano, R.A., 2015. Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. Transp. Res. Part B 76, 1–26.

de Abreu e Silva, J., Sottile, E., Cherchi, E., 2014. Effects of land use patterns on tour type choice: application of a hybrid choice model. Transp. Res. Rec.: J. Transp. Res. Board 2453, 100–108.

Ellaway, A., Macintyre, S., Hiscock, R., Kearns, A., 2003. In the driving seat: psychosocial benefits from private motor vehicle transport compared to public transport. Transp. Res. Part F 6 (3), 217–231.

European Commission, DG-REGIO, 2006. Study on strategic evaluation on transport investment priorities under structural and cohesion funds for the programming period 2007–2013. Report No. 2005.CE.16.AT.014.

Franzese, R.J., Hays, J.C., Schaffer, L.M., 2010. Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. APSA 2010 Annual Meeting Paper.

Gärling, T., Gillholm, R., Gärling, A., 1998. Reintroducing attitude theory in travel behavior research: the validity of an interactive interview procedure to predict car use. Transportation 25 (2), 129–146.

Gates, K.M., Molenaar, P., Hillary, F.G., Slobounov, S., 2011. Extended unified SEM approach for modeling event-related fMRI data. NeuroImage 54 (2), 1151–1158.

Glerum, A., Atasoy, B., Monticone, A., Bierlaire, M., 2011. Adjectives qualifying individuals' perceptions impacting on transport mode preferences. In: Presented at the 2nd International Choice Modelling Conference, Leeds, UK.

Golob, T.F., 2003. Structural equation modeling for travel behavior research. Transp. Res. Part B 37 (1), 1–25.

Gopinath, D.A., 1994. Modeling heterogeneity in discrete choice processes: application to travel demand. Ph.D. Thesis. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Green, R.S., Smorodinsky, S., Kim, J.J., McLaughlin, R., Ostro, B., 2004. Proximity of California public schools to busy roads. Environm. Health Perspect. 12, 61–66.

Grow, H., Saelens, B., Kerr, J., Durant, N., Norman, G., Sallis, J., 2008. Where are youth active? Roles of proximity, active transport, and built environment. Medicine+ Sci. Sports+ Exercise 40(12), 2071.

Heinen, E., Handy, S., 2012. Similarities in attitudes and norms and the effect on bicycle commuting: evidence from the bicycle cities Davis and Delft. Int. J. Sustain. Transp. 6 (5), 257–281.

Hess, S., Shires, J., Jopson, A., 2013. Accommodating underlying pro-environmental attitudes in a rail travel context: application of a latent variable latent class specification. Transp. Res. Part D 25, 42–48.

Hiscock, R., Macintyre, S., Ellaway, A., Kearns, A., 2002. Means of transport and ontological security: do cars provide psychosocial benefits to their users? Transp. Res. Part D 7, 119–135.

Hoshino, T., Bentler, P.M., 2013. Bias in factor score regression and a simple solution. In: De Leon, A.R., Chough, K.C. (Eds.), Analysis of Mixed Data: Methods & Applications, CRC Press, Taylor & Francis Group, Boca Raton, FL, pp. 43–61

Hunecke, M., Haustein, S., Grischkat, S., Böhler, S., 2007. Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. J. Environ. Psychol. 27 (4), 277–292.

Jin, R., 2007. Korea Ranks 40th in Car Ownership. The Korea Times (accessed 25.0315) <http://www.koreatimes.co.kr/www/news/biz/2007/12/123_15561.html>.

Johansson, M.V., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. Transp. Res. Part A 40 (6), 507–525.

Jöreskog, K.G., 1977. Factor analysis by least squares and maximum likelihood methods. In: Enslein, K., Ralston, A., Wilf, H.S. (Eds.), Statistical Methods for Digital Computers. John Wiley & Sons,, New York.

Judkins, B., Presser, L., 2008. Division of eco-friendly household labor and the marital relationship. J. Soc. Personal Relat. 25 (6), 923–941.

Kamargianni, M., 2014. Development of Hybrid Models of Teenager's Travel Behavior to School and to After-School Activities. PhD thesis, University of the Aegean.

Kamargianni, M., Polydoropoulou, A., 2013. Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. Transp. Res. Rec.: J. Transp. Res. Board 2382, 151–161.

Kamargianni, M., Polydoropoulou, A., 2014. Social networking effect on net generation's trip making behaviour: findings from a latent class model. In: Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., January.

Kamargianni, M., Ben-Akiva, M., Polydoropoulou, A., 2014. Integrating social interaction into hybrid choice models. Transportation 41 (6), 1263–1285.

Katsikatsou, M., 2013. Composite likelihood estimation for latent variable models with ordinal and continuous, or ranking variables. Doctoral thesis, Uppsala University.

La Paix, L., Bierlaire, M., Cherchi, E., Monzón, A., 2013. How urban environment affects travel behaviour: integrated choice and latent variable model for travel schedules. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State of the Art and the State of Practice, Edward Elgar Publishing, pp. 211–228.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4 (1), 103–120.

McNamara, D., Caulfield, B., 2013. Examining the impact of carbon price changes under a personalised carbon trading scheme for transport. Transp. Policy 30, 238–253.

Mitra, R., Buliung, R.N., 2012. Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. J. Transp. Geogr. 20 (1), 51–61.

Morikawa, T., 1989. Incorporating stated preference data in travel demand analysis. Doctoral Dissertation, Massachusetts Institute of Technology.

Nelson, M., Neumark-Stzainer, D., Hannan, P.J., Sirard, J.R., Story, M., 2006. Longitudinal and secular trends in physical activity and sedentary behavior during adolescence. Pediatrics 118 (6), e1627–e1634.

Olson, J.M., Zanna, M.P., 1993. Attitudes and attitude change. Annu. Rev. Psychol. 44 (1), 117–154.

Outwater, M., Castleberry, S., Shiftan, Y., Ben-Akiva, M., Zhou, Y., Kuppam, A., 2003. Attitudinal market segmentation approach to mode choice and ridership forecasting: structural equation modeling. Transp. Res. Rec.: J. Transp. Res. Board 1854, 32–42.

Polydoropoulou, A., Kamargianni, M., Tsirimpa, A., 2014. Car use addiction vs. ecological consciousness: which one prevails on mode choice behavior? In: Roorda, M., Miller, E. (Eds.), Travel Behaviour Research: Current Foundations, Future Prospects, Lulu Publishers, USA.

Rabin, M., 1998. Psychology and economics. J. Econ. Lit. 36 (1), 11–46.

Rieser-Schüssler, N., Axhausen, K.W., 2012. Investigating the influence of environmentalism and variety seeking on mode choice. In: Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., January.

Shiftan, Y., Outwater, M.L., Zhou, Y., 2008. Transit market research using structural equation modeling and attitudinal market segmentation. Transp. Policy 15 (3), 186–195.

Spissu, E., Pinjari, A.R., Bhat, C.R., Pendyala, R.M., Axhausen, K.W., 2009. An analysis of weekly out-of-home discretionary activity participation and time-use behavior. Transportation 36 (5), 483–510.

Stapleton, D.C., 1978. Analyzing political participation data with a MIMIC Model. Sociol. Methodol. 9, 52–74.

Statistical Service of Cyprus, 2014. Statistical Abstract 2013. General Statistics, Series I, Report No. 59, Republic of Cyprus <http://www.cystat.gov.cy/mof/cystat/statistics.nsf/All/8B6B6163A4A61D3BC225753E0033D015/$file/ABSTRACT-2013-EN-171214.pdf?OpenElement>.

Sundblad, E.L., Biel, A., Gärling, T., 2007. Cognitive and affective risk judgements related to climate change. J. Environ. Psychol. 27 (2), 97–106.

Temme, D., Paulssen, M., Dannewald, T., 2008. Incorporating latent variables into discrete choice models – a simultaneous estimation approach using SEM software. BuR-Bus. Res. 1 (2), 220–237.

Tyrinopoulos, Y., Antoniou, C., 2013. Factors affecting modal choice in urban mobility. Eur. Transport Res. Rev. 5 (1), 27–39.

Vij, A., Walker, J., 2014. Hybrid choice models: the identification problem. In: Hess, S., Daly, A. (Eds.), Handbook of Choice Modelling, Edward Elgar Publishing, pp. 519–564.

Walker, J.L., 2001. Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Doctoral Dissertation. Massachusetts Institute of Technology.

Wardman, M., Chintakayala, P., de Jong, G., Ferrer, D., 2012. European wide meta-analysis of values of travel time.

Weikert, M., Motl, R.W., Suh, Y., McAuley, E., Wynn, D., 2010. Accelerometry in persons with multiple sclerosis: measurement of physical activity or walking mobility? J. Neurol. Sci. 290 (1–2), 6–11.

Zhu, X., Lee, C., 2008. Walkability and safety around elementary schools: economic and ethnic disparities. Am. J. Prev. Med. 34 (4), 282–290.

Zhu, X., Lee, C., 2009. Correlates of walking to school and implications for public policies: survey results from parents of elementary school children in Austin, Texas. J. Public Health Policy 30, 177–202.