Transportation Research Part B 90 (2016) 192–217

Contents lists available at ScienceDirect

Transportation Research Part B

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

How, when and why integrated choice and latent variable models are latently useful

Akshay Vij[a,*], Joan L. Walker[b,1]

[a] Institute for Choice, University of South Australia, Level 13, 140 Arthur Street, North Sydney, NSW 2060, Australia
[b] University of California at Berkeley, 111 McLaughlin Hall, Berkeley, CA 94720-1720, United States

ARTICLE INFO

Article history:
Received 9 June 2015
Revised 22 April 2016
Accepted 26 April 2016
Available online 11 May 2016

Keywords:
Hybrid choice models
Integrated choice and latent variable models
Discrete choice models

ABSTRACT

Integrated Choice and Latent Variable (ICLV) models are an increasingly popular extension to discrete choice models that attempt explicitly to model the cognitive process underlying the formation of any choice. This study was born from the discovery that an ICLV model can in many cases be reduced to a choice model without latent variables that fits the choice data at least as well as the original ICLV model from which it was obtained. The failure of past studies to recognize this fact raised concerns about other benefits that have been claimed with regards to the framework. With the objective of addressing these concerns, this study undertakes a systematic comparison between the ICLV model and an appropriately specified reduced form choice model. We derive analytical proofs regarding the benefits of the framework and use synthetic datasets to corroborate any conclusions drawn from the analytical proofs. We find that the ICLV model can under certain conditions lead to an improvement in the analyst's ability to predict outcomes to the choice data, allow for the identification of structural relationships between observable and latent variables, correct for bias arising from omitted variables and measurement error, reduce the variance of parameter estimates, and abet practice and policy, all in ways that would not be possible using the reduced form choice model. We synthesize these findings into a general process of evaluation that can be used to assess what gains, if any, might be had from developing an ICLV model in a particular empirical context.

1. Introduction

Traditional models of disaggregate decision-making have long ignored the question of why we want what we want. Human needs have been treated as given, and attention has largely centered on the expression of these needs in terms of behavior in the marketplace. As a consequence, traditional models of disaggregate decision-making have focused on observable variables, such as product attributes, socioeconomic characteristics, market information and past experience, as determinants of choice, at the expense of the biological, psychological and sociological reasons underlying the formation of individual preferences (McFadden, 1986). This idealized representation of consumers as optimizing black boxes with predetermined wants and needs is at odds with findings from studies in the social sciences that have attempted explicitly to map the cognitive path that leads consumers from observable inputs to their observed choices in the marketplace. These studies

* Corresponding author. Tel.: +61 8 8302 0817.
E-mail addresses: vij.akshay@gmail.com (A. Vij), joanwalker@berkeley.edu (J.L. Walker).
[1] Tel.: + 1 510 642 6897.

http://dx.doi.org/10.1016/j.trb.2016.04.021
0191-2615/

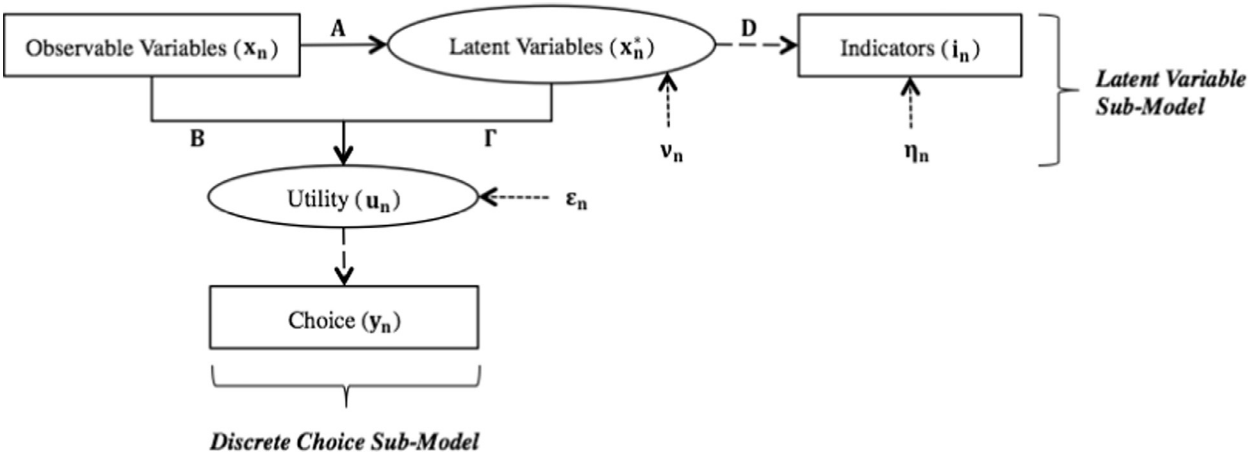

Fig. 1. The ICLV model framework (adapted from Ben-Akiva et al., 2002).

have consistently shown that latent constructs such as attitudes, norms, perceptions, affects and beliefs can often override the influence of observable variables on disaggregate behavior (see, for example, Bamberg and Schmidt, 2001; Gärling et al., 2003; Anable, 2005).

Integrated Choice and Latent Variable (ICLV) models overcome these deficiencies by allowing for the incorporation of latent behavioral constructs within the framework employed by traditional models of disaggregate decision-making. ICLV models were first proposed two-and-a-half decades ago by McFadden (1986) and Train et al. (1987) and popularized by later studies such as Ashok et al. (2002) and Ben-Akiva et al. (2002). Rapid strides in optimization techniques and computational power and the ready availability of estimation software such as Python Biogeme (Bierlaire, 2003) and Mplus (Muthén and Muthén, 2011) have since contributed to a veritable explosion in the number of studies estimating ICLV models. In the context of transportation and logistics, ICLV models have been applied to the study of travel mode choice (Paulssen et al., 2014), route choice (Bhat et al., 2015), car ownership (Daziano and Bolduc, 2013a), departure time (Thorhauge et al., 2015), freight (Bergantino et al., 2013), etc.

Much progress has been made in terms of model development and estimation (see, for example, Bhat and Dubey, 2014; Daziano, 2015 for recent methodological advances on the subject), but concerns remain regarding the value of the framework to econometricians, practitioners and policy-makers. On one hand, ICLV models appear to be powerful methods with which to enhance existing representations of decision-making. They allow for the proper integration of psychometric data within extant model frameworks and provide statistical tools with which to test complex theories of behavior, such as the Theory of Interpersonal Behavior (Triandis, 1977) and the Theory of Planned Behavior (Azjen, 1991). On the other, questions have been raised regarding the practical benefits of the framework (see, for example, Chorus and Kroesen, 2014). Does an ICLV model fit the choice data better than a simpler choice model without latent variables? Can findings from an ICLV model be used for policy analysis in ways that are not already possible using choice models without latent variables? Both sides of the debate have their proponents, but a clear verdict remains elusive.

The objective of this study is to systematically evaluate the benefits of the ICLV model framework in comparison with a more traditional choice model without latent variables, using a set of criteria based on statistical considerations and relevance to practice and policy. The study derives analytical proofs regarding the benefits, or lack thereof, of ICLV models over choice models without latent variables for each criterion and uses synthetic datasets to corroborate any conclusions drawn from the analytical proofs. Through most of the paper, we limit our attention to the form of the model that has most commonly been employed in the literature, but wherever possible, we outline how findings from the paper might be extrapolated to other variations. Our process of evaluation is general, and can be used to assess the benefits of any ICLV model framework over an appropriately specified reduced form model.

The paper is structured as follows: Section 2 describes the ICLV model framework in greater detail; Section 3 compares the framework with a choice model without latent variables in terms of its ability to predict outcomes to the choice data; Section 4 explores the usefulness of additional parameters identified by the framework; Sections 5 and 6 compare the bias and variance of parameter estimates obtained by the two model forms, respectively; Section 7 discusses the benefits of ICLV models to practice and policy; and Section 8 concludes the paper with a summary of key findings and directions for future research.

2. The ICLV model framework

Fig. 1 illustrates the ICLV model framework. In the general formulation, two components can be distinguished: a multinomial discrete choice model and a latent variable model. Each of these sub-models consists of a structural and a measurement component. In the discrete choice component, the alternatives' utilities may depend on both observed and latent attributes of the alternatives and characteristics of the decision makers. Consistent with the random utility maximization model, utility as a theoretical construct is operationalized by assuming that individuals choose the alternative with the greatest utility. The latent variable part is rather flexible in that it allows for both simultaneous relationships between the latent variables and MIMIC-type models where observed exogenous variables influence the latent variables. Such a specification enables the researcher to disentangle the direct and indirect effects of observed as well as latent variables on the alternatives' utilities. The latent variables themselves are assumed to be measured by multiple indicators, such as responses to Likert-scale survey questions.

Mathematically, the model is typically represented using the following set of four equations:

$$\mathbf{u_n} = \mathbf{B}\mathbf{x_n} + \mathbf{\Gamma}\mathbf{x_n^*} + \boldsymbol{\varepsilon}_\mathbf{n} \tag{1}$$

$$\mathbf{x_n^*} = \mathbf{A}\mathbf{x_n} + \boldsymbol{\nu}_\mathbf{n} \tag{2}$$

$$\mathbf{i_n} = \mathbf{D}\mathbf{x_n^*} + \boldsymbol{\eta}_\mathbf{n} \tag{3}$$

$$y_{nj} = \begin{cases} 1 & \text{if } u_{nj} \geq u_{nj'} \text{ for } j' \in \{1, \ldots, J\} \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where $\mathbf{u_n}$ is the $(J \times 1)$ vector of utilities of each of the J alternatives, as perceived by decision-maker n, $\mathbf{x_n}$ is the $(K \times 1)$ vector of observable explanatory variables and $\mathbf{x_n^*}$ is the $(M \times 1)$ vector of latent explanatory variables, $\mathbf{B}$ and $\mathbf{\Gamma}$ are the $(J \times K)$ and $(J \times M)$ matrices of model parameters denoting sensitivities to the observable and latent variables, respectively, and $\boldsymbol{\varepsilon}_\mathbf{n}$ is the $(J \times 1)$ vector denoting the stochastic component of the utility specification; $\mathbf{A}$ is the $(M \times K)$ matrix of model parameters denoting the structural relationship between the latent and observable variables, and $\boldsymbol{\nu}_\mathbf{n}$ is the $(M \times 1)$ vector denoting the stochastic component of that relationship; $\mathbf{i_n}$ is the $(R \times 1)$ vector of indicators used to measure the latent variables, assumed to represent deviations from the mean, $\mathbf{D}$ is the $(R \times M)$ matrix of model parameters denoting the sensitivities of the measurement indicators to the latent variables, and $\boldsymbol{\eta}_\mathbf{n}$ is the $(R \times 1)$ vector denoting the stochastic component of the measurement equation; and y_{nj} is the choice indicator, equal to one if decision-maker n chose alternative j, and zero otherwise. The stochastic components $\boldsymbol{\varepsilon}_\mathbf{n}$, $\boldsymbol{\nu}_\mathbf{n}$ and $\boldsymbol{\eta}_\mathbf{n}$ are assumed to be mutually independent. For the sake of generality, we have assumed that the vector of observable explanatory variables entering Eqs. (1) and (2) is the same. However, if the analyst wishes to use different subsets of observable explanatory variables in Eqs. (1) and (2), the appropriate columns of $\mathbf{B}$ and $\mathbf{A}$ may be constrained to be zero vectors.

Eqs. (1)–(4) are usually how ICLV models are specified in practice (see, for example, Train et al., 1987; Ben-Akiva et al., 2002; Daziano and Bolduc, 2013a). However, there are many ways in which this original specification may be modified, depending upon the empirical context. For example, Abou-Zeid et al. (2010), Hess and Beharry-Borg (2012), and Bhat et al. (2015) estimate ICLV models with interactions between the latent and observable explanatory variables in the utility specification. Yáñez et al. (2010) estimate an ICLV model where $\mathbf{\Gamma}$, or the matrix of taste parameters associated with the latent variables, is allowed to vary randomly across the sample population. Paulssen et al. (2014) estimate an ICLV model with hierarchical relationships between the latent variables, such that the structural equation for a subset of the latent variables includes other latent variables as explanatory variables. Other possible variations include non-linear specifications for the utility function, the incorporation of observable explanatory variables in the measurement equation for the latent variables, structural relationships between subsets of observable explanatory variables, etc. While these many different model specifications may be more appropriate under certain contexts, a discussion of each of these variations is beyond the scope of this study. However, findings presented here should offer general lessons for each of these variations.

For the sake of notational simplicity, we have assumed that the alternatives faced by decision-makers are the same across the sample population. Though the analytical proofs derived in subsequent sections will be for this special case, each of the proofs can be extrapolated straightforwardly to the more general case where different decision-makers may be faced with different choice sets. Different distributional assumptions about each of the stochastic variables can lead to different forms of the ICLV model. Typically, most models in the literature make one of three broadly generalizable assumptions about the vector $\boldsymbol{\varepsilon}_\mathbf{n}$: (1) each element of $\boldsymbol{\varepsilon}_\mathbf{n}$, denoted ε_{nj}, is i.i.d. Gumbel across alternatives and decision-makers with location zero and scale one, resulting in a multinomial logit kernel for the discrete choice sub-model; (2) the vector $\boldsymbol{\varepsilon}_\mathbf{n}$ is distributed normally with a mean vector of zeros and covariance matrix given by $\mathbf{\Omega_n}$, resulting in the multinomial probit kernel for the discrete choice sub-model; or (3) the vector $\boldsymbol{\varepsilon}_\mathbf{n}$ is a mixture between normally distributed and Gumbel distributed vectors, resulting in the mixed logit kernel. We will be working under the first assumption, since it is the most popular, though recent work has argued in favor of the second assumption (see, for example, Bhat and Dubey, 2014). However, equivalent proofs can be derived quite easily under the other assumptions. Assuming that ε_{nj} is i.i.d. Gumbel across alternatives and decision-makers with location zero and scale one, conditional on the latent variables, the probability that decision-maker n chooses alternative j may be derived from Eq. (4) to yield the following functional form:

$$P\left(y_{nj} = 1 \middle| \mathbf{x_n}, \mathbf{x_n^*}; \mathbf{B}, \mathbf{\Gamma}\right) = \frac{\exp\left(\boldsymbol{\beta}_{\mathbf{j*}}\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j*}}\mathbf{x_n^*}\right)}{\sum_{j'=1}^{J} \exp\left(\boldsymbol{\beta}_{\mathbf{j'*}}\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j'*}}\mathbf{x_n^*}\right)} \tag{5}$$

where $\boldsymbol{\beta}_{\mathbf{j}^*}$ and $\boldsymbol{\gamma}_{\mathbf{j}^*}$ are the $(1 \times K)$ and $(1 \times M)$ vectors corresponding to the jth rows of $\mathbf{B}$ and $\boldsymbol{\Gamma}$, respectively. Eq. (5) may be combined iteratively over alternatives to yield the following conditional probability of observing the vector of choices $\mathbf{y_n}$ for decision-maker n:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{B}, \boldsymbol{\Gamma}) = \prod_{j=1}^{J} \left[P\left(y_{nj} = 1|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{B}, \boldsymbol{\Gamma}\right)\right]^{y_{nj}} \tag{6}$$

With regards to the measurement indicators, we have assumed that the indicators represent continuous response variables, as is standard practice in the literature. This need not always be the case. Findings from this study still hold for models with measurement indicators that may best be represented as discrete response variables, but extending the proofs to include these special cases is not necessarily straightforward. For the sake of concision, we have left the derivations for these cases up to the reader. In cases where the measurement indicators are treated as continuous response variables, the vector $\boldsymbol{\eta}_{\mathbf{n}}$ is usually assumed to be distributed normally with a mean vector of zeros and covariance matrix denoted by $\boldsymbol{\Psi}$, assumed to be invariant across decision-makers (but this need not always be the case). Under these assumptions, conditional on the latent variables, the probability distribution function associated with the measurement indicators may be formulated as follows:

$$f_{\mathbf{i}}(\mathbf{i_n}|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{D}, \boldsymbol{\Psi}) = (2\pi)^{-\frac{R}{2}}|\boldsymbol{\Psi}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{i_n} - \mathbf{D}\mathbf{x_n^*})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{i_n} - \mathbf{D}\mathbf{x_n^*})\right) \tag{7}$$

where $|\boldsymbol{\Psi}|$ denotes the determinant of $\boldsymbol{\Psi}$. The latent variables can be formulated as either nonparametric or parametric random variables. A nonparametric formulation would be more akin to a latent class choice model (LCCM), and is beyond the scope of this paper. We will be limiting our attention to the case where the latent variables are represented as parametric random variables. We assume further that the latent variable can be represented using a linear in parameters formulation, as given by Eq. (2). When a linear in parameters formulation is used, the vector $\boldsymbol{\nu}_{\mathbf{n}}$ is usually assumed to be distributed normally with a mean vector of zeros and covariance matrix denoted by $\boldsymbol{\Phi}$, and the probability distribution function associated with the latent variables can be expressed as follows:

$$f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{x_n}; \mathbf{A}, \boldsymbol{\Phi}) = (2\pi)^{-\frac{M}{2}}|\boldsymbol{\Phi}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{x_n^*} - \mathbf{A}\mathbf{x_n})^{\mathrm{T}}\boldsymbol{\Phi}^{-1}(\mathbf{x_n^*} - \mathbf{A}\mathbf{x_n})\right) \tag{8}$$

Taking advantage of the conditional independence of the choice and measurement indicators and marginalizing over the distribution of the latent variables, Eqs. (6)–(8) may be combined to yield the joint unconditional probability distribution function for the choice and measurement indicators as follows:

$$f_{\mathbf{y},\mathbf{i}}(\mathbf{y_n}, \mathbf{i_n}|\mathbf{x_n}; \mathbf{B}, \boldsymbol{\Gamma}, \mathbf{D}, \boldsymbol{\Psi}, \mathbf{A}, \boldsymbol{\Phi}) = \int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{B}, \boldsymbol{\Gamma}) f_{\mathbf{i}}(\mathbf{i_n}|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{D}, \boldsymbol{\Psi}) f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{x_n}; \mathbf{A}, \boldsymbol{\Phi}) d\mathbf{x_n^*} \tag{9}$$

Eq. (9) may be combined iteratively over all decision-makers to yield the likelihood function for the sample population. The unknown model parameters $\mathbf{B}, \boldsymbol{\Gamma}, \mathbf{D}, \boldsymbol{\Psi}, \mathbf{A}$ and $\boldsymbol{\Phi}$ are estimated by maximizing the likelihood function for each of these parameters. Traditionally, maximum simulated likelihood estimation has been used to recover parameter estimates, but a number of studies in the recent past have proposed alternative methods for estimation (see, for example, Daziano and Bolduc, 2013a; Bhat and Dubey, 2014). Over subsequent sections, we examine and compare different properties of the model parameter estimates thus recovered. Since these different estimation methods are asymptotically equivalent, the results presented in this study apply regardless of whichever method is used for parameter estimation. Note that not all elements of the unknown model parameters $\mathbf{B}, \boldsymbol{\Gamma}, \mathbf{D}, \boldsymbol{\Psi}, \mathbf{A}$ and $\boldsymbol{\Phi}$ are theoretically identifiable, and appropriate constraints need to be imposed by the analyst prior to model estimation. The theoretical identification of ICLV models is a complex subject that is beyond the scope of this study, and for recent work on the topic, the reader is referred to Raveau et al. (2012), Daziano and Bolduc (2013b), Bhat and Dubey (2014), and Vij and Walker (2014). Through the remainder of the paper, we will be working under the assumption that appropriate identifying constraints have already been imposed on each of the model parameters.

3. Goodness of fit

We develop discrete choices models to better explain and predict individual choice behavior. The goodness of fit of a model is a measure of how well it explains and predicts outcomes to the dependent variable of interest. For discrete choice models, with or without latent variables, goodness of fit is a measure of the model's ability to explain and predict outcomes to the choice indicators. And most measures of fit are defined as some function of the likelihood of observing the dependent variable evaluated at the parameter estimates for either the estimation sample or a holdout sample.

For ICLV models, there are two ways in which the likelihood of observing choice outcomes may be constructed. The first approach formulates the choice probability as a function solely of the observable variables $\mathbf{x_n}$, derived by marginalizing Eq. (6) over the distribution of latent variables $\mathbf{x_n^*}$ as given by the structural equation:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}; \mathbf{B}, \boldsymbol{\Gamma}, \mathbf{A}, \boldsymbol{\Phi}) = \int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}, \mathbf{x_n^*}; \mathbf{B}, \boldsymbol{\Gamma}) f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{x_n}; \mathbf{A}, \boldsymbol{\Phi}) d\mathbf{x_n^*} \tag{10}$$

Note that the measurement indicators $\mathbf{i_n}$ are used to estimate the model parameters $\mathbf{B}, \boldsymbol{\Gamma}, \mathbf{A}$ and $\boldsymbol{\Phi}$, but they are not used to predict outcomes to the choice indicators.

The second approach formulates the choice probability as a function of both the observable variables $\mathbf{x_n}$ and the measurement indicators $\mathbf{i_n}$, and may be derived itself in one of two ways. The analyst may use Bayes theorem to derive the conditional probability from the joint probability distribution function:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n},\mathbf{i_n};\mathbf{B},\mathbf{\Gamma},\mathbf{D},\mathbf{\Psi},\mathbf{A},\mathbf{\Phi}) = \frac{f_{\mathbf{y},\mathbf{i}}(\mathbf{y_n},\mathbf{i_n}|\mathbf{x_n};\mathbf{B},\mathbf{\Gamma},\mathbf{D},\mathbf{\Psi},\mathbf{A},\mathbf{\Phi})}{\int_{\mathbf{x}^*} f_{\mathbf{i}}(\mathbf{i_n}|\mathbf{x_n},\mathbf{x_n^*};\mathbf{D},\mathbf{\Psi}) f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{x_n};\mathbf{A},\mathbf{\Phi})d\mathbf{x_n^*}} \tag{11}$$

Alternatively, the analyst may marginalize Eq. (6) over the distribution of latent variables $\mathbf{x_n^*}$ as given indirectly by the measurement equation:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n},\mathbf{i_n};\mathbf{B},\mathbf{\Gamma},\mathbf{D},\mathbf{\Psi},\mathbf{\Phi}) = \int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n},\mathbf{x_n^*};\mathbf{B},\mathbf{\Gamma}) f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{i_n};\mathbf{D},\mathbf{\Psi},\mathbf{\Phi})d\mathbf{x_n^*} \tag{12}$$

where $f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{i_n};\mathbf{D},\mathbf{\Psi},\mathbf{\Phi})$ is the probability distribution function for the latent variables, conditional on the measurement indicators $\mathbf{i_n}$ and the model parameters $\mathbf{D}$, $\mathbf{\Psi}$ and $\mathbf{\Phi}$, derived using (3), the measurement equation for the latent variables. Guttman (1955) shows that the latent variables may be expressed as the following function of the measurement indicators:

$$\mathbf{x_n^*} = \mathbf{\Lambda i_n} + (\mathbf{\Phi} - \mathbf{\Lambda D \Phi})\boldsymbol{\xi}_\mathbf{n} \tag{13}$$

where $\mathbf{\Lambda} = \mathbf{\Phi D}^{\mathrm{T}}(\mathbf{D\Phi D}^{\mathrm{T}} + \mathbf{\Psi})^{-1}$; and $\boldsymbol{\xi}_\mathbf{n}$ is an $(M \times 1)$ vector of random variables with mean equal to the $(M \times 1)$ zero vector and variance equal to the $(M \times M)$ identity matrix, such that $\mathbf{i_n}$ and $\boldsymbol{\xi}_\mathbf{n}$ are uncorrelated. For a discussion on the form of the equation in the context of factor analysis, the reader is referred to Steiger (1979), and for an analogous discussion in the context of ICLV models, the reader is referred to Train et al. (1987). Under standard normality assumptions, the conditional probability distribution function may be given by:

$$f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{i_n};\mathbf{D},\mathbf{\Psi},\mathbf{\Phi}) = (2\pi)^{-\frac{M}{2}}|\mathbf{\Phi} - \mathbf{\Lambda D\Phi}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{x_n^*} - \mathbf{\Lambda i_n})^{\mathrm{T}}(\mathbf{\Phi} - \mathbf{\Lambda D\Phi})^{-1}(\mathbf{x_n^*} - \mathbf{\Lambda i_n})\right) \tag{14}$$

Though Eq. (11) is the more direct method for calculating choice probabilities as a function of both observable explanatory variables and measurement indicators, studies that sequentially estimate ICLV models implicitly employ Eqs. (12)–(14) to calculate goodness of fit measures for the model, and we have included them in our discussion for the sake of completeness.

When evaluating the ability of the ICLV model to predict outcomes to the choice data, the decision on whether to use the measurement indicators or not is often dictated by the objectives of the study. If we are interested in using the model to predict behavior under hypothetical scenarios not contained in the observed data, and forecasts for the measurement indicators are expected to be unavailable for these scenarios, as is typically the case, then the goodness of fit measures for the model should be calculated without these indicators as well, or else any comparison with choice models without latent variables based on these measures would be a misleading exercise. This is the recommended approach in the literature (Ben-Akiva et al., 2002), and the approach most commonly used by studies (see, for example, Yáñez et al., 2010; Hess and Beharry-Borg, 2012; Daziano and Bolduc, 2013a). If we are interested in using the models to explain behavior and the process underlying it, or if forecasts for the measurement indicators are expected to be available for prediction, measurement indicators may be used to determine fit. As mentioned before, this is the approach implicitly taken by studies that sequentially estimate ICLV models, where scores for the latent variables are first determined using some variation of a factor analysis, and these scores are subsequently treated as observable (random) variables in the choice model (see, for example, Domarchi et al., 2008; Galdames et al., 2011; Maldonado-Hinarejos et al., 2014).

In general, one would expect more complex models to fit the data better than less complex models. However, as we demonstrate in this section, an ICLV model can often be reduced to a choice model without latent variables that fits the choice data at least as well as the original ICLV model framework from which it was obtained. We limit our attention initially to the first approach, where the marginal choice probability is expressed solely as a function of observable explanatory variables, as given by Eq. (10). Since the probability function contains an integral that does not have a closed form solution, it is usually approximated using the following simulator:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n};\mathbf{B},\mathbf{\Gamma},\mathbf{A},\mathbf{\Phi}) \approx \frac{1}{Q}\sum_{q=1}^{Q} f_{\mathbf{y}}\left(\mathbf{y_n}|\mathbf{x_n},\mathbf{x_{nq}^*};\mathbf{B},\mathbf{\Gamma}\right) \tag{15}$$

where $\mathbf{x_{nt}^*}$ represents the q^{th} draw from $f_{\mathbf{x}^*}(\mathbf{x_n^*}|\mathbf{x_n};\mathbf{A},\mathbf{\Phi})$ and Q is the number of draws. For more information on simulation methods, the reader is referred to Train (2009). To see that the probability of observing the choice indicators, as given by the ICLV model, can be mimicked by a choice model without latent variables, substitute (2), the structural equation of the latent variables, in (15):

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n};\mathbf{B},\mathbf{\Gamma},\mathbf{A},\mathbf{\Phi}) &\approx \frac{1}{Q}\sum_{q=1}^{Q} f_{\mathbf{y}}\left(\mathbf{y_n}|\mathbf{x_n},\mathbf{x_{nq}^*};\mathbf{B},\mathbf{\Gamma}\right) \\ &= \frac{1}{Q}\sum_{q=1}^{Q}\prod_{j=1}^{J}\left[\frac{\exp\left(\beta_{j*}x_n + \gamma_{j*}x_{nq}^*\right)}{\sum_{j'=1}^{J}\exp\left(\beta_{j'*}x_n + \gamma_{j'*}x_{nq}^*\right)}\right]^{y_{nj}} \end{aligned}$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\prod_{j=1}^{J}\left[\frac{\exp\left(\boldsymbol{\beta}_{\mathbf{j}*}\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j}*}(\mathbf{A}\mathbf{x_n} + \boldsymbol{\nu}_{\mathbf{nq}})\right)}{\sum_{j'=1}^{J}\exp\left(\boldsymbol{\beta}_{\mathbf{j'}*}\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j'}*}(\mathbf{A}\mathbf{x_n} + \boldsymbol{\nu}_{\mathbf{nq}})\right)}\right]^{y_{nj}}$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\prod_{j=1}^{J}\left[\frac{\exp\left(\left(\boldsymbol{\beta}_{\mathbf{j}*} + \boldsymbol{\gamma}_{\mathbf{j}*}\mathbf{A}\right)\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j}*}\boldsymbol{\nu}_{\mathbf{nq}}\right)}{\sum_{j'=1}^{J}\exp\left(\left(\boldsymbol{\beta}_{\mathbf{j'}*} + \boldsymbol{\gamma}_{\mathbf{j'}*}\mathbf{A}\right)\mathbf{x_n} + \boldsymbol{\gamma}_{\mathbf{j'}*}\boldsymbol{\nu}_{\mathbf{nq}}\right)}\right]^{y_{nj}} \tag{16}$$

where $\boldsymbol{\nu}_{\mathbf{nq}}$ is the q^{th} draw from $N(0, \boldsymbol{\Phi})$. At this stage, we have eliminated the vector of latent explanatory variables $\mathbf{x}^*_{\mathbf{nq}}$. What we have is the simulated probability of observing the choice indicators $\mathbf{y_n}$ as a function solely of the observable explanatory variables $\mathbf{x_n}$. Introduce the $(J \times K)$ matrix of parameters $\mathbf{T} = \mathbf{B} + \boldsymbol{\Gamma}\mathbf{A}$, such that the jth row of $\mathbf{T}$, $\boldsymbol{\tau}_{\mathbf{j}*} = \boldsymbol{\beta}_{\mathbf{j}*} + \boldsymbol{\gamma}_{\mathbf{j}*}\mathbf{A}$. Introduce the $(J \times 1)$ vector of random variables $\boldsymbol{\omega}_{\mathbf{n}} = \boldsymbol{\Gamma}\boldsymbol{\nu}_{\mathbf{n}}$, distributed normally across alternatives with a mean vector of zeros and covariance matrix $\mathbf{Z} = \boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^{\mathrm{T}}$, such that $\boldsymbol{\omega}_{\mathbf{nq}}$ is the q^{th} draw from $N(0, \mathbf{Z})$ and ω_{nqj} is the j^{th} element of $\boldsymbol{\omega}_{\mathbf{nq}}$. Substituting these new parameters in (16):

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}; \mathbf{T}, \mathbf{Z}) \approx \frac{1}{Q}\sum_{q=1}^{Q}\prod_{j=1}^{J}\left[\frac{\exp\left(\boldsymbol{\tau}_{\mathbf{j}*}\mathbf{x_n} + \omega_{nqj}\right)}{\sum_{j'=1}^{J}\exp\left(\boldsymbol{\tau}_{\mathbf{j'}*}\mathbf{x_n} + \omega_{nqj'}\right)}\right]^{y_{nj}} \tag{17}$$

But (17) is simply the simulated likelihood function for a mixed logit model without latent variables and with the following utility specification:

$$\mathbf{u_n} = \mathbf{T}\mathbf{x_n} + \boldsymbol{\omega}_{\mathbf{n}} + \boldsymbol{\varepsilon}_{\mathbf{n}} \tag{18}$$

Therefore, for any values of the model parameter vectors $\mathbf{B}, \boldsymbol{\Gamma}, \mathbf{A}$ and $\boldsymbol{\Phi}$ for the ICLV model, we can construct a mixed logit model without latent variables with parameters $\mathbf{T}$ and $\mathbf{Z}$ that yields the same distribution of the dependent variable $\mathbf{y_n}$ as a function of the vector of observable variables $\mathbf{x_n}$:

$$f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}; \mathbf{B}, \boldsymbol{\Gamma}, \mathbf{A}, \boldsymbol{\Phi}) \approx f_{\mathbf{y}}(\mathbf{y_n}|\mathbf{x_n}; \mathbf{T}, \mathbf{Z}) \tag{19}$$

Eq. (19) can be construed as a specific example of the more general result given by McFadden and Train (2000) that states that a mixed logit model can approximate to any arbitrary accuracy choice probabilities given by any other model derived from the theory of random utility maximization. In deriving (19), we have made a number of assumptions about the discrete choice and latent variable sub-models. In cases where the discrete choice sub-model has a mixed logit or multinomial probit kernel, the corresponding reduced form choice model without latent variables will be a mixed logit or multinomial probit model, respectively. In cases where the structural equation of the latent variable model has a non-linear form, the resulting utility specification for the reduced form model will tend to have a similar non-linear form. In cases where the latent variables are interacted with observable variables, the resulting utility specification for the reduced form model will have analogous interactions between different sets of observable variables. Analogous proofs can be derived for each of these cases following the method described here for the model given by Eqs. (1)–(4). And the proof applies to choice probabilities for the estimation sample, the holdout sample, and any hypothetical scenario that the analyst may wish to evaluate. For any ICLV model used to predict outcomes to the choice indicators as a function solely of observable explanatory variables (where the model is estimated using both choice and measurement indicators, but the measurement indicators are not used during prediction), we can construct a choice model without latent variables that predicts the same exact probability distribution across outcomes.

What happens when the choice probabilities are expressed as a function of both observable explanatory variables and measurement indicators? In cases where the observable explanatory variables are good predictors of the latent variables, the inclusion of measurement indicators does not offer any new information with regards to the choice outcomes, and the use of the ICLV model framework should not result in an improvement in fit. In practice, ICLV models often suffer from weak structural equations where the observable explanatory variables are poor predictors of the latent variables, and the latent variables are, in a sense, truly latent. In such cases, the inclusion of measurement indicators may result in an improvement in fit. One could argue that the measurement indicators could be included directly in the choice model without latent variables, but such an approach has been criticized on the grounds of endogeneity and causal misrepresentation (Ben-Akiva et al., 2002). However, any goodness of fit measure calculated using the measurement indicators is meaningful only if the analyst is interested in using the model to explain behavior, or the analyst is interested in using the model to predict behavior and forecasts for the measurement indicators are available for prediction, but the measure does not offer insight on the model's ability to predict choice outcomes in cases where forecasts for measurement indicators are expected to be unavailable, as is usually the case.

3.1. Monte Carlo experiment I

We present corroborating evidence for (19) using synthetic data generated through a Monte Carlo experiment. A Monte Carlo experiment is especially useful because the true parameters underlying the data generating process are known, and the credibility of the analytical proof can be evaluated under a wide variety of conditions, leading to more generalizable results that are not specific to any one dataset. For the purpose of the experiment, we construct a hypothetical model of bicycle

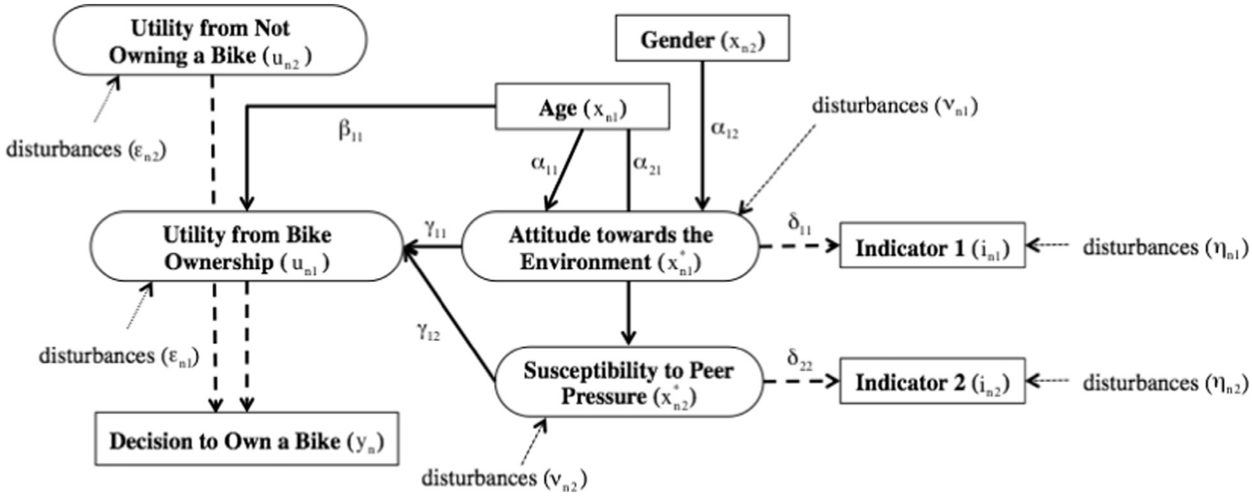

Fig. 2. A hypothetical model of bicycle ownership.

ownership, as illustrated in Fig. 2. The utility of owning a bicycle is hypothesized to be some function of an individual's age, her level of environmentalism and her susceptibility to peer influence. The latter two variables are assumed to be latent variables that can be measured through responses to, say, Likert-scale questions. For example, level of environmentalism may be measured by asking the individual her degree of agreement or disagreement with the statement, "If things continue on their present course, we will soon experience a major environmental catastrophe (Kirk, 2010)." Similarly, susceptibility to peer influence may be measured by asking the individual to what degree the following statement applies to her, "I go along with my friends just to keep them happy (Steinberg and Monahan, 2007)." Environmentalism is hypothesized to be a function itself of the individual's age and gender, and susceptibility to peer influence is hypothesized to be a function of the individual's age alone. The model may be summarized mathematically by the following set of six equations:

$$u_{n1} = \beta_{11}x_{n1} + \beta_{13} + \gamma_{11}x^*_{n1} + \gamma_{12}x^*_{n2} + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim \mathrm{GEV}(0, 1, 0) \tag{20}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \mathrm{GEV}(0, 1, 0) \tag{21}$$

$$x^*_{n1} = \alpha_{11}x_{n1} + \alpha_{12}x_{n2} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{22}$$

$$x^*_{n2} = \alpha_{21}x_{n1} + \alpha_{23} + \nu_{n2}, \quad \nu_{n2} \sim N(0, \phi_{22}) \tag{23}$$

$$i_{n1} = x^*_{n1} + \eta_{n1}, \quad \eta_{n1} \sim N(0, 1) \tag{24}$$

$$i_{n2} = x^*_{n2} + \eta_{n2}, \quad \eta_{n2} \sim N(0, 1) \tag{25}$$

where GEV(0, 1, 0) denotes the generalized extreme value distribution with location and shape parameters equal to zero and scale parameter equal to one; and N(0, 1) denotes a normal distribution with mean zero and variance one. Consistent with practices in the literature (cf. Ben-Akiva et al., 2002), we have selected a multinomial logit kernel for the discrete choice sub-model, we have assumed that the covariance matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are diagonal matrices, and we have set the location and scale for each of the latent variables through the measurement equations. Synthetic datasets are generated for the proposed model framework as follows: age (x_{n1}) is specified as an ordered categorical variable having a discrete uniform distribution between 0 and 5 where each number is understood to denote a particular age group, and the higher the number the greater the range of ages that belong to that group; and gender (x_{n2}) is specified as a Bernoulli random variable with mean 0.5. Following the methodology proposed by Williams and Ortúzar (1982) and the approach outlined by Raveau et al. (2010), values for each element of the model parameter matrices $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{D}$, $\mathbf{\Psi}$, $\mathbf{A}$ and $\mathbf{\Phi}$ are chosen such that they satisfy three conditions. First, the model should be theoretically identifiable. This is achieved by imposing appropriate constraints to fix the location and scale of each of the latent variables. Second, the part-worth utilities of each of the explanatory variables, as represented by the product between that variable and the corresponding parameter, should be comparable in terms of magnitude. If this is not the case, one of the attributes could potentially dominate the utility function, and it may be hard to empirically isolate the effect of other variables. And third, the scale of the model is set such that the error rate for the data is roughly 25%, i.e. one in four simulated decision-makers change their choice because of the stochastic component, thereby ensuring that the decision-making process is neither completely deterministic nor completely stochastic. 100 datasets each are generated for

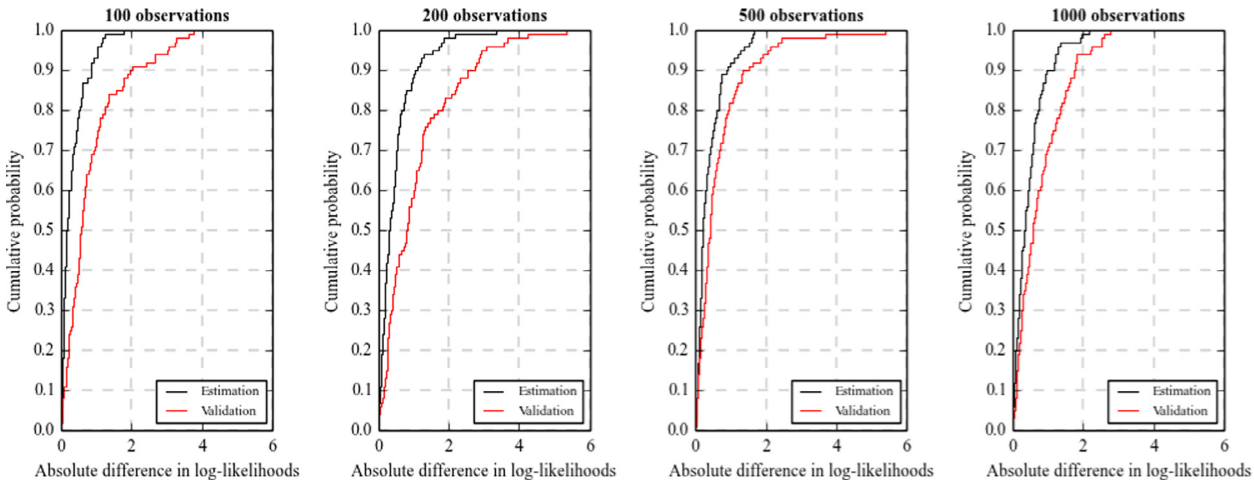

Fig. 3. A plot of the cumulative distribution function over 100 datasets of the absolute difference in the log-likelihood of choice outcomes at convergence, as given by the ICLV model of (20)–(25) and the mixed logit model of (26) and (27) for the estimation and validation samples, where the log-likelihood for the ICLV models is calculated using Eq. (10).

100, 200, 500 and 1000 pseudo-observed decision-makers hypothesized to behave according to the decision-making process described above, resulting in a total of 400 datasets.

Substituting (22) and (23) in (20) and (21), we get the following form for the reduced form mixed logit model without latent variables:

$$u_{n1} = \tau_{11}x_{n1} + \tau_{12}x_{n2} + \tau_{13} + \omega_{n1} + \varepsilon_{n1}, \quad \omega_{n1} \sim N(0, \zeta_{11}), \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{26}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{27}$$

For each of the 400 datasets, we estimate an ICLV model with the specification given by (20)–(25) and a mixed logit model with the specification given by (26) and (27).[1] All models are estimated using maximum simulated likelihood estimation with the software package Python Biogeme (Bierlaire, 2003) using pseudo-random draws. Each pair of ICLV and ML models estimated using one of the 100 datasets corresponding to a particular number of observations is subsequently used to predict outcomes for observations belonging to a different dataset with the same number of observations. The simulated log-likelihood for each of these datasets serves as a measure of either model's forecasting power, or the ability to predict outcomes to out-of-sample data.

Fig. 3 plots the cumulative distribution function of the absolute difference in the value of the log-likelihood function at convergence between the two models for the estimation and validation samples with 100, 200, 500 and 1000 observations (where the log-likelihood for the ICLV model is calculated as a function solely of observable explanatory variables). Though the data was generated using the ICLV model, the mixed logit model can be seen to fit the data just as well. For both the estimation and validation samples corresponding to a particular number of observations, we test the null hypothesis that the absolute difference in the value of the log-likelihood function at convergence between the two models is zero. The corresponding t-statistic for the one-tailed test for the estimation samples is 0.91, 0.92, 0.95 and 1.04 for each of the four sets of 100 datasets corresponding to 100, 200, 500 and 1000 observations, respectively. The analogous t-statistic for the validation samples is 0.99, 1.05, 0.82 and 1.15, respectively. In each case, the t-statistic is less than 1.282, the corresponding threshold value for a one-tailed test with a significance level of 10% (i.e. the p-value is greater than 0.10). The log-likelihoods for the two models are not identical because of simulation noise. Maximum simulated likelihood is consistent if the number of draws Q rises faster than the square root of the sample size $\sqrt{N}$. Since the number of draws is held constant across the datasets, the difference in log-likelihoods, as measured by the t-statistic, can be seen to increase with sample size, with the exception of the validation sample for 500 datasets. In general, the difference in log-likelihoods is greater for the validation samples, and the cumulative distribution function for the validation samples is to the right of the corresponding distribution for the estimation samples across each of the four sets of datasets. To conclude, we fail to reject the null hypothesis that

[1] The variance of the error component ω_{n1} for the mixed logit model is theoretically unidentified. There are two ways in which the problem can be addressed. The combined error term $\omega_{n1} + \varepsilon_{n1}$ could be approximated by a single random variable with an Extreme Value distribution, as argued for example by Guevara (2010, Chap. 3), resulting in a logit specification for the reduced form model. The drawback to the approach is that the scale of the utilities of the logit model thus estimated is different from the scale of the utilities of the true model, and the estimated reduced form parameters cannot be compared directly to the true values. The alternative is to keep ω_{n1} in the reduced form model, but constrain ζ_{11} to be equal to some constant, in our case the true value for the parameter. The benefit of the approach is that the scale of the utilities is the same for the estimated and true models, and the estimated reduced form parameters can be compared directly to the true values. This is the approach that we employ throughout this paper.

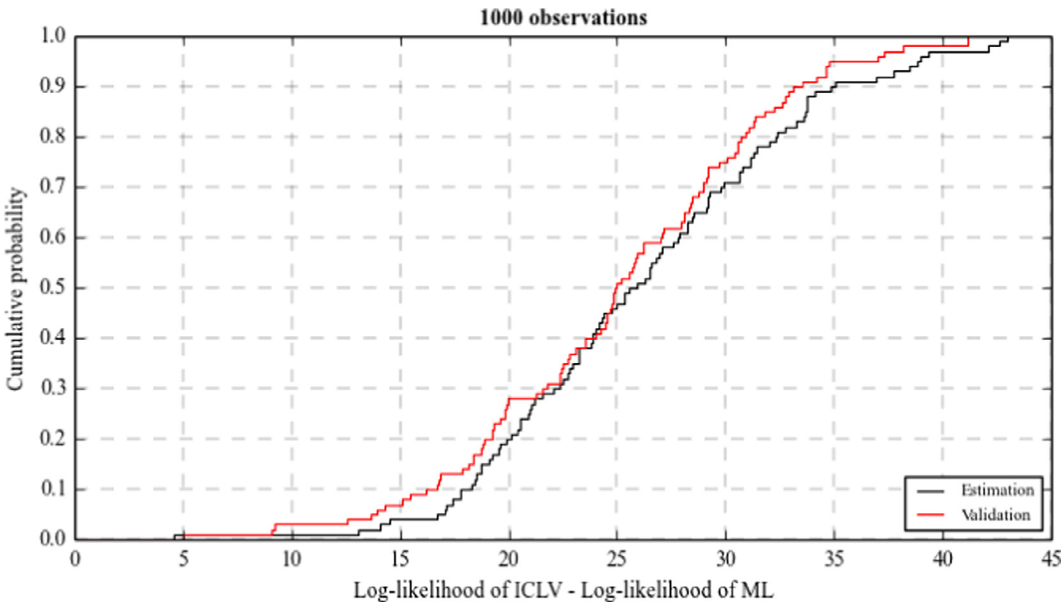

Fig. 4. A plot of the cumulative distribution function over 100 datasets, each with 1000 observations, of the difference in log-likelihood at convergence between the ICLV model given by (20)–(25) and the sub-optimal mixed logit (ML) model given by (28) and (29) for the estimation and validation samples, where the log-likelihood for the ICLV models is calculated using Eq. (10).

ICLV models do not result in an improvement in fit with respect to the choice data over simpler models without latent variables.

3.2. Monte Carlo experiment II

Despite the findings presented in previous subsections, a number of studies in the literature have used (10) to compare ICLV models to choice models without latent variables and concluded that the former provide a better fit to the choice data (see, for example, Ben-Akiva et al., 2002; Yáñez et al., 2010; Hess and Beharry-Bong, 2012; Daly et al., 2012; Kamargianni and Polydoropoulou, 2013). In each of these cases, the conclusion can be attributed to comparisons with choice models without latent variables that failed additionally to include observable variables otherwise included in the ICLV model through the structural component of the latent variable sub-model. In the context of the hypothetical model of bicycle ownership presented in Section 3.2, this is equivalent to comparing the ICLV model to a mixed logit model that does not include age or gender as explanatory observable variables, or the random error term (resulting in a multinomial logit model). For example, consider the following specification for a mixed logit model where age has not been included as an explanatory variable:

$$u_{n1} = \tau_{12}x_{n2} + \tau_{13} + \omega_{n1} + \varepsilon_{n1}, \quad \omega_{n1} \sim N(0, \zeta_{11}), \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{28}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{29}$$

Fig. 4 plots the cumulative distribution function of the difference in the value of the log-likelihood function at convergence between the ICLV model specification given by Eqs. (20)–(25), calculated using Eq. (10), and the mixed logit model given by Eqs. (28) and (29) for the 100 datasets generated in Section 3.1 with 1000 observations each for both the estimation and validation samples (the results were more or less identical across the four sets of 100 datasets with 100, 200, 500 and 1000 observations). As is apparent, the mixed logit model fits much worse than the ICLV model. For both the estimation and validation samples, we test the null hypothesis that the difference in the value of the log-likelihood function at convergence between the two models is zero. The *t*-statistic for the estimation and validation samples is 3.72 and 3.62, respectively, and in both cases we accept the alternative hypothesis that the log-likelihood at convergence for the ICLV model is greater than the log-likelihood at convergence for the ML model with a confidence of 99.9%. In such cases, the improvement in fit does not result from the inclusion of latent variables to the ICLV model, but rather from the omission of observable variables from the choice model without latent variables.

Some studies have justified omission on the grounds of statistical significance: the omitted variables when included in the ICLV model through the structural component of the latent variable sub-model were found to be statistically significant, but when these same variables were included in the choice model without latent variables through the utility specification they were found to be statistically insignificant (see, for example, Scagnolari et al., 2015). Therefore, it has been argued, these variables were retained in the ICLV model but omitted from the choice model without latent variables. We

Table 1
Estimation results for the ICLV model given by Eqs. (19)–(24), the reduced form mixed logit model given by Eqs. (26) and (27), and the sub-optimal mixed logit model given by Eqs. (28) and (29).

Parameter	True value	ICLV		Reduced form mixed logit		Sub-optimal mixed logit	
		Est.	*p*-value	Est.	*p*-value	Est.	*p*-value
Utility specification of recycling							
Constant	−3.00	−3.20	0.02	−0.14	0.80	−0.46	0.21
Attitude towards environment	0.90	0.93	0.01	–	–	–	–
Susceptibility to peer pressure	0.70	0.84	0.06	–	–	–	–
Age	−0.83	−1.07	0.02	−0.12	0.44	–	–
Gender	–	–	–	0.99	0.06	1.02	0.05
Structural equation of attitude towards environment							
Age	0.30	0.26	0.00	–	–	–	–
Gender	1.00	1.19	0.00	–	–	–	–
Structural equation of susceptibility to peer pressure							
Age	0.80	0.86	0.00	–	–	–	–
Log-likelihood of the choice model			**−66.96**		**−66.99**		**−67.29**

reason that this argument is flawed. The difference in fit cannot be significant if the omitted variable was originally insignificant in the choice model without latent variables. Since the only difference between the two models is the inclusion of the variable in the ICLV model, the likelihood ratio test between the two models is equivalent to a t-test on the parameter associated with the omitted variable. Therefore, if the parameter was found to be insignificant in the choice model without latent variables, then the difference in fit as indicated by the likelihood ratio test should not be significant either.

However, ignoring goodness of fit for now, how might such a situation arise at all and what is an appropriate response? In order to address these questions, we return to the hypothetical model of bicycle ownership. Age is posited to influence the decision to own a bicycle through its indirect influence on level of environmentalism and susceptibility to peer pressure, and through some residual direct influence on the utility of bicycle ownership itself. For certain values of the model parameters, it is plausible that the three factors might negate each other and the cumulative effect of age on bicycle ownership is small (in fact, this is true for any set of model parameters that satisfy the relationship $\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21} \to 0$).

To show that this can happen, we synthesize a dataset with 100 observations using the ICLV model specification given by (20)–(25) and the parameter values listed in Table 1 (the observable variables age and gender are simulated using the same distributions as the Monte Carlo experiment in Section 3.1). We use this dataset to recover parameter estimates for the original ICLV model specification, the reduced form mixed logit model given by (26) and (27), and the mixed logit model given by (28) and (29). Note that each of the three parameters associated with age in the ICLV model is statistically significant at the 5% level, but the solitary parameter associated with age in the reduced form mixed logit model has a *p*-value of 0.44. However, as asserted in the previous paragraph, omitting age from the mixed logit model does not result in a statistically significant loss in fit. In any case, the appropriate response in such a situation is not to omit age from the choice model without latent variables (and risk arriving at incorrect conclusions based on an unfair comparison between the ICLV model and the choice model without latent variables), but to discuss possible hypotheses for why the associated parameters might be significant in the ICLV model but insignificant in the choice model without latent variables.

3.3. Monte Carlo experiment III

In some situations, the analyst might find that the ICLV model actually performs worse than the reduced form choice model without latent variables (see, for example, Kløjgaard and Hess, 2014). Different ICLV model specifications can result in the same reduced form choice model without latent variables. For example, in the case of the hypothetical model of bicycle ownership, consider an alternative ICLV model where the utility of owning a bicycle is hypothesized to be a function solely of a decision-maker's attitude towards the environment, which in turn is hypothesized to be a function of both age and gender and is measured using the same indicator as before. The model specification is as follows:

$$u_{n1} = \beta_{13} + \gamma_{11}x^*_{n1} + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim \mathrm{GEV}(0, 1, 0) \tag{30}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \mathrm{GEV}(0, 1, 0) \tag{31}$$

$$x^*_{n1} = \alpha_{11}x_{n1} + \alpha_{12}x_{n2} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{32}$$

$$i_{n1} = x^*_{n1} + \eta_{n1}, \quad \eta_{n1} \sim N(0, 1) \tag{33}$$

The reader should check that substituting (32) in (30) and (31) yields the same reduced form mixed logit model as that given by (26) and (27), even though the true underlying ICLV model specification is different in the two cases. The number of parameters in the choice sub-model of the ICLV model given by (30)–(33) are two (β_{13} and γ_{11}) whereas the number of

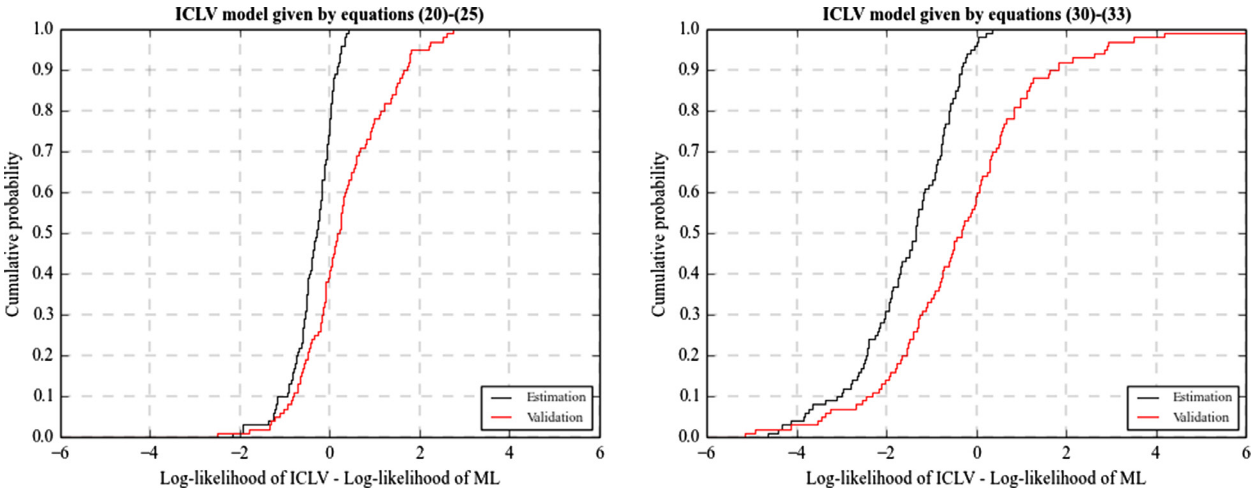

Fig. 5. A plot of the cumulative distribution function of the difference in log-likelihood at convergence for 100 datasets, each with 1000 observations, between (a) the ICLV model given by (20)–(25) and the reduced form mixed logit (ML) model given by (26) and (27); and (b) the ICLV model given by (30)–(33) and the reduced form mixed logit (ML) model given by (26) and (27), where the log-likelihood for the ICLV models is calculated using Eq. (10).

parameters in the reduced form mixed logit model are three (τ_{11}, τ_{12} and τ_{13}). In other words, the choice sub-model of the ICLV model is in fact a restricted version of the reduced form mixed logit model, and the log-likelihood at convergence for the ICLV model can at best be equal to that of the mixed logit model.

Fig. 5 plots the cumulative distribution function of the difference in the value of the log-likelihood function at convergence between the ICLV model specification given by (30)–(33), and the mixed logit model given by (26) and (27) for the 100 datasets generated in Section 3.1 corresponding to 1000 observations each (the plots were similar for 100, 200 and 500 observations, and are not included to avoid repetition). As a point of reference, we also plot the cumulative distribution function of the difference in the value of the log-likelihood function at convergence between the ICLV model specification given by (20)–(25) and the mixed logit model given by (26) and (27) for the same 100 datasets. In both cases, the log-likelihood function for the ICLV model was calculated using Eq. (10). The plot on the right shows that the ICLV model given by (30)–(33) fits the estimation sample systematically worse than the reduced form mixed logit model, and unlike the plot on the left the difference in log-likelihoods, though small, cannot be attributed to simulation noise alone. Interestingly, the same cannot be said about the validation sample, where the ICLV model given by (30)–(33) and the reduced form mixed logit model seem to perform equally well. For both the estimation and validation samples, we test the null hypothesis that the difference in the value of the log-likelihood function at convergence between the two models is zero, against the alternative hypothesis that the log-likelihood for the ICLV model is less than that for the ML model. The *t*-statistic for the estimation and validation samples is −1.41 and −0.18, respectively. In the case of the estimation sample, we accept the alternative hypothesis with a confidence of 90%, but in the case of the validation sample, we fail to reject the null hypothesis. It appears that, at least for the example constructed here, the difference in fit does not translate into poorer out-of-sample predictions.

The difference in fit for the estimation sample can be addressed by including either age or gender as a separate explanatory variable in the choice sub-model of the ICLV model through either Eq. (30) or (31), thereby increasing the number of parameters in the choice sub-model of the ICLV model to three as well. Such an anomaly presents itself when some observable variables included in the ICLV model through the structural component of the latent variable sub-model are omitted from the utility specification of the choice sub-model. Some studies have argued that a loss in fit occurs because the likelihood function of the ICLV model is fitting parameters both to the choice indicators $\mathbf{y_n}$ and the measurement indicators $\mathbf{i_n}$, as opposed to just the choice indicators $\mathbf{y_n}$ in the case of the reduced form mixed logit model. This can be construed as an alternative explanation to the one provided here. The latent variables help introduce correlation between the choice and measurement indicators, but the correlation thus imposed may lead to a loss in the ability of the model to predict outcomes to the choice indicators. The question then is: what is gained from introducing the correlation?

3.4. Monte Carlo experiment IV

Studies that have sequentially estimated ICLV models have usually reported improvements in the model's ability to predict outcomes to the choice data (e.g. Popuri et al., 2011; Córdoba and Jaramillo, 2012; Maldonado-Hinarejos et al., 2014). The sequential approach comprises four steps. First, $\mathbf{D}$ and $\boldsymbol{\Psi}$, the parameters corresponding to the measurement equation of the latent variables, are estimated using some form of factor analysis. Next, the probability distribution function of the latent variables, conditional on the measurement indicators, is constructed using Eq. (14) under some assumptions about $\boldsymbol{\Phi}$, the covariance matrix of the latent variables. Third, the estimated distribution is regressed on observable explanatory variables,

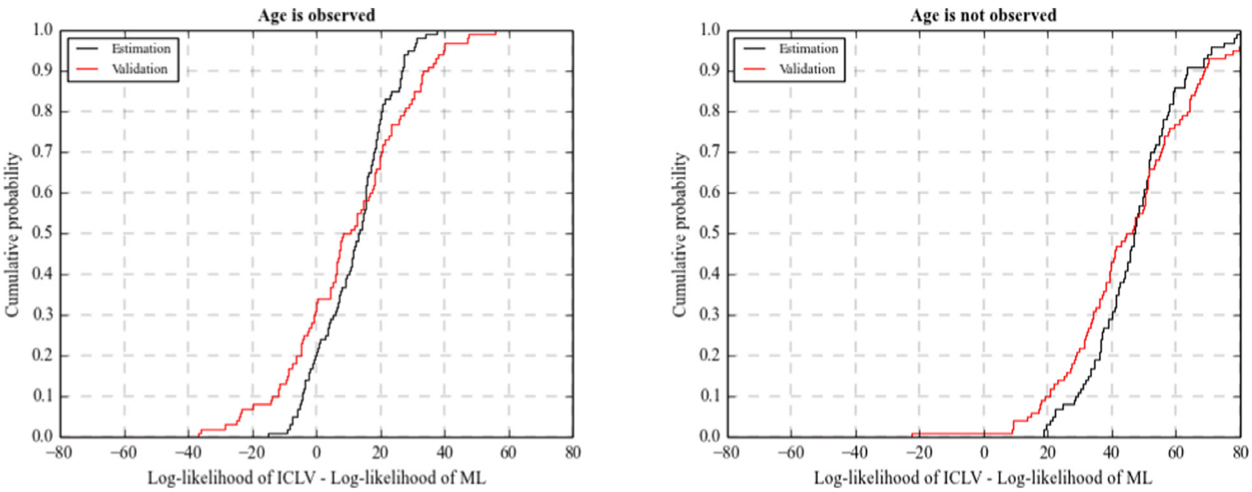

Fig. 6. A plot of the cumulative distribution function of the difference in log-likelihood at convergence for 100 datasets, each with 1000 observations, between (a) the ICLV model given by (20)–(25) and the reduced form mixed logit (ML) model given by (26) and (27); and (b) the ICLV model given by Eqs. (34)–(39) and the reduced form mixed logit (ML) model given by (28) and (29), where the log-likelihood for the ICLV models is calculated using Eq. (11).

resulting in estimates for **A**, the parameters corresponding to the structural equation of the latent variables (though most studies tend to skip this step). And finally, the choice probabilities are marginalized over the estimated distribution of the latent variables, and the resulting marginal likelihood function, conditional on the observable explanatory variables and the measurement indicators, as given by Eq. (12), is maximized to obtain estimates for **B** and **Γ**, the parameters corresponding to the choice model.[2]

The sequential approach typically evaluates the likelihood of the choice data as a function of both the observable variables and the measurement indicators, while the simultaneous approach typically evaluates the likelihood of the choice data as a function solely of the observable variables. Note that this is how both approaches are usually implemented in practice, but there is nothing about either approach that precludes them from using one or the other method to evaluate the likelihood function at convergence for the choice data. As mentioned previously, if the observable explanatory variables are good predictors of the latent variables, then the two methods should, on average, yield the same value, regardless of the estimation approach. In practice, ICLV models often suffer from weak structural equations where the observable explanatory variables are not good predictors of the latent variables, and the latent variables are, to an extent, truly latent. As mentioned before, in such cases the use of measurement indicators can indeed result in an improvement in fit over a choice model without latent variables.

For the sake of illustration, consider the datasets generated in Section 3.1 using the ICLV model specification given by (20)–(25). How does an ICLV model with the same specification estimated using this data compare to the reduced form mixed logit model given by (26) and (27)? By design, the observable explanatory variables are good predictors of the latent variables. Therefore, the log-likelihood for the ICLV model calculated as a function solely of the observable explanatory variables should be approximately equal to the log-likelihood calculated using both the observable explanatory variables and the measurement indicators. But we showed in Section 3.1 that the former is asymptotically equal to the log-likelihood for the mixed logit model. Therefore, the same should apply to the latter, and the use of the measurement indicators should not result in a statistically significant improvement in fit over a choice model without latent variables.

The left-hand plot of Fig. 6 shows the cumulative distribution function of the difference in the log-likelihood of the ICLV model given by (20)–(25), calculated using (11), and the corresponding reduced form mixed logit model given by (26) and (27) for the 100 datasets generated in Section 3.1 with 1000 observations each. Though the difference in log-likelihoods has a large spread for both the estimation and validation samples, the *t*-statistic for the two samples is 1.03 and 0.53, respectively, and for both we fail to reject the null hypothesis that the difference is zero. The spread of the distribution is a function of the variance of the stochastic components of the structural and measurement equations. But on average, the two models fit the data equally well, and the use of the measurement equation does not result in a statistically significant improvement in the model's ability to predict choice outcomes. Since the observable explanatory variables are good predictors of the latent variables, and the latent variables are measured using the measurement indicators, the observable explanatory variables

[2] Most studies use principal components analysis in the first step (e.g. Banerjee et al., 2012; Popuri et al., 2011; Wang and Chen, 2012), which assumes that $R = M$, or $\mathbf{i_n}$ and $\mathbf{x_n^*}$ are of the same dimension; $\boldsymbol{\Psi}$, the covariance matrix of the measurement indicators, is a zero matrix; and $\boldsymbol{\Phi}$ is the $(M \times M)$ identity matrix. Under these assumptions, the probability distribution function of the latent variables, as given by (13), is degenerate, and the latent variable scores are given by $\hat{\mathbf{D}}^{-1}\mathbf{i_n}$, where $\hat{\mathbf{D}}$ is the estimate for $\mathbf{D}$. The benefit of using principal components analysis is that the likelihood function given by (12) no longer requires the evaluation of an integral, and the estimated scores can be treated as observable variables. The drawback to the method is that it requires the analyst to make a number of unnecessarily restrictive assumptions about model structure that may not hold true.

are able to explain most of the variance in the measurement indicators. And therefore, the inclusion of the measurement indicators does not offer new information with regard to the choice indicators.

For the same datasets, what if we did not have access to the age of each individual in the sample population during model estimation, and we estimated an ICLV model with the following specification:

$$u_{n1} = \beta_{13} + \gamma_{11}x^*_{n1} + \gamma_{12}x^*_{n2} + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{34}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{35}$$

$$x^*_{n1} = \alpha_{12}x_{n2} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{36}$$

$$x^*_{n2} = \alpha_{23} + \nu_{n2}, \quad \nu_{n2} \sim N(0, \phi_{22}) \tag{37}$$

$$i_{n1} = x^*_{n1} + \eta_{n1}, \quad \eta_{n1} \sim N(0, 1) \tag{38}$$

$$i_{n2} = x^*_{n2} + \eta_{n2}, \quad \eta_{n2} \sim N(0, 1) \tag{39}$$

where age has been omitted from each of the equations. How would this model compare to the analogous mixed logit model given by (28) and (29), where age is not included as an explanatory variable as well? In the case of the ICLV model, the structural equations for the latent variables are quite different from the true structural equations underlying the data generating process, but the measurement equations are identical. Therefore, the measurement indicators are likely to bring in information about the choice indicators that is not already contained in the observable explanatory variables, and the log-likelihood for the ICLV model calculated as a function of both the observable explanatory variables and the measurement indicators is likely to be greater than the log-likelihood calculated solely as a function of the observable explanatory variables. But we have established through (19) and earlier experiments that the latter should be the same as the log-likelihood for the mixed logit model. Therefore, we expect that the former will likely be greater than the log-likelihood for the mixed logit model.

The right-hand plot of Fig. 6 shows the cumulative distribution function of the difference at convergence in the log-likelihood for the ICLV model, as given by Eq. (11), and the mixed logit model, for the 100 datasets generated in Section 3.1 with 1000 observations each. For all but one of these datasets, the log-likelihood for the ICLV model is found to be greater than the log-likelihood for the mixed logit model. The t-statistic for the difference is 3.37 for the estimation sample and 2.23 for the validation sample, and we accept the alternative hypothesis that the log-likelihood for the ICLV model is greater than that for the mixed logit model with a confidence of 99.5% and 97.5%, respectively. Therefore, we conclude that in cases where observable explanatory variables are poor predictors of the latent variables, the use of the measurement equations can indeed result in an improvement in the ICLV model's ability to predict choice outcomes. In fact, the difference in the log-likelihood values given by (10) and (11) can be used as a measure of the latency of the latent variables.

3.5. Summary on goodness of fit

There are two ways in which an ICLV model can be used to predict outcomes to the choice data. The first method formulates the likelihood of choice data as a function solely of observable explanatory variables, and the second method formulates the likelihood of choice data as a function of both observable explanatory variables and measurement indicators. The method used is often determined by the objectives of the study.

In cases where we are interested in using the ICLV model to predict outcomes to the choice data, and forecasts for the measurement indicators are expected to be unavailable, the choice likelihood for the ICLV model must necessarily be expressed solely as a function of observable explanatory variables. As long as the structural equations for the latent variables have a parametric form, the analyst can substitute the equation in the utility specification of the discrete choice sub-model to obtain a reduced form choice model without latent variables that explains the choice data at least as well as the ICLV model from which it was originally derived. In these cases, it is important to demonstrate what benefits, if any, might be had from adopting the ICLV framework, recognizing that improvement in fit is not one of them.

For certain model specifications, it could be argued that the structure imposed by an ICLV model may lead the analyst to a reduced form choice model that may never have been evaluated had the analyst not estimated the original choice model with latent variables. Consider an ICLV model specification where the discrete choice sub-model contains interactions between observable and latent variables (see, for example, Abou-Zeid et al., 2010; Hess and Beharry-Borg, 2012; Bhat et al., 2015). The utility specification for the equivalent reduced form choice model would comprise interactions between different pairs of observable explanatory variables. When $M \ll K$, or the number of latent variables is much smaller than the number of observable explanatory variables, the number of possible interactions that can be introduced between different pairs of observable explanatory variables is much larger than the number of interactions that can be introduced between latent and observable variables, and the reduced form choice model can be much more complex than model specifications typically tested in the literature. In such cases, the ICLV model helps lend structure to the underlying sources of heterogeneity, and it may be fair to say that, in the absence of latent variables to inform the process of model development, an analyst exploring the space of all possible model specifications may never chance upon the appropriate reduced form model specification. Similar arguments could be made with regards to models that employ complex correlation structures, such as spatial and social interactions, where again the reduced form model specification might be somewhat convoluted (see, for example, Kassahun et al., 2015).

And finally, in cases where forecasts for the measurement indicators are expected to be available, or in cases where we are interested only in using the ICLV model to understand the cognitive process underlying decision-making, the choice likelihood for the ICLV model may be calculated as an additional function of the measurement indicators. If the model has strong structural equations, and the observable explanatory variables are good predictors of the latent variables, then on average the choice likelihood of the ICLV model should be the same as the choice likelihood of the reduced form model, though results may vary from case to case. But if the model suffers from weak structural equations, and the observable explanatory variables are not good predictors of the latent variables, then the choice likelihood of the ICLV model may indeed be better than the choice likelihood of the reduced form model, and the use of the latent variable framework may lead to statistically significant improvements in fit.

4. Identification of latent variable effects and parameter decomposition

ICLV models were conceived with the intention of enriching existing representations of decision-making through the inclusion of latent biological, sociological and psychological constructs. Discrete choice models have long used sociodemographic variables and random error components as proxies for these latent behavioral constructs, but without the ability to say exactly what these variables might be proxies for. By incorporating additional data through the use of measurement indicators, the ICLV model can help identify additional parameters associated with the latent explanatory variables, test for the effect of these latent variables on observable choices, and lend structure and meaning to underlying differences in behavior.[3] To see that this is the case, we substitute (2) in (1) to get the following reduced form for the utility specification:

$$\mathbf{u_n} = (\mathbf{B} + \mathbf{\Gamma A})\mathbf{x_n} + \mathbf{\Gamma \upsilon_n} + \boldsymbol{\varepsilon}_\mathbf{n} \tag{40}$$

A more traditional choice model without latent variables would only be able to identify the $(J \times K)$ matrix $\mathbf{T} = \mathbf{B} + \mathbf{\Gamma A}$. However, the ICLV model can help decompose the influence of the vector of observable explanatory variables $\mathbf{x_n}$, as denoted by $\mathbf{T}$, into different constituent effects, as represented by the matrices $\mathbf{B}$, $\mathbf{\Gamma}$ and $\mathbf{A}$. Through the measurement equation given by (3), these effects can be explicitly linked to specific behavioral constructs. Similarly, an error components choice model would be able to identify the covariance matrix $\mathbf{Z} = \mathbf{\Gamma \Phi \Gamma}^\mathrm{T}$, but would not be able to ascribe the underlying source of heterogeneity to differences in the latent variables.

Take, for example, the hypothetical model of bicycle ownership first presented in Section 3.1, given by (20)–(25). Eqs. (22) and (23) may be substituted in (20) and (21) to get the following utility specification for the reduced form model:

$$u_{n1} = (\beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21})x_{n1} + (\gamma_{11}\alpha_{12})x_{n2} + (\beta_{13} + \gamma_{11}\alpha_{13} + \gamma_{12}\alpha_{23}) + (\gamma_{11}\nu_{n1} + \gamma_{12}\nu_{n2} + \varepsilon_{n1}) \tag{41}$$

$$u_{n2} = \varepsilon_{n2} \tag{42}$$

As we did in Section 3.1, define the parameters τ_{11}, τ_{12}, τ_{13} and ζ_{11} as follows:

$$\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21} \tag{43}$$

$$\tau_{12} = \gamma_{11}\alpha_{12} \tag{44}$$

$$\tau_{13} = \beta_{13} + \gamma_{11}\alpha_{13} + \gamma_{12}\alpha_{23} \tag{45}$$

$$\zeta_{11} = \sqrt{\gamma_{11}^2 + \gamma_{12}^2} \tag{46}$$

where τ_{11} and τ_{12} capture the influence of age and gender on the utility of bicycle ownership, respectively, τ_{13} is the alternative-specific constant, and ζ_{11} is the variance of the error component. Eq. (43) shows how the ICLV model may be employed to decompose the influence of age on the utility of bicycle ownership into three constituent effects: the indirect influence of age through its effect on attitudes towards the environment ($\gamma_{11}\alpha_{11}$) and susceptibility to peer pressure ($\gamma_{12}\alpha_{21}$), and some unexplained residual effect (β_{11}). Similarly, (44) states that the influence of gender on the utility of bicycle ownership may be attributed entirely to its influence on attitudes towards the environment ($\gamma_{11}\alpha_{12}$). Eq. (45) shows how the ICLV model may be employed to decompose the alternative-specific constant τ_{13} in the reduced form mixed logit model into three constituent effects: the mean population effect of attitudes towards the environment ($\gamma_{11}\alpha_{13}$), the mean population effect of susceptibility to peer pressure ($\gamma_{12}\alpha_{23}$), and some residual that captures the mean of that which is either unobservable or purely random (β_{13}). It is important to recognize that for the ICLV model, the two latent variables are

[3] Note that this section does not discuss the theoretical or empirical identification of ICLV models, the word identification is used in the context of the ICLV model's ability to *identify* the direct and indirect effects of different observable and latent variables on the choice outcome of interest.

no longer unobservable variables whose mean population effect would otherwise have been subsumed by the alternative-specific constant τ_{13} in the reduced form mixed logit model. And finally, (46) shows how the ICLV model can decompose the variance of the stochastic component into separate elements that correspond to random heterogeneity with regards to attitudes towards the environment and susceptibility to peer pressure.

In general, ICLV models offer greater explanatory power than choice models without latent variables by allowing the analyst to decompose the influence of observable variables into constituent effects, each of which can subsequently be attributed to some latent construct. Parameter decomposition is a powerful tool with which to pry open black-box representations of the decision-making process such as those usually employed by traditional choice models. However, as mentioned in Section 3, ICLV models in practice may suffer from weak structural equations where observable variables are poor predictors of the latent variables, and weak measurement equations where responses to the appropriate indicators are either not available or not requested. For parameter decomposition to be of value to the analyst, the data must allow for the empirical identification of each of the different constituent effects associated with each observable variable. If the data is insufficient, in terms of either quantity or quality, then the benefits of using the framework are likely to be limited.

5. Bias of the parameter estimates

It would appear that the inclusion of latent variables and measurement indicators should help correct for bias due to omitted variables or measurement error (see, for example, Walker et al., 2010; Guevara, 2015). As we argue in this section, this is not true in cases where the model framework is used under traditional assumptions of causality, and the observable variables entering the structural equation of the latent variables are assumed to causally determine the latent variables. However, if the analyst is willing to relax some of these assumptions, and use the structural equation merely to estimate the covariance between the observable and latent variables, then the framework can potentially be used to correct for bias. Note that the reduced form choice model can only estimate the composite parameter $\mathbf{T} = \mathbf{B} + \mathbf{\Gamma A}$, and not the individual parameters $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Gamma}$. In the previous section, we argued that the inclusion of measurement indicators in the ICLV model can help decompose the influence of the observable explanatory variables on the choice indicators into constituent components, each of which can be attributed to different latent constructs. In this section, we contend that under certain cases, the inclusion of measurement indicators can, at least theoretically, help control for bias, where the difference between $\mathbf{T}$ on one hand, and $\mathbf{B}$ or $\mathbf{\Gamma}$ on the other, can be construed as a form of bias. Section 5.1 discusses the issue in the context of omitted variables and Section 5.2 discusses the issue in the context of measurement error. However, as we concluded before, the practical ability of the ICLV model to control for bias in both cases will be determined by the availability of appropriate measurement indicators.[4]

5.1. Omitted variable bias

Suppose that the data generating process for the choice indicators is given by Eqs. (1) and (4), such that $\mathbf{x_n}$ and $\mathbf{x_n^*}$ are both uncorrelated with $\boldsymbol{\varepsilon}_\mathbf{n}$. If we fail to observe $\mathbf{x_n^*}$, the variable is implicitly absorbed by the stochastic component of the utility. If $\mathbf{x_n}$ and $\mathbf{x_n^*}$ are correlated through (2), and we estimate a choice model with the utility specification $\mathbf{u_n} = \mathbf{T x_n} + \boldsymbol{\omega}_\mathbf{n} + \boldsymbol{\varepsilon}_\mathbf{n}$ on the choice indicators generated by the process given by (1) and (4), the stochastic component is no longer independent of $\mathbf{x_n}$, the variable $\mathbf{x_n}$ is endogenous, and $\mathbf{T} \neq \mathbf{B}$. In fact, since $\mathbf{x_n}$ and $\mathbf{x_n^*}$ are correlated, $\mathbf{A}$ is a non-zero matrix, and the difference between $\mathbf{T}$ and $\mathbf{B}$ may be given by $\mathbf{\Gamma A}$.

Assume now that we have access to measurement indicators $\mathbf{i_n}$, such that they are correlated with $\mathbf{x_n^*}$ through (3), and $\mathbf{x_n^*}$ and $\boldsymbol{\eta}_\mathbf{n}$ are uncorrelated. If we estimate an ICLV model with the specification given by Eqs. (1)–(4), using the choice indicators generated by the process given by (1) and (4) and the measurement indicators generated by the process given by (3), treating $\mathbf{x_n^*}$ as latent, if $\mathbf{x_n}$ and $\boldsymbol{\upsilon}_\mathbf{n}$ are uncorrelated, we should obtain unbiased estimates for each of the model parameters $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Gamma}$.

Whether the difference between $\mathbf{T}$ and $\mathbf{B}$ can be regarded as bias or not will depend on the causal relationship between the latent and observable explanatory variables. In most cases, the cross-sectional nature of the data does not allow for the identification of causal relationships between different pairs of variables, only associations can be identified, and the internal validity of the relationship is left to the judgment of the analyst (Gujarati, 2004, pp. 22–23). If there are reasons to believe that $\mathbf{x_n^*}$ is causally determined by $\mathbf{x_n}$, then the reduced form model can reliably estimate the cumulative effect of $\mathbf{x_n}$ on choice, as denoted by $\mathbf{T}$, but cannot decompose it like the ICLV model. However, if there are reasons to believe that $\mathbf{x_n}$ and $\mathbf{x_n^*}$ are merely correlated, but there is no causal relationship, then the effect of $\mathbf{x_n}$ on choice as estimated by the reduced form model, still given by $\mathbf{T}$, is biased, and the true effect is given by $\mathbf{B}$.

The reader should note that under the latter case the structural equation for the latent variables does not represent a causal relationship between $\mathbf{x_n}$ and $\mathbf{x_n^*}$. The causal relationship of interest is between $\mathbf{x_n}$ and $\mathbf{y_n}$, and the structural equation is merely being used to identify the covariance between $\mathbf{x_n}$ and $\mathbf{x_n^*}$. This is a significant departure from the usual assumptions of causality underlying the ICLV model framework, where $\mathbf{x_n^*}$ is assumed to be causally determined by $\mathbf{x_n}$. Therefore, when drawing inferences from the model, the analyst needs to be aware of the distinction, and any post-estimation analysis needs to be adapted accordingly (see, for example, the discussion that follows in Section 7.2). In fact, the use of the

[4] The discussion contained in this section was inspired by comments from three anonymous reviewers.

Table 2
Sample mean (and *t*-statistic) of parameter estimates for the ICLV model given by (20)–(25) and the reduced form mixed logit model given by (26) and (27), where the null hypothesis is that the mean parameter estimate equals the true value.

Variable	Model parameter	True value	Model	Number of observations			
				100	200	500	1000
Age	β_{11}	−0.45	ICLV	−9.77 (−0.29)	−0.93 (−0.69)	−0.47 (−0.09)	−0.47 (−0.18)
			Mixed logit	–	–	–	–
	τ_{11}	0.38	ICLV	2.19 (0.24)	0.40 (0.13)	0.38 (−0.05)	0.38 (0.00)
			Mixed logit	0.37 (−0.04)	0.38 (0.02)	0.38 (−0.05)	0.38 (−0.05)
Gender	τ_{12}	0.90	ICLV	6.56 (0.26)	1.15 (0.48)	0.89 (−0.03)	0.92 (0.18)
			Mixed logit	0.89 (−0.01)	0.92 (0.05)	0.91 (0.02)	0.92 (0.12)
Constant	τ_{13}	−0.15	ICLV	−1.81 (−0.16)	−0.29 (−0.32)	−0.12 (0.13)	−0.14 (0.07)
			Mixed logit	−0.12 (0.06)	−0.17 (−0.04)	−0.13 (0.07)	−0.14 (0.05)

structural equation in this case may be likened to two-stage least squares regression, where the first stage regresses the endogenous explanatory variable on the instrument, even when there might be no causal relationship between the two.

As before, we corroborate the assertion using synthetic data. For the one hundred datasets corresponding to 100, 200, 500 and 1000 observations, generated in Section 3.1 using the ICLV model specification given by (20)–(25), Table 2 enumerates the sample mean of the estimates for β_{11} as recovered by the ICLV model specification, and τ_{11}, τ_{12} and τ_{13} as recovered by both the ICLV model specification and the corresponding reduced form mixed logit model given by (26) and (27). Note that the estimates for the parameters τ_{11}, τ_{12} and τ_{13} are not directly given by the ICLV model but have to be calculated using (43)–(45) and the estimates for **A, B** and $\mathbf{\Gamma}$.

The mean estimate for β_{11} as recovered by the ICLV model can be seen to converge to the true value, and for each of the four sets of 100 datasets, we fail to reject the null hypothesis that the mean parameter estimate is equal to the true value. The mixed logit model cannot identify β_{11}, but can only identify the composite parameter $\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21}$. How bad this limitation might be will be determined by the causal relationship between the latent and observable explanatory variables. For example, is it reasonable to assume that attitudes towards the environment and susceptibility to peer pressure are causally determined by a decision maker's age? Or are statistically significant relationships between these variables reflective merely of correlation? If the former is true, then as described in the previous section, τ_{11} captures the cumulative effect of age on bicycle ownership, arising from its indirect effect on attitudes towards the environment and susceptibility to peer pressure, and some residual effect. The mixed logit model can yield an unbiased estimate of this cumulative effect (see following paragraph), but it cannot decompose it like the ICLV model. However, if there is no causal relationship between these variables, only correlation, then estimates from the mixed logit model may lead to the spurious conclusion that age has a positive effect on bicycle ownership ($\tau_{11} > 0$), when in fact the opposite is true ($\beta_{11} < 0$). The ICLV model, on the other hand, can identify both τ_{11} and β_{11}, and depending on the judgment of the analyst, one or the other parameter may be interpreted as the effect of age on bicycle ownership, and the difference between the two parameters may or may not be interpreted as bias.

The mean estimates for τ_{11}, τ_{12} and τ_{13} from both model forms can be seen to converge to their true values, and in each case, we fail to reject the null hypothesis that the mean parameter estimates are equal to the true values. However, estimates from the reduced form mixed logit model seem to converge faster, likely due to differences in model complexity.[5] For example, at 100 observations the mean estimates from the ICLV model are off by several orders of magnitude. In fact, the ICLV model was not identified empirically for seven of the one hundred datasets corresponding to 100 observations, indicating that 100 observations are likely not enough to support estimation of the particular specification. At 200, 500 and 1000 observations, the difference in mean parameter estimates between the two models is small and rapidly diminishing. This may be an artifact arising out of the particulars of the ICLV model specification used to generate the data. In general, estimates for **T** from both model forms will be consistent. In some cases, a greater number of observations may be required to support estimation of the ICLV model, but in others, the additional information available to the ICLV model through the measurement indicators may necessitate the collection of a fewer number of total observations.

5.2. Measurement error bias

Suppose that the data generating process for the choice indicators is given by the following set of equations:

$$\mathbf{u_n} = \mathbf{\Gamma x_n^*} + \boldsymbol{\varepsilon}_\mathbf{n} \tag{47}$$

$$y_{nj} = \begin{cases} 1 & \text{if } u_{nj} \geq u_{nj'} \text{ for } j' \in \{1, \ldots, J\} \\ 0 & \text{otherwise} \end{cases} \tag{48}$$

[5] As pointed out by a reviewer, the two models do not have the same number of model parameters and are not estimated using the same data. Therefore their small sample properties should expectedly differ.

where the variables and parameters have the same definitions as before, and $\mathbf{x_n^*}$ is uncorrelated with $\boldsymbol{\varepsilon}_\mathbf{n}$. Assume that $\mathbf{x_n^*}$ is not observed, but is measured using $\mathbf{x_n}$, such that the error in measurement $\boldsymbol{\nu}_\mathbf{n}$ may be given as follows:

$$\boldsymbol{\nu}_\mathbf{n} = \mathbf{x_n^*} - \mathbf{x_n} \tag{49}$$

Substituting (49) in (47), we get the following expression:

$$\mathbf{u_n} = \boldsymbol{\Gamma}(\mathbf{x_n} + \boldsymbol{\nu}_\mathbf{n}) + \boldsymbol{\varepsilon}_\mathbf{n} = \boldsymbol{\Gamma}\mathbf{x_n} + (\boldsymbol{\Gamma}\boldsymbol{\nu}_\mathbf{n} + \boldsymbol{\varepsilon}_\mathbf{n}) \tag{50}$$

where the stochastic component is now represented by the combined term $\boldsymbol{\Gamma}\boldsymbol{\nu}_\mathbf{n} + \boldsymbol{\varepsilon}_\mathbf{n}$. If $\boldsymbol{\nu}_\mathbf{n}$ is uncorrelated with $\mathbf{x_n^*}$, then it can be shown that $\boldsymbol{\nu}_\mathbf{n}$ is correlated with $\mathbf{x_n}$ (and the covariance between the two is given by the variance of the measurement error). If we estimate a choice model with the utility specification $\mathbf{u_n} = \mathbf{T}\mathbf{x_n} + \boldsymbol{\omega}_\mathbf{n} + \boldsymbol{\varepsilon}_\mathbf{n}$ on the choice data generated by the process given by (47) and (48), the stochastic component of the utility specification is no longer independent of $\mathbf{x_n}$, the variable $\mathbf{x_n}$ is endogenous, and $\mathbf{T} \neq \boldsymbol{\Gamma}$. In fact, since $\mathbf{x_n}$ and $\mathbf{x_n^*}$ are correlated, $\mathbf{A}$ is a non-zero matrix, we know that $\mathbf{B}$ is the zero matrix, and $\mathbf{T} = \boldsymbol{\Gamma}\mathbf{A}$.

Assume now that we have access to measurement indicators $\mathbf{i_n}$, such that they are correlated with $\mathbf{x_n^*}$ through (3), and $\mathbf{x_n^*}$ and $\boldsymbol{\eta}_\mathbf{n}$ are uncorrelated. As before, if we estimate an ICLV model with the specification given by (1)–(4), using the choice indicators generated by the process given by (47) and (48) and the measurement indicators generated by the process given by (3), treating $\mathbf{x_n^*}$ as latent, if $\mathbf{x_n}$ and $\boldsymbol{\upsilon}_\mathbf{n}$ are uncorrelated, we should obtain unbiased estimates for each of the model parameters $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\Gamma}$.

In this case, there is no causal relationship between $\mathbf{x_n}$ and $\mathbf{x_n^*}$, the two variables represent the observed and true values for the same set of explanatory variables, and the only parameter of interest is $\boldsymbol{\Gamma}$. The reduced form model can estimate $\mathbf{T}$, but not $\boldsymbol{\Gamma}$. Since $\mathbf{T} = \boldsymbol{\Gamma}\mathbf{A}$, estimates from the model are biased by a factor equal to $\mathbf{A}$. The ICLV model, on the other hand, can identify $\boldsymbol{\Gamma}$, yielding an unbiased estimate for the effect of the explanatory variables of interest on the choice indicators. As we mentioned in the previous section, the structural equation for the latent variables does not represent a causal relationship between $\mathbf{x_n}$ and $\mathbf{x_n^*}$, but is merely being used to identify the covariance between the observed and true values for the explanatory variables. Therefore, the same caveats apply.

We corroborate the assertion by synthesizing data for a hypothetical model of bicycle ownership, summarized by the following set of four equations:

$$u_{n1} = \beta_{12}x_{n2} + \beta_{13} + \gamma_{11}x_{n1}^* + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{51}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{52}$$

$$x_{n1} = x_{n1}^* + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{53}$$

$$i_{n1} = x_{n1}^* + \eta_{n1}, \quad \eta_{n1} \sim N(0, 1) \tag{54}$$

where u_{n1} denotes the utility of owning a bicycle and u_{n2} denotes the utility of not owning one, as perceived by decision-maker n; x_{n1}^* and x_{n1} denote the decision-maker's true and observed age, respectively; x_{n2} denotes the decision-maker's gender; and i_{n1} denotes a measurement indicator that is correlated with the decision-maker's true age, such as income.[6] The variables denoting a decision-maker's true age and gender are simulated using the same distributions as the Monte Carlo experiment in Section 3.1, and values for each of the model parameters are selected using the same methodology as before. Note that in our general discussion, we had constrained $\mathbf{B}$ to be the zero matrix, to keep the analysis as straightforward as possible. In the simulation exercise, we are allowing elements of $\mathbf{B}$ that correspond to observable variables that do not have measurement error to be non-zero. This does not change any of the conclusions arrived at in the general analysis.

100 datasets each are generated for 100, 200, 500 and 1000 pseudo-observed decision-makers hypothesized to behave according to the decision-making process described above, resulting in a total of 400 datasets. For each of the 400 datasets, we estimate an ICLV model with the following specification:

$$u_{n1} = \beta_{12}x_{n2} + \beta_{13} + \gamma_{11}x_{n1}^* + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{55}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{56}$$

$$x_{n1}^* = \alpha_{11}x_{n1} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{57}$$

$$i_{n1} = x_{n1}^* + \eta_{n1}, \quad \eta_{n1} \sim N(0, 1) \tag{58}$$

[6] Perhaps a better example would be the case where income is the variable with measurement error, and measures such as number of cars owned, or number of bedrooms in current place of residence, may be treated as indicators of the true income. However, for the sake of consistency, we stay with the example of age.

Table 3
Sample mean (and *t*-statistic) of parameter estimates for the ICLV model given by (55)–(58) and the reduced form mixed logit model given by (59) and (60), where the null hypothesis is that the mean parameter estimate equals the true value.

Variable	Model parameter	True value	Model	Number of observations			
				100	200	500	1000
Age	γ_{11}	0.30	ICLV	0.30 (0.00)	0.30 (0.02)	0.30 (0.01)	0.30 (−0.02)
			Mixed logit	–	–	–	–
	τ_{11}	0.20	ICLV	0.20 (0.00)	0.21 (0.06)	0.20 (0.08)	0.20 (0.03)
			Mixed logit	0.19 (−0.07)	0.21 (0.12)	0.20 (0.02)	0.20 (−0.08)
Gender	β_{12} or τ_{12}	−1.00	ICLV	−1.07 (−0.15)	−1.03 (−0.11)	−1.00 (0.00)	−1.02 (−0.16)
			Mixed logit	−1.07 (−0.14)	−1.04 (−0.13)	−1.00 (0.02)	−1.03 (−0.19)
Constant	τ_{13}	−0.60	ICLV	−0.59 (0.03)	−0.62 (−0.08)	−0.63 (−0.15)	−0.60 (−0.02)
			Mixed logit	−0.57 (0.07)	−0.64(−0.12)	−0.62 (−0.11)	−0.59 (0.06)

where x^*_{n1} is treated as the latent estimate of the decision-maker's age, and x_{n1} is treated as the measured estimate of the same. Similarly, we estimate a mixed logit model with the following specification:

$$u_{n1} = \tau_{11}x_{n1} + \tau_{12}x_{n2} + \tau_{13} + \omega_{n1} + \varepsilon_{n1}, \quad \omega_{n1} \sim N(0, \zeta_{11}), \quad \varepsilon_{n1} \sim \text{GEV}(0, 1, 0) \tag{59}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim \text{GEV}(0, 1, 0) \tag{60}$$

Table 3 enumerates the sample mean of the estimates for γ_{11} as recovered by the ICLV model specification, and τ_{11}, τ_{12} and τ_{13} as recovered by both the ICLV model specification and the reduced form mixed logit. Note that the estimates for the parameters τ_{11}, τ_{12} and τ_{13} are not directly given by the ICLV model but have to be calculated indirectly from the estimates for **A, B** and $\mathbf{\Gamma}$. In each case, we fail to reject the null hypothesis that the mean parameter estimates are equal to the true values. However, due to measurement error, the mixed logit model risks underestimating the effect of age on bicycle ownership ($\tau_{11} < \gamma_{11}$), while the inclusion of the measurement indicator allows the ICLV model to yield an unbiased estimate for the same. With regards to the other model parameters, estimates from both models can be seen to converge to their true values, and in each case, we fail to reject the null hypothesis that the mean parameter estimates are equal to the true values.

6. Variance of the parameter estimates

It is widely believed that the additional information available to ICLV models through the incorporation of measurement indicators leads to a reduction in the variance of parameter estimates. However, evidence to support the belief is, at best, limited. A study on the adoption of electric cars by Glerum (2014) uses a holdout sample to compare the choice probabilities for the chosen alternative as predicted by an ICLV model with those from an equivalent choice model without latent variables. The study finds that the difference between the average values for the 5% and 95% confidence bounds on the predicted choice probabilities is 17.3% for the ICLV model and 18.5% for the choice model without latent variables, indicating that the ICLV model gives "slightly more accurate choice probabilities" than the choice model without latent variables. Similar arguments have been forwarded by other studies with respect to other model frameworks that fall within the broader family of hybrid choice models, but differ from the ICLV framework. For example, Ben-Akiva and Boccara (1995) and Abou-Zeid and Ben-Akiva (2014) have argued that the use of additional indicators to measure latent variables can "improve the efficiency of the model" parameters thus estimated. However, the example employed by both studies formulates the latent variable as a discrete construct, resulting in an LCCM and not an ICLV model. Abou-Zeid (2009) reports similar findings for a hybrid choice model where measures of happiness and wellbeing are used as additional indicators of the utility associated with different alternatives. In other words, the utility of different alternatives is used to predict outcomes to both choice and measurement indicators. In such cases, the use of measurement indicators is akin to the use of a greater number of observations to estimate the same model. Just as the variance of the parameter estimates decreases with sample size, it decreases also with the use of these additional indicators. However, this formulation differs from the general ICLV model framework where the utility of different alternatives is used solely to predict outcomes to the choice indicators. With regards specifically to the ICLV model as defined in Section 2, Glerum (2014) is the only study that the authors are aware of that offers some evidence of a decrease in variance, but their findings are not conclusive.

Since a reduced form model cannot estimate **A, B** and $\mathbf{\Gamma}$, we limit our attention to estimates for **T**. We show that ICLV models may produce more efficient estimates for **T** through an analysis of the covariance matrix of observable variables, based on the seminal work by Jöreskog (1978). The covariance matrix of observable variables offers insights to the structure underlying observable data and provides a mechanism for identifying unknown model parameters. Using notation not specific to any particular model form, the sample covariance matrix may be expressed as a function of unknown model parameters, denoted $\boldsymbol{\theta}$, as follows:

$$\sigma_{ij} = f(\theta) \tag{61}$$

where σ_{ij} is the $(i, j)^{\text{th}}$ element of the sample covariance matrix $\mathbf{\Sigma}$. Each element of the sample covariance matrix provides a unique equation in the model parameters. If a particular parameter θ can be determined from solving some subset of the equations given by (61), the parameter is identified. In some cases, a parameter can be determined in multiple ways using different sets of equations. This gives rise to overidentifying conditions on $\mathbf{\Sigma}$ that must hold if the assumed model specification is the true data generating process. If the sample covariance matrix uniquely determines each element of the vector of model parameters $\boldsymbol{\theta}$, then the model specification as a whole is identified. An analysis of the covariance matrix is useful to determine which subset of observable variables offers information about which subset of model parameters, and how the availability of more information through the inclusion of additional observable variables may lead to parameter estimates with smaller variance. In restricting our attention to the sample covariance matrix, we are implicitly assuming that the distribution of observable variables can be described sufficiently in terms of the first and second order moments, and that additional information about the model parameters contained in higher order moments may be ignored.

For a choice model without latent variables, the observable variables comprise the vector of choice indicators $\mathbf{y_n}$ and the vector of explanatory variables $\mathbf{x_n}$. The covariance matrix could be formulated directly with regards to these two vectors, but due to the nonlinearity of the choice sub-model, instead of working with the vector of choice indicators $\mathbf{y_n}$ it is easier to work with the vector of utilities $\mathbf{u_n}$. Since only the differences in utilities are identifiable, and not the absolute levels themselves, we introduce the $((J-1)\times 1)$ vector of the differences in utilities, denoted $\Delta\mathbf{u_n}$, where Δ is the linear operator that transforms the J utilities into $(J-1)$ utility differences taken with respect to the J^{th} alternative. Δ is a $((J-1)\times J)$ matrix that consists of a $((J-1)\times(J-1))$ identity matrix with a column vector of -1's appended as the J^{th} column. Though Δ performs the differences with respect to the last alternative for each choice situation, our analysis is indifferent to whichever alternative is used as the base. Therefore, for a choice model without latent variables, the observable variables may be redefined as the vector of utility differences $\Delta\mathbf{u_n}$ and the vector of explanatory variables $\mathbf{x_n}$. For an ICLV model, the observable variables comprise the additional vector of measurement indicators $\mathbf{i_n}$. We show here that the inclusion of the measurement indicators in the ICLV model offers additional information about the reduced form model parameters $\mathbf{T}$, information that would not be available to a choice model without latent variables, and this additional information can produce parameter estimates with potentially smaller variance.

Under the assumptions outlined in Section 2 for the general ICLV model framework, the variance of the vector of utility differences $\Delta\mathbf{u_n}$, and the covariance between the vector of utility differences $\Delta\mathbf{u_n}$ and the vector of explanatory variables $\mathbf{x_n}$, can be parameterized as follows:

$$\begin{aligned}\text{var}(\Delta\mathbf{u_n}) &= \Delta(\mathbf{B}+\mathbf{\Gamma A})E[(\mathbf{x_n}-E[\mathbf{x_n}])(\mathbf{x_n}-E[\mathbf{x_n}])^{\text{T}}](\mathbf{B}+\mathbf{\Gamma A})^{\text{T}}\Delta^{\text{T}}+\Delta\mathbf{\Gamma\Phi\Gamma}^{\text{T}}\Delta^{\text{T}}+k\Delta\Delta^{\text{T}} \\ &= \Delta\mathbf{T}\text{var}(\mathbf{x_n})\mathbf{T}^{\text{T}}\Delta^{\text{T}}+\Delta\mathbf{Z}\Delta^{\text{T}}+k\Delta\Delta^{\text{T}}\end{aligned} \tag{62}$$

$$\text{cov}(\Delta\mathbf{u_n},\mathbf{x_n}) = \Delta(\mathbf{B}+\mathbf{\Gamma A})\text{var}(\mathbf{x_n}) = \Delta\mathbf{T}\text{var}(\mathbf{x_n}) \tag{63}$$

where k is a constant, equal to the variance of a standard Gumbel distributed random variable. For a choice model without latent variables, (62) and (63) offer the only information available to identify the matrices of reduced form model parameters $\mathbf{T}$ and $\mathbf{Z}$. For an ICLV model, the analyst has access to additional information through the inclusion of the vector of measurement indicators $\mathbf{i_n}$:

$$\text{var}(\mathbf{i_n}) = \mathbf{DA}\text{var}(\mathbf{x_n})\mathbf{A}^{\text{T}}\mathbf{D}^{\text{T}}+\mathbf{D\Phi D}^{\text{T}}+\mathbf{\Psi} \tag{64}$$

$$\text{cov}(\mathbf{i_n},\mathbf{x_n}) = \mathbf{DA}\text{var}(\mathbf{x_n}) \tag{65}$$

$$\text{cov}(\Delta\mathbf{u_n},\mathbf{i_n}) = \Delta(\mathbf{B}+\mathbf{\Gamma A})\text{var}(\mathbf{x_n})\mathbf{A}^{\text{T}}\mathbf{D}^{\text{T}}+\Delta\mathbf{\Gamma\Phi D}^{\text{T}} \tag{66}$$

The reader's attention is directed towards (66), which indicates that the covariance between the measurement indicators and the differences in utilities (as measured indirectly through the choice indicators) can be expressed as a direct function of, among other variables, the matrix of reduced form model parameters $\mathbf{T}=\mathbf{B}+\mathbf{\Gamma A}$. In other words, the inclusion of the measurement indicators $\mathbf{i_n}$ in the ICLV model may provide potentially additional information about $\mathbf{T}$. Depending on the specification for each of the unknown model parameters, and the resulting form of (62)–(66), in some cases (66) may be used to identify one or more additional model parameters, and in others it may be used to impose overidentifying conditions on the sample covariance matrix. If the former is true, estimates for $\mathbf{T}$ from the ICLV model and the reduced form choice model should have the same variance. But if the latter is true, estimates from the ICLV model may have lower variance.

Consider, for the sake of illustration, the two ICLV model specifications given by (20)–(25) and (30)–(33). For the latter model specification, the reader should verify that $\mathbf{A}$ can be completely identified from (65), $\mathbf{B}$ and $\mathbf{\Gamma}$ can be completely identified from (63), $\mathbf{\Phi}$ can be completely identified by (64), $\mathbf{D}$ and $\mathbf{\Psi}$ are constant matrices that do not need to be estimated, (62) helps to set the scale for the utility differences, and (66) serves to impose overidentifying restrictions on the sample covariance matrix. For the former model specification, the reader should similarly verify that $\mathbf{A}$ can be completely identified from (65), $\mathbf{\Phi}$ can be completely identified by (64), $\mathbf{D}$ and $\mathbf{\Psi}$ are constant matrices that do not need to be estimated, and (62) helps to set the scale for the utility differences. However, (63) alone is no longer sufficient to identify $\mathbf{B}$ and $\mathbf{\Gamma}$, and (66) is needed as well to ensure that the model specification can be identified. As a result, estimates from the ICLV model given by (20)–(25) and the corresponding reduced form model given by (26) and (27) should have equal variance. But estimates

Table 4
Sample standard error of parameter estimates for the ICLV model given by (20)–(25) and the reduced form mixed logit model given by (26) and (27), and the F-statistic under the null hypothesis that the variances of the parameter estimates from the two models are equal.

Parameter	Model and test statistic	Number of observations			
		100	200	500	1000
Age (τ_{11})	ICLV	7.67	0.17	0.09	0.06
	Mixed logit	0.16	0.12	0.08	0.06
	F-stat	∞	1.93	1.05	1.17
Gender (τ_{12})	ICLV	21.67	0.52	0.25	0.16
	Mixed logit	0.55	0.41	0.27	0.17
	F-stat	∞	1.58	0.74	0.85
Constant (τ_{13})	ICLV	10.37	0.45	0.25	0.16
	Mixed logit	0.51	0.39	0.25	0.16
	F-stat	∞	1.31	0.89	0.96

Table 5
Sample standard error of parameter estimates for the ICLV model given by (30)–(33) and the reduced form mixed logit model given by (26) and (27), and the F-statistic under the null hypothesis that the variances of the parameter estimates from the two models are equal.

Parameter	Model and test statistic	Number of observations			
		100	200	500	1000
Age (τ_{11})	ICLV	0.16	0.09	0.06	0.04
	Mixed logit	0.21	0.13	0.08	0.06
	F-stat	0.56	0.52	0.56	0.48
Gender (τ_{12})	ICLV	0.41	0.28	0.20	0.13
	Mixed logit	0.61	0.45	0.29	0.19
	F-stat	0.46	0.38	0.46	0.52
Constant (τ_{13})	ICLV	0.52	0.35	0.22	0.15
	Mixed logit	0.63	0.41	0.27	0.20
	F-stat	0.68	0.70	0.69	0.61

from the ICLV model given by (30)–(33) should have lower variance than the corresponding reduced form model given by (26) and (27).

As before, we corroborate the result using Monte Carlo experiments. For the one hundred datasets corresponding to 100, 200, 500 and 1000 observations, generated in Section 3.1 using the ICLV model specification given by (20)–(25), Table 4 enumerates the sample standard error of the estimates for the three parameters τ_{11}, τ_{12} and τ_{13} as recovered indirectly by the ICLV model specification, using Eqs. (43)–(45) and the estimates for $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\Gamma}$, and directly by the corresponding reduced form mixed logit model given by (26) and (27). To facilitate a comparison between the sample standard errors for the two models, we calculate the ratio of the sample variances. Given that there are 100 datasets each corresponding to either model, the ratio has an F distribution with both degrees of freedom equal to 99. The null hypothesis that the sample variances for the two models are equal can be rejected in favor of the alternative hypothesis that the sample variances are different with a confidence of 95% when the test statistic is less than 0.67 or greater than 1.49. For 100 and 200 observations, the ICLV model has higher sample standard errors, indicating that a greater number of observations is needed to support estimation. However, as the number of observations grows, the difference in sample standard errors decreases, and for 500 and 1000 observations, the difference is statistically insignificant, indicating that estimates from the two models are asymptotically of equal variance.

Contrast these results with those for the analogous datasets generated using the ICLV model specification given by (30)–(33), shown in Table 5. Given that the ratio of the sample variances has an F distribution with both degrees of freedom equal to 99, the null hypothesis that the sample variances for the two models are equal can be rejected in favor of the alternative hypothesis that the sample variance for the ICLV model is less than the sample variance for the mixed logit model with a confidence of 95% when the test statistic is less than 0.72. As the table shows, the sample standard errors for the ICLV model are smaller than those for the reduced form mixed logit model, and in all cases, the difference is statistically significant. Therefore, it is fair to conclude that the variance of estimates from the ICLV model will be at least as low as the variance of estimates from the reduced form choice model, and in some cases, depending upon the model structure, it might actually be lower.

7. Relevance to practice and policy

The use of ICLV models for policy analysis has come under recent criticism, most notably by Chorus and Kroesen (2014), who have argued that implications usually derived from ICLV models that seek “to influence latent variables and as a consequence choice behavior… are not supported by the cross-sectional nature of the data in combination with the endogenous nature of the latent variables.” As with any econometric framework, ICLV models can be done well and they can be done

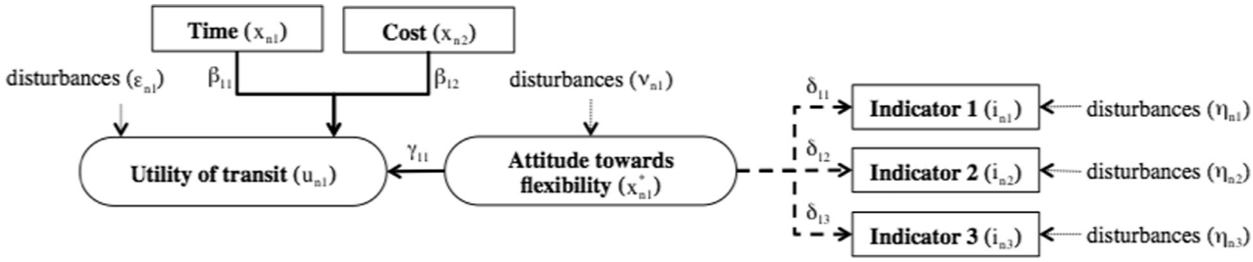

Fig. 7. A component of a hypothetical model of travel mode choice behavior.

poorly. Though the concerns raised by Chorus and Kroesen (2014) are legitimate, they are not unique to ICLV models. First, latent variables are no more likely to be endogenous than observable variables. Take, for example, the case of travel mode choice behavior, where the destination for a trip is often determined by the level-of-service of the transportation network ("if I were walking I may have gone to a different store to do groceries than if I were driving or taking the bus"), but travel mode choice models routinely treat choice of destination as an exogenous input. Second, not all latent variables are equally endogenous. Indeed, more mutable constructs like attitudes and perceptions may be more likely to be endogenous to the choice behavior of interest, due to learning effects, cognitive dissonance, etc., but constructs such as social norms and values that are stable over longer periods of time are much less likely to be influenced by short-term behavior. Third, while they may not be straightforward, methods do exist to control for endogenous variables in the context of discrete choice models, with or without latent variables (see, for example, Villas-Boas and Winer, 1999; Berry et al., 1998; Petrin and Train, 2010). And finally, the use of longitudinal datasets and dynamic frameworks could indeed help improve the state-of-the-practice with regards to choice models with latent variables, but the same is true for choice models without latent variables.

Much work still remains to be done to establish the benefits of the framework to practice and policy, but ICLV models do offer distinct advantages over choice models without latent variables. Unlike simpler choice models, ICLV models provide a mathematical framework for testing and applying complex theories of behavior, and lend structure and meaning to underlying sources of heterogeneity. Section 7.1 illustrates how ICLV models allow the analyst to quantify the impact of latent constructs on observable behavior through measures such as marginal rates of substitution and demand elasticities; and Section 7.2 discusses how ICLV models may be employed for scenario and policy analyses in ways in which choice models without latent variables would prove inadequate. The discussion contained within these subsections is by no means a comprehensive overview of the benefits that can be expected from the framework. The objective is to demonstrate through a series of concrete examples how findings from the framework may be leveraged for practice and policy in a manner that goes beyond the obvious.

7.1. Quantifying the effect of latent constructs

ICLV models enable the analyst to quantify the impact of latent constructs on observable behavior. The marginal rate of substitution is the most commonly used measure to compare the influence exerted by different explanatory variables. The marginal rate of substitution between any two explanatory variables, defined as the ratio of the marginal utility with respect to the two variables, is a measure of the trade-off that a decision-maker is willing to make between the two variables. Since the units of a latent variable are often unclear, marginal rates of substitution where one or both of the explanatory variables are latent can be difficult to interpret. However, a simultaneous comparison across the choice and measurement sub-models can provide a useful basis for quantifying and interpreting the latent variable's effect, relative to other variables. Take, for example, the component of the hypothetical model of travel mode choice behavior shown in Fig. 7, where the utility of public transportation and the measurement equation for the latent variable corresponding to the first indicator are given as follows:

$$u_{n1} = \beta_{11}x_{n1} + \beta_{12}x_{n2} + \gamma_{11}x^*_{n1} + \varepsilon_{n1} \tag{67}$$

$$i_{n1} = \delta_{11}x^*_{n1} + \eta_{n1} \tag{68}$$

where x_{n1} and x_{n2} denote the travel time and cost incurred by public transit; and x^*_{n1} denotes how much an individual cares about flexibility. Assume that the first measurement indicator is a Likert-scale statement that asks individuals to indicate on a five-point scale how important it is to them that a means of transport is available right away, from 'not at all important' to 'very important'. Say that we observe two individuals whose responses to the statement differ by one point, such that individual 1 cares less about the immediate availability of a travel mode than individual 2. Assuming that travel cost (x_{n2}) is measured in dollars, from (67) we know that a unit change in an individual's attitude towards flexibility has the same effect on the utility of public transit as a change in the cost of public transit by $\frac{\gamma_{11}}{\beta_{12}}$ dollars. Similarly, from (68) we know that a unit change in an individual's attitude towards flexibility changes the outcome to the measurement indicator by δ_{11} units (where the equation is implicitly treating the indicator as a continuous response variable). Putting these two pieces of information, and assuming further that $\beta_{12} \le 0$, $\gamma_{11} \le 0$ and $\delta_{11} \ge 0$, as one would expect in practice, ceteris paribus, individual 1 would be willing to pay $\frac{\gamma_{11}}{\beta_{12}\delta_{11}}$ dollars more to use public transport than individual 2. The effect of the latent variable can similarly be interpreted in terms of travel times, or using one of the other measurement indicators.

Another way to quantify the effect of a (latent) variable is through the calculation of demand elasticities. A decision-maker's elasticity of demand for a particular alternative with respect to an explanatory variable, latent or observed, is defined as the percentage change in the probability of choosing that alternative in response to a one percent change in the explanatory variable. Since elasticities are normalized for the unit of the explanatory variable, they are often easier to interpret with respect to latent variables than marginal rates of substitution. For example, in the case of the hypothetical model of travel mode choice behavior, if the choice model is logit then the elasticity of demand of decision-maker n for public transit with respect to attitude towards flexibility, denoted $e_{1x^*_{n1}}$, conditional on the latent variables, is given as follows:

$$e_{1x^*_{n1}} = \left(\frac{x^*_{n1}}{P(y_{n1} = 1|\mathbf{x_n}, \mathbf{x^*_n}; \mathbf{B}, \boldsymbol{\Gamma})} \right) \frac{\partial P(y_{n1} = 1|\mathbf{x_n}, \mathbf{x^*_n}; \mathbf{B}, \boldsymbol{\Gamma})}{\partial x^*_{n1}} = \gamma_{11} x^*_{n1} [1 - P(y_{n1} = 1|\mathbf{x_n}, \mathbf{x^*_n}; \mathbf{B}, \boldsymbol{\Gamma})] \tag{69}$$

where y_{n1} equals one if the decision-maker chose public transit, and zero otherwise; and $P(y_{n1} = 1|\mathbf{x_n}, \mathbf{x^*_n}; \mathbf{B}, \boldsymbol{\Gamma})$ is the choice model probability that the decision-maker chooses public transit. For a derivation of the above expression, the reader is referred to Train (2009, pp. 57–60). Since the latent variable is a random variable with a probability distribution function, we can take the expectation of Eq. (69) with respect to the latent variables, using either the structural or the measurement equation to derive the probability density function for the latent variables. Different from individual elasticities of demand, the aggregate elasticity of demand for a particular alternative with respect to an explanatory variable, latent or observed, is defined as the percentage change in the demand for the alternative across the sample population in response to a one percent change in the explanatory variable for all decision-makers in the sample population. Again, since the latent variable is a random variable with a probability distribution function, a one percent change in the latent variable may be interpreted as a one percent change in the mean of the latent variable, and aggregate elasticities may subsequently be computed using simulation and sample enumeration.

Marginal rates of substitution and demand elasticities provide useful ways of quantifying the effects of observable and latent constructs and comparing their relative influence. Unfortunately, when evaluating the effect of latent constructs on observable behavior, directionality and statistical significance have tended to dominate the discussion, and the magnitude of the effect has often been ignored. It is hoped that future studies report these quantitative measures more frequently.

7.2. Scenario and policy analyses

Discrete choice models are often employed to predict changes in the demand of different alternatives in response to changes in one or more explanatory variables. Scenarios that predict changes in observable alternative attributes or decision-maker characteristics can be simulated using choice models without latent variables, though forecasts may differ from those for the ICLV model in cases where parameter estimates from the reduced form choice model are believed to be biased. For example, say we wish to evaluate the impact of an ageing population on bicycle ownership, and we estimate an ICLV model given by (20)–(25) and the corresponding reduced form mixed logit model given by (26) and (27). When discussing omitted variable bias in Section 5.1, we had argued that parameter interpretation is determined by the assumed causal relationship between the latent and observable explanatory variables. Here we contend that differences in interpretation can lead to consequent differences in forecasts from the two models. Say that attitudes towards the environment and susceptibility to peer pressure are causally determined by a decision-maker's age. As the population ages over time, these latent variables will change, and consequently, so will the probability that a particular decision-maker owns a bicycle. The mixed logit model captures this effect directly through the composite parameter $\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21}$, while the ICLV model captures it indirectly through the influence of age on each of the latent constructs ($\gamma_{11}\alpha_{11}$ and $\gamma_{12}\alpha_{21}$) and some residual direct influence (β_{11}). However, both models yield the same estimate for the cumulative effect (τ_{11}), and forecasts from the two models should be identical. Say now that there is no causal relationship between the latent variables and age, only correlation. Changes in age over time should have no effect on the latent variables, and the effect on the probability of bicycle ownership should be mediated only through the parameter β_{11}. The ICLV model can be employed to simulate this process, but a mixed logit model cannot (because the latter can only identify τ_{11}, and not β_{11}). As a result, forecasts from the mixed logit model will differ from those from the ICLV model, and likely be biased. Analogous arguments can be made for cases where parameter estimates are biased due to measurement error.

Scenarios that predict a change in one or more latent explanatory variables cannot obviously be simulated using choice models without latent variables. For example, in the context of bicycle ownership, policy-makers may wish to know the response from a dedicated information campaign on the benefits of environmentalism that is expected to lead to an increase in mean environmental attitudes across the population of interest. Scenarios such as these are often employed by ICLV studies to demonstrate the benefits of the framework over simpler representations (see, for example, Daziano and Bolduc, 2013a). If the latent variable is measured using standardized indicators used by other studies in the past, they also provide some basis for saying how the latent variable might change in the future. For example, the International Social Survey Program (ISSP) has conducted cross-national surveys on, among other topics, environmental attitudes at varying points in time since 1985, including as recently as 2010. While some degree of contextualization is often unavoidable, the reliability

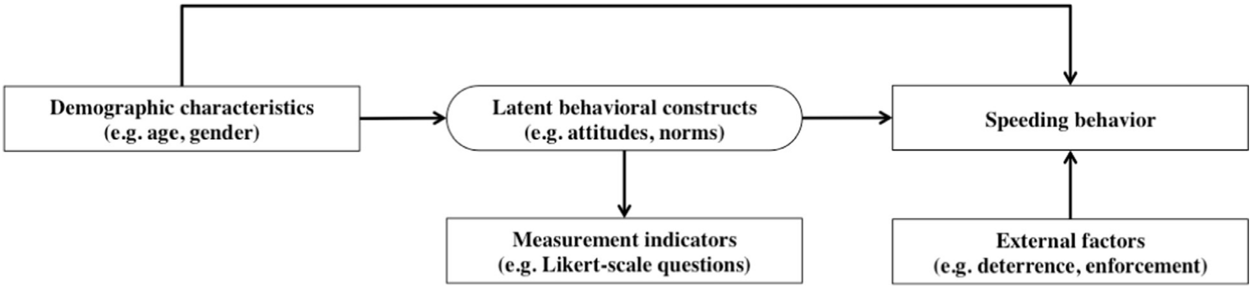

Fig. 8. Framework for a hypothetical model of risky speeding behavior.

and validity of standardized indicators has usually been established by previous research, and their use allows for easier comparison with past research and serves as a point of reference for future research. In addition, forecasts for standardized indicators are more likely to be available, which allows for the use of the measurement equation during model evaluation and prediction. Therefore, as far as possible, we recommend the use of standardized indicators for the measurement of latent variables.

Apart from scenario analysis, there are a number of other ways in which findings from ICLV models can be used for the design of policy. As an illustration, consider the simplified framework for a model of speeding behavior shown in Fig. 8, based loosely on the Theory of Planned Behavior (Azjen, 1991) that has been applied with great success to this particular context (see, for example, Stead et al., 2005; Elliott and Thomson, 2010; Scagnolari et al., 2015). Imagine that the authorities of some metropolitan region are planning a video campaign to discourage risky speeding behavior among drivers, and they wish to adapt scripts for the following two advertisements, taken from a 3-year media campaign in Scotland, as described in Stead et al. (2005):

Advertisement A: "A male driver in his 30 s was featured with two significant others likely to be particularly influential, a female partner and a male friend/colleague (Parker and Stradling, 2001). The ad begins with the female partner, at home, describing how 'he becomes a different person, totally unrecognizable'. The family is then shown in the car, with the driver speeding and his partner protesting as the speed of the car jolts their young son's neck, in the back of the car. She wishes her partner could 'see things through her eyes'. A male friend/colleague of the driver then addresses the camera, also expressing his disapproval of his friend's 'boy racer' behavior. The two friends are then shown in the car, where the friend is annoyed because he spills juice down his sweater as a result of the driver's hasty acceleration. The ad closes with the driver alone in the car and the voice-over 'Put yourself in the passenger seat. If you don't, others won't'."

Advertisement B: "Three different drivers and driving scenarios are depicted in the ad, with a nursery rhyme-style voiceover on the theme of the children's game 'Simon Says'. The three scenarios illustrate the pressure of being in a flow of traffic going at 40 m.p.h. in a 30 m.p.h. limit, the pressure of being late for work, and the more direct pressure of an impatient driver (a 'white van man') behind. In the latter scenario, the driver nearly hits a cyclist as a result of being distracted and pressurized by the white van driver. The ad closes with the strapline 'Be your own man'."

Advertisement A is specifically designed to target drivers' normative beliefs, and advertisement B is specifically designed to target drivers' perceived behavioral control. There are three broadly generalizable ways in which the design and use of these advertisements can be tailored based on findings from an ICLV model of speeding behavior in the metropolitan region of interest. First, by quantifying the relative impact of different behavioral constructs, the model can help identify which constructs to focus on in the design of intervention campaigns. For example, if the study finds that the influence of normative beliefs on speeding intention and behavior far outweighs the influence of perceived behavioral control, then the allocation of air time should favor advertisement A over advertisement B, and vice versa. The exact allocation itself could be determined such that it is roughly proportional to the relative impacts of the two constructs. Second, by quantifying the relationship between the demographic characteristics and the behavioral constructs, the study can help identify which messages to target at what segments of the population. For example, if the study finds that women are more likely to speed because of lower perceived behavioral control than men, then the script for advertisement B should be adapted to target women (change the strapline, choose a female lead actor, etc.). And third, by quantifying the association between each of the behavioral constructs and the indicators used to measure them, the study can help identify which specific points to focus on during an advertisement designed around a particular behavioral construct. For example, if the study finds that perceived behavioral control is a strong predictor of how easy or difficult drivers find to respect the speed limit when they are in a hurry, but a weak predictor of how easy or difficult drivers find to respect the speed limit when other drivers are driving too fast, then the first scenario from advertisement B (the pressure of being in a flow of traffic going at 40 m.p.h. in a 30 m.p.h. limit) may need either to be replaced by a different scenario that focuses on a more relevant point, or removed to give more air time to the second scenario ("the pressure of being late for work").

This is just one example of how findings from an ICLV model can be used for the design of information campaigns that seek to change behavior through their impact on 'soft' constructs, like attitudes, perceptions, etc. Unfortunately, many studies in the past that have employed ICLV models have failed to conclusively demonstrate what tangible gains might be had from adopting the framework. Most have tended to simplify significantly the cognitive theories motivating the use of ICLV models, and much of the behavioral richness captured originally in these theories through the complex interplay between different latent psychological constructs has often been lost as a consequence of these simplifications. While such work is appropriate and valuable as the methods are being developed, the ICLV methodology is now mature enough that the discussion needs to be redirected to emphasize added value. Studies need to show how the greater insights into the decision-making process offered by the ICLV model can be used to inform policy and generate forecasts in unobvious ways that would not be possible using choice models without latent variables.

8. Conclusions

For most part of the last century, discrete choice models have based their representation of the decision-making process upon the neoclassical abstraction of decision-makers as rational self-interested actors engaged in a continuous process of evaluating the costs and benefits associated with any decision in the marketplace as they strive to maximize their personal wellbeing. Early applications of the framework were limited by their emphasis on the direct impact of observable variables on the decision-making process and their concurrent indifference to the influence exerted by latent variables underlying the process. As a matter of convenience, the task of breaking open these black-box representations was left to studies in the behavioral and social sciences. Much progress has been made in these fields over the last several decades. Detailed survey instruments, such as Likert-scale items asking respondents to rate their level of agreement or disagreement on a symmetric agree–disagree scale, have been developed to measure the many different latent constructs postulated to motivate observable behavior. Rich theories of cognitive decision-making have been proposed and tested across a wide spectrum of disciplines that include travel demand analysis, healthcare, education, marketing sciences, etc. (see, for example, Triandis, 1977; Homer and Kahle, 1988; Azjen, 1991). However, the reliance of most of these studies on some variation of the Structural Equations Modeling (SEM) approach precluded integration with studies in econometrics and related disciplines for a long time. Though SEM is particularly suited to the estimation of models with causal relationships between multiple observable and latent variables, it differs substantially from the neoclassical framework of random utility maximization used by discrete choice models. The ICLV model reconciles these methodological differences by combining the factor analytic approach used by behavioral and social scientists with discrete choice methods popularly employed by econometricians. By allowing for the explicit incorporation of psychometric data and latent constructs within existing representations of the decision-making process, the ICLV model frees the analyst from restrictions arising out of simplifying assumptions made by earlier models.

Notwithstanding these benefits, the practical value of the ICLV model has been questioned (see, for example, Daly and Hess, 2012; Chorus and Kroesen, 2014; Chorus and Hess, 2015). This study was born from the discovery that an ICLV model can in many cases be reduced to a choice model without latent variables that fits the choice data at least as well as the original ICLV model from which it was obtained. The failure of studies in the past to recognize this retrospectively simple truth raised concern over the other benefits that were claimed with regards to the use of ICLV models. With the objective of addressing some of these concerns, this study undertook a measured reexamination of the benefits of ICLV models vis-à-vis more traditional choice models without latent variables. Our findings may be summarized through five key points. First, the ICLV model can lead to an improvement in the analyst's ability to predict outcomes to the choice data in cases where forecasts for the measurement indicators are expected to be available and the observable explanatory variables are poor predictors of the latent constructs, or in cases where it can justifiably be shown that the reduced form choice model specification would not have been arrived at in the absence of latent variables to guide the process of model development. Second, the ICLV model allows for the identification of structural relationships between variables that could not otherwise be identified by a choice model without latent variables, providing a mathematical framework for testing and applying complex theories of behavior, and lending structure and meaning to underlying sources of heterogeneity. Third, under traditional assumptions of causality, parameter estimates from the ICLV model and the reduced form model should be asymptotically equal, but in cases where these assumptions may be relaxed, the ICLV model can help correct for bias arising from omitted variables and measurement error. Fourth, the ICLV model can in some cases reduce the variance of the parameter estimates, and consequently, increase the precision of policy outputs such as marginal rates of substitution and aggregate forecasts. And finally, the ICLV model can quantify the impact of latent constructs on observable behavior through measures such as willingness to pay and demand elasticities, predict changes in behavior as a function of changes in one or more such latent constructs over time, and abet the design of policies seeking to change behavior, all in ways that would not be possible using simpler choice models without latent variables.

For the benefit of future research, we propose a general process of evaluation that studies can use to demonstrate what tangible gains, if any, might be had from adopting the ICLV model framework in a particular context. When comparing an ICLV model with the appropriate reduced form choice model, if a study can show that at least one of the following four criteria are met, then the use of the ICLV model in that context ought to be justified (as a reminder, we include references to the corresponding sections in this paper that discussed each of these criteria):

- The structure imposed by the ICLV model results in a reduced form choice model specification that may not have been considered in the absence of latent variables to guide the process of model development (refer to Section 3).

- The greater insights into the decision-making process offered by the ICLV model can be demonstrably used to inform practice and policy in unobvious ways that would not be possible using the reduced form model (refer to Sections 4 and 7).
- Estimation results from the ICLV model lead the analyst to interpret the effect of explanatory variables on choice outcomes differently than the reduced form model (refer to Section 5). Here, the analyst must carefully verify that the intention to correct for bias has actually made a difference.
- The inclusion of additional information in the form of measurement indicators produces model outputs, such as demand elasticities and market predictions, that have smaller standard errors than corresponding outputs from the reduced form model (refer to Section 6). We leave it to the analyst to determine whether a gain in precision is worth the additional effort required to develop an ICLV model.

If used appropriately, the ICLV model can be a powerful tool to have in an ever expanding toolbox of models. But as analysts who develop statistical models of disaggregate behavior, we need to be more conscientious in terms of the data that we collect, more thoughtful with regards to the design of the model framework, and more creative in illustrating its value to practitioners and policy-makers.

Acknowledgments

We wish to thank Andrew Daly, Stephane Hess and Caspar Chorus for giving us the opportunity to present findings from earlier iterations of this study at workshops held during the 13th International Conference on Travel Behavior Research in Toronto, July 2012 and the 4th International Choice Modelling Conference in Austin, May 2015. Feedback received during the workshop helped us improve the presentation of our ideas. This study has greatly profited from additional conversations with Moshe Ben-Akiva, Michel Bierlaire, David Brownstone, David Bunch, Ricardo Daziano, Joffre Swait, and many others. And finally, we wish to express our gratitude for the constructive comments and criticisms proffered by Chandra Bhat and five anonymous reviewers. Of course, any mistakes are the fault of the authors alone.

References

Abou-Zeid, M., 2009. Measuring and Modeling Activity and Travel Well-being (Ph.D. dissertation). Massachusetts Institute of Technology.
Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C., Hess, S., 2010. Attitudes and value of time heterogeneity. Applied Transport Economics – A Management and Policy Perspective. De Boeck Publishing, pp. 523–545.
Abou-Zeid, M., Ben-Akiva, M., 2014. Hybrid choice models. In: Hess, S., Daly, A. (Eds.), Handbook of Choice Modelling. Edward Elgar, Cheltenham.
Anable, J., 2005. 'Complacent car addicts' or 'aspiring environmentalists'? Identifying travel behaviour segments using attitude theory. Transport Policy 12, 65–78.
Ashok, K., Dillon, W.R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. Journal of Marketing Research 39 (1), 31–46.
Azjen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision 50 (2), 179–211.
Bamberg, S., Schmidt, P., 2001. Theory-driven evaluation of an intervention to reduce the private car-use. Journal of Applied Social Psychology 31 (6), 1300–1329.
Banerjee, I., Kanafani, A., Walker, J., 2012. Motorized vehicle purchase in Surat, India. Transportation Research Record 2317, 24–31.
Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12 (1), 9–24.
Ben-Akiva, M., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H.S. (Ed.), In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges. Elsevier, Amsterdam, pp. 431–470.
Bergantino, A.S., Bierlaire, M., Catalano, M., Migliore, M., Amoroso, S., 2013. Taste heterogeneity and latent preferences in the choice behaviour of freight transport operators. Transport Policy 30, 77–91.
Berry, S., Levinsohn, J., Pakes, A., 1998. Differentiated Products Demand Systems From a Combination of Micro and Macro Data: The New Car Market, Report no. W6481. National Bureau of Economic Research.
Bhat, C.R., Dubey, S.K., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. Transportation Research Part B 67, 68–85.
Bhat, C.R., Dubey, S.K., Nagel, K., 2015. Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. Transportation Research Part B 78, 341–363.
Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference. Ascona, Switzerland.
Chorus, C.G., Kroesen, M., 2014. On the (im-)possibility of deriving transport policy implications from hybrid choice models. Transport Policy 36, 217–222.
Chorus, C.G., Hess, S., 2015. Behavioral choice modelling: much ado about nothing? In: Proceedings of the Workshop at the 4th International Choice Modelling Conference. Austin, United States.
Córdoba, J.E., Jaramillo, G.P., 2012. Inclusion of the latent personality variable in multinomial logit models using the 16PF psychometric test. Procedia – Social and Behavioral Sciences 54, 169–178.
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 behaviour. Transportation 39 (2), 267–297.
Daly, A., Hess, S., 2012. Latent reality: recognizing the limits of observation. In: Proceedings of the Workshop at the 13th International Conference on Travel Behavior Research. Toronto, Canada.
Daziano, R.A., Bolduc, D., 2013a. Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. Transportmetrica A: Transport Science 9 (1), 74–106.
Daziano, R.A., Bolduc, D., 2013b. Covariance, identification, and finite-sample performance of the MSL and Bayes estimators of a logit model with latent attributes. Transportation 40 (3), 647–670.
Daziano, R.A., 2015. Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. Transportation Research Part B 76, 1–26.
Domarchi, C., Tudela, A., González, A., 2008. Effect of attitudes, habit and affective appraisal on mode choice: an application to university workers. Transportation 35 (5), 585–599.
Elliott, M.A., Thomson, J.A., 2010. The social cognitive determinants of offending drivers' speeding behaviour. Accident Analysis & Prevention 42 (6), 1595–1605.

Galdames, C., Tudela, A., Carrasco, J.A., 2011. Exploring the role of psychological factors in mode choice models by a latent variables approach. Transportation Research Record 2230, 68–74.
Gärling, T., Fujii, S., Gärling, A., Jakobsson, C., 2003. Moderating effects of social value orientation on determinants of pro-environmental behavior intention. Journal of Environmental Psychology 23, 1–9.
Glerum, A., 2014. Static and dynamic mathematical models of behavior. PhD Thesis. École polytechnique fédérale de Lausanne.
Guevara, C.A., 2010. Endogeneity and Sampling of Alternatives in Spatial Choice Models (Ph.D. dissertation). Massachusetts Institute of Technology.
Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transportation Research Part A 82, 240–254.
Gujarati, D.N., 2004. Basic Econometrics, 4th edition The McGraw-Hill Companies.
Guttman, L., 1955. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. British Journal of Statistical Psychology 22 (1), 79–81.
Hess, S., Beharry-Borg, N., 2012. Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in Tobago. Environmental and Resource Economics 52 (1), 109–131.
Homer, P.M., Kahle, L.R., 1988. A structural equation test of the value-attitude-behavior hierarchy. Journal of Personality and Social Psychology 54 (4), 638–646.
Jöreskog, K.G., 1978. Structural analysis of covariance and correlation matrices. Psychometrika 43 (4), 443–477.
Kamargianni, M., Polydoropoulou, A., 2013. Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. Transportation Research Record 2382, 151–161.
Kassahun, H.T., Thorsen, B.J., Jacobsen, J.B., 2015. Accounting for spatial-social influence in environmental management decisions and endogeneity in a panel discrete choice model. In: Proceedings of the 4th International Choice Modeling Conference. Austin, US.
Kirk, J.L., 2010. Sustainable Environments and Pro-Environmental Behavior (Master's thesis). University of Nebraska, Lincoln.
Kløjgaard, M.E., Hess, S., 2014. Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: Testing the benefits of a hybrid choice model approach. Social Science & Medicine 114, 138–150.
Maldonado-Hinarejos, R., Sivakumar, A., Polak, J.W., 2014. Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. Transportation 41 (6), 1287–1304.
McFadden, D.L., 1986. The choice theory approach to marketing research. Marketing Science 5 (4), 275–297.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of applied Econometrics 15 (5), 447–470.
Muthén, L.K., Muthén, B.O., 2011. Mplus User's Guide, Sixth edition Muthén & Muthén, Los Angeles, CA.
Parker, D., Stradling, S., 2001. Influencing Driver Attitudes and Behaviour, Road Safety Research Report 17. Department of the Environment, Transport and the Regions, London.
Paulssen, M., Temme, D., Vij, A., Walker, J.L., 2014. Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transportation 41 (4), 873–888.
Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. Journal of Marketing Research 47 (1), 3–13.
Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., Lee, A., 2011. Importance of traveler attitudes in the choice of public transportation to work: findings from the Regional Transportation Authority Attitudinal Survey. Transportation 38 (4), 643–661.
Raveau, S., Daziano, R.A., Yáñez, M.F., Bolduc, D., Ortúzar, J., de, D., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: Some new findings. Transportation Research Record 2156, 131–139.
Raveau, S., Yáñez, M.F., Ortúzar, J.deD., 2012. Practical and empirical identifiability of hybrid discrete choice models. Transportation Research Part B 46 (10), 1374–1383.
Scagnolari, S., Walker, J., Maggi, R., 2015. Young drivers' night-time mobility preferences and attitude toward alcohol consumption: a hybrid choice model. Accident Analysis & Prevention 83, 74–89.
Stead, M., Tagg, S., MacKintosh, A.M., Eadie, D., 2005. Development and evaluation of a mass media theory of planned behaviour intervention to reduce speeding. Health Education Research 20 (1), 36–50.
Steiger, J.H., 1979. Factor indeterminacy in the 1930′s and the 1970′s some interesting parallels. Psychometrika 44 (2), 157–167.
Steinberg, L., Monahan, K.C., 2007. Age differences in resistance to peer influence. Developmental Psychology 43 (6), 1531–1543.
Thorhauge, M., Rich, J., Cherchi, E., 2015. Departure Time Choice: Modelling Individual Preferences, Intention and Constraints (Doctoral dissertation). Technical University of Denmark, Department of Transport, Traffic Modelling.
Train, K.E., McFadden, D.L., Goett, A.A., 1987. Consumer attitudes and voluntary rate schedules for public utilities. The Review of Economics and Statistics 69 (3), 383–391.
Train, K.E., 2009. Discrete Choice Models with Simulation. Cambridge University Press, Cambridge.
Triandis, H.C., 1977. Interpersonal Behaviour. Brooks/Cole, Monterey, CA.
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.
Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Management Science 45 (10), 1324–1338.
Walker, J., Li, J., Srinivasan, S., Bolduc, D., 2010. Travel demand models in the developing world: correcting for measurement errors. Transportation Letters 2 (4), 231–243.
Wang, T., Chen, C., 2012. Attitudes, mode switching behavior, and the built environment: A longitudinal study in the Puget Sound Region. Transportation research part A: policy and practice 46 (10), 1594–1607.
Williams, H.C.W.L., Ortúzar, J., de, D., 1982. Behavioural theories of dispersion and the misspecification of travel demand models. Transportation Research Part B 16 (3), 167–219.
Yáñez, M.F., Raveau, S., Ortúzar, J., de, D., 2010. Inclusion of latent variables in mixed logit models: modelling and forecasting. Transportation Research Part A 44 (9), 744–753.