Transportation Research Part A 78 (2015) 473–488

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

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

Article history:
Received 11 October 2014
Received in revised form 10 June 2015
Accepted 22 June 2015
Available online 7 July 2015

Keywords:
Integrated Choice Latent Variable (ICLV) models
Multinomial probit (MNP)
Walking
Cycling
School transportation

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.

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

* Corresponding author. Tel.: +44 20 3108 5942.
E-mail addresses: m.kamargianni@ucl.ac.uk (M. Kamargianni), subbits@gmail.com (S. Dubey), polydor@aegean.gr (A. Polydoropoulou), bhat@mail.utexas.edu (C. Bhat).
[1] Tel.: +1 512 471 4535; fax: +1 512 475 8744.
[2] Tel.: +30 22710 35236.

http://dx.doi.org/10.1016/j.tra.2015.06.011
0965-8564/

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.