Statistical properties of Integrated Choice and Latent Variable models

WORKING PAPER

June 9, 2015

Akshay Vij (corresponding author)
Institute for Choice
University of South Australia
Level 13, 140 Arthur Street
North Sydney, NSW 2060
vij.akshay@gmail.com
+61 08 8302 0817

Joan L. Walker
University of California at Berkeley
111 McLaughlin Hall
Berkeley, CA 94720-1720
joanwalker@berkeley.edu
+1 510 642 6897

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. Though the ICLV model has been employed extensively by studies across a wide spectrum of disciplines, the value of the framework has remained unclear. On one hand, ICLV models allow for the proper integration of psychometric data and provide a framework with which to test the influence of latent variables, such as attitudes and perceptions, on observable behavior. On the other, questions have been raised regarding their value to econometricians, practitioners and policy-makers. This study undertakes a systematic evaluation of the statistical properties of the ICLV framework, and how they compare to a reduced form choice model without latent variables. We derive easily generalizable analytical proofs regarding the statistical benefits, or lack thereof, of ICLV models over choice models without latent variables and use synthetic datasets to validate any conclusions drawn from the analytical proofs. In terms of goodness of fit and the consistency of parameter estimates, we find that ICLV models do not offer improvements over reduced form choice models without latent variables. However, ICLV models allow for the identification of structural relationships between observable variables that could not be identified using choice models without latent variables, and parameter estimates from the ICLV model are shown to be potentially more efficient than equivalent estimates from a reduced form choice model without latent variables. Given the limited nature of these practical benefits, we argue that studies that use ICLV models need to show either that 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, or that 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.

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 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 expanded on by 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 (Prato et al., 2012), car ownership (Daziano and Bolduc, 2013), air travel (Fleischer et al., 2012), freight (Ben-Akiva et al., 2008), etc.

Though much progress has been made in terms of model development and estimation (see, for example, Bhat and Dubey, 2014 and Daziano, 2015 for recent methodological advances on the subject), existing studies have failed to demonstrate conclusively 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. Does an ICLV model fit the data better than a choice model without latent variables? Are parameter estimates from the ICLV model more efficient? Both sides of the debate have their proponents, but a clear verdict remains elusive.

The objective of this study is to evaluate systematically the statistical properties of the ICLV model framework, and how they compare with a more traditional choice model without latent variables based on four criteria: goodness of fit with respect to the choice data, identification of structural relationships between variables, and the consistency and efficiency of parameter estimates. The study derives analytical proofs regarding the benefits, or lack thereof, of ICLV models over choice models without latent variables for each of the four criteria, and uses synthetic datasets to validate any conclusions drawn from the analytical proofs.

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 consistency and efficiency of parameter estimates obtained by the two model forms, respectively; and Section 7 concludes the paper with a summary of key findings and a discussion of potential directions for future research.

2. The ICLV Model Framework

Figure 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 may be represented using the following set of four equations:

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

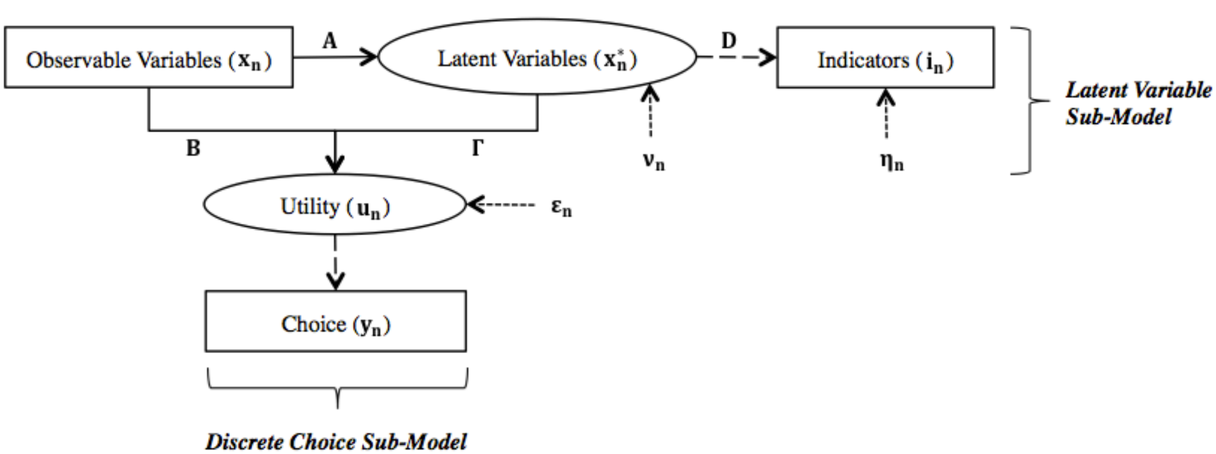
$$\mathbf{x_n^*} = \mathbf{A}\mathbf{x_n} + \boldsymbol{\nu}_\mathbf{n} \quad (2)$$

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

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

, where $\mathbf{u_n}$ is the (J×1) vector of utilities of each of the J alternatives, as perceived by decision-maker n, $\mathbf{x_n}$ is the (K×1) vector of observable explanatory variables and $\mathbf{x_n^*}$ is the (M×1) vector of latent explanatory variables, $\mathbf{B}$ and $\mathbf{\Gamma}$ are the (J×K) and (J×M) matrices of model parameters denoting sensitivities to the observable and latent variables, respectively, and $\boldsymbol{\varepsilon}_\mathbf{n}$ is the (J×1) vector denoting the stochastic component of the utility specification; $\mathbf{A}$ is the (M×K) matrix of model parameters denoting the structural relationship between the latent and observable variables, and $\boldsymbol{\nu}_\mathbf{n}$ is the (M×1) vector denoting the stochastic component of that relationship; $\mathbf{i_n}$ is the (R×1) vector of indicators used to measure the latent variables, assumed without loss of generality to represent deviations from the mean, $\mathbf{D}$ is the (R×M) matrix of model parameters denoting the sensitivities of the measurement indicators to the latent variables, and $\boldsymbol{\eta}_\mathbf{n}$ is the (R×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.

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



For the sake of notational simplicity, we've 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 $\boldsymbol{\Omega}_\mathbf{n}$, 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. 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 equation (4) to yield the following functional form:

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

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

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

With regards to the measurement indicators, we've 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 isn't 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 (though this assumption too can be relaxed without loss of generality). Under these assumptions, conditional on the latent variables, the probability distribution function associated with the measurement indicators may be formulated as follows:

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

, 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 results in a latent class choice model (LCCM) and a parametric formulation results in the ICLV model, the subject of this paper. Therefore, we will be limiting our attention to the latter case. We assume further that the latent variable can be represented using a linear in parameters formulation, as given by equation (2). While this is usually the case in the literature, our results hold under more general parametric formulations as well. 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:

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

Taking advantage of the conditional independence of the choice and measurement indicators and marginalizing over the distribution of the latent variables, equations (6), (7) and (8) may be combined to yield the joint unconditional probability distribution function for the choice and measurement indicators as follows:

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

Equation (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, 2013 and Bhat and Dubey, 2014). Over the course of the subsequent sections, we examine and compare the different statistical 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.

3. Goodness of Fit

The goodness of fit of a statistical model is a measure of how well the model explains observable data. For discrete choice models, with or without latent variables, goodness of fit is a measure of the model's ability to predict outcomes to the choice indicators, usually defined as some function of the likelihood evaluated at the parameter estimates. 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 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. Section 3.1 derives the analytical proof for our assertion; Sections 3.2, 3.3 and 3.4 validate the proof using synthetic datasets; and Section 3.5 summarizes key findings.

3.1 Analytical Proof of Model Equivalence

In order to facilitate a comparison between choice models with and without latent variables, we must necessarily restrict our attention to the likelihood of observing the vector of choices $\mathbf{y}$, which for the ICLV model may be derived by marginalizing equation (9) over the vector of indicators $\mathbf{i}_n$ as follows:

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}_n|\mathbf{x}_n;\mathbf{B},\boldsymbol{\Gamma},\mathbf{A},\boldsymbol{\Phi}) &= \int_{\mathbf{i}} f_{\mathbf{y},\mathbf{i}}(\mathbf{y}_n,\mathbf{i}_n|\mathbf{x}_n;\mathbf{B},\boldsymbol{\Gamma},\mathbf{D},\boldsymbol{\Psi},\mathbf{A},\boldsymbol{\Phi})d\mathbf{i}_n \\ &= \int_{\mathbf{i}}\int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y}_n|\mathbf{x}_n,\mathbf{x}_n^*;\mathbf{B},\boldsymbol{\Gamma})f_{\mathbf{i}}(\mathbf{i}_n|\mathbf{x}_n,\mathbf{x}_n^*;\mathbf{D},\boldsymbol{\Psi})f_{x^*}(\mathbf{x}_n^*|\mathbf{x}_n;\mathbf{A},\boldsymbol{\Phi})d\mathbf{x}_n^*\,d\mathbf{i}_n \\ &= \int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y}_n|\mathbf{x}_n,\mathbf{x}_n^*;\mathbf{B},\boldsymbol{\Gamma})f_{x^*}(\mathbf{x}_n^*|\mathbf{x}_n;\mathbf{A},\boldsymbol{\Phi})\left(\int_{\mathbf{i}} f_{\mathbf{i}}(\mathbf{i}_n|\mathbf{x}_n,\mathbf{x}_n^*;\mathbf{D},\boldsymbol{\Psi})d\mathbf{i}_n\right)d\mathbf{x}_n^* \\ &= \int_{\mathbf{x}^*} f_{\mathbf{y}}(\mathbf{y}_n|\mathbf{x}_n,\mathbf{x}_n^*;\mathbf{B},\boldsymbol{\Gamma})f_{x^*}(\mathbf{x}_n^*|\mathbf{x}_n;\mathbf{A},\boldsymbol{\Phi})d\mathbf{x}_n^* \end{aligned} \tag{10}$$

Since the likelihood function contains an integral that does not have a closed form solution, it may be approximated using the following simulator:

$$\mathrm{f_y}(\mathbf{y_n}|\mathbf{x_n};\mathbf{B},\mathbf{\Gamma},\mathbf{A},\mathbf{\Phi}) \approx \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathrm{f_y}(\mathbf{y_n}|\mathbf{x_n},\mathbf{x^*_{nt}};\mathbf{B},\mathbf{\Gamma}) \tag{11}$$

, where $\mathbf{x^*_{nt}}$ represents the $\mathrm{t^{th}}$ draw from $\mathrm{f_{x^*}}(\mathbf{x^*_n}|\mathbf{x_n};\mathbf{A},\mathbf{\Phi})$ and T is the number of draws. For more information on simulation methods, the reader is referred to Train (2009). The proof is quite straightforward. We begin by substituting equation (2), the structural equation of the latent variables, in equation (11):

$$\begin{aligned}
\mathrm{f_y}(\mathbf{y_n}|\mathbf{x_n};\mathbf{B},\mathbf{\Gamma},\mathbf{A},\mathbf{\Phi}) &\approx \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathrm{f_y}(\mathbf{y_n}|\mathbf{x_n},\mathbf{x^*_{nt}};\mathbf{B},\mathbf{\Gamma}) \\
&= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\prod_{j=1}^{\mathrm{J}}\left[\frac{\exp\left(\boldsymbol{\beta}_{\mathbf{j*}}\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j*}}\mathbf{x^*_{nt}}\right)}{\sum_{j'=1}^{\mathrm{J}}\exp\left(\boldsymbol{\beta}_{\mathbf{j'*}}\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j'*}}\mathbf{x^*_{nt}}\right)}\right]^{y_{nj}} \\
&= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\prod_{j=1}^{\mathrm{J}}\left[\frac{\exp\left(\boldsymbol{\beta}_{\mathbf{j*}}\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j*}}(\mathbf{A}\mathbf{x_n}+\boldsymbol{\nu}_{\mathbf{nt}})\right)}{\sum_{j'=1}^{\mathrm{J}}\exp\left(\boldsymbol{\beta}_{\mathbf{j'*}}\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j'*}}(\mathbf{A}\mathbf{x_n}+\boldsymbol{\nu}_{\mathbf{nt}})\right)}\right]^{y_{nj}} \\
&= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\prod_{j=1}^{\mathrm{J}}\left[\frac{\exp\left((\boldsymbol{\beta}_{\mathbf{j*}}+\boldsymbol{\gamma}_{\mathbf{j*}}\mathbf{A})\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j*}}\boldsymbol{\nu}_{\mathbf{nt}}\right)}{\sum_{j'=1}^{\mathrm{J}}\exp\left((\boldsymbol{\beta}_{\mathbf{j'*}}+\boldsymbol{\gamma}_{\mathbf{j'*}}\mathbf{A})\mathbf{x_n}+\boldsymbol{\gamma}_{\mathbf{j'*}}\boldsymbol{\nu}_{\mathbf{nt}}\right)}\right]^{y_{nj}}
\end{aligned} \tag{12}$$

, where $\boldsymbol{\nu}_{\mathbf{nt}}$ is the $\mathrm{t^{th}}$ draw from $\mathrm{N}(\mathbf{0},\mathbf{\Phi})$. At this stage, we have eliminated the vector of latent explanatory variables $\mathbf{x^*_{nt}}$. What we have is the simulated probability of observing the vector of choice indicators $\mathbf{y_n}$ as a function solely of the vector of observable explanatory variables $\mathbf{x_n}$. Introduce the $(\mathrm{J}\times\mathrm{K})$ matrix of parameters $\mathbf{T}=\mathbf{B}+\mathbf{\Gamma}\mathbf{A}$, such that the $\mathrm{j^{th}}$ row of $\mathbf{T}$, $\boldsymbol{\tau}_{\mathbf{j*}}=\boldsymbol{\beta}_{\mathbf{j*}}+\boldsymbol{\gamma}_{\mathbf{j*}}\mathbf{A}$. Introduce the $(\mathrm{J}\times 1)$ vector of random variables $\boldsymbol{\omega}_{\mathbf{n}}=\mathbf{\Gamma}\boldsymbol{\nu}_{\mathbf{n}}$, distributed normally across alternatives with a mean vector of zeros and covariance matrix $\mathbf{Z}=\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}^{\mathrm{T}}$, such that $\boldsymbol{\omega}_{\mathbf{nt}}$ is the $\mathrm{t^{th}}$ draw from $\mathrm{N}(\mathbf{0},\mathbf{Z})$ and ω_{ntj} is the $\mathrm{j^{th}}$ element of $\boldsymbol{\omega}_{\mathbf{nt}}$. Substituting these new parameters in equation (12), we get:

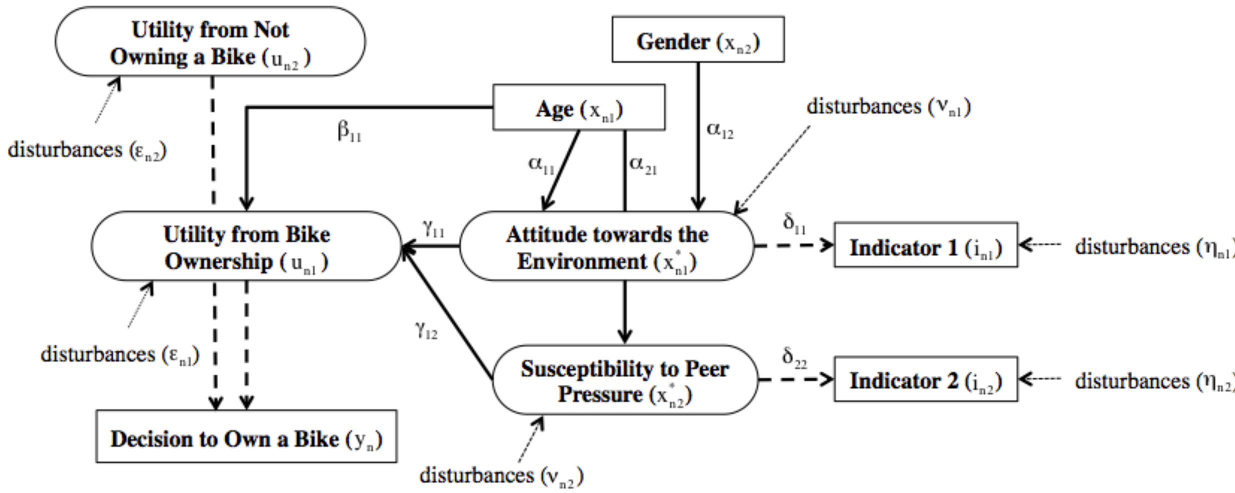
$$\mathrm{f_y}(\mathbf{y_n}|\mathbf{x_n};\mathbf{T},\mathbf{Z}) \approx \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\prod_{j=1}^{\mathrm{J}}\left[\frac{\exp\left(\boldsymbol{\tau}_{\mathbf{j*}}\mathbf{x_n}+\omega_{\mathrm{ntj}}\right)}{\sum_{j'=1}^{\mathrm{J}}\exp\left(\boldsymbol{\tau}_{\mathbf{j'*}}\mathbf{x_n}+\omega_{\mathrm{ntj'}}\right)}\right]^{y_{nj}} \tag{13}$$

But equation (13) is simply the simulated likelihood function for a mixed logit model without latent variables and with the following utility specification:

$$\mathbf{u_n}=\mathbf{T}\mathbf{x_n}+\boldsymbol{\omega}_{\mathbf{n}}+\boldsymbol{\varepsilon}_{\mathbf{n}} \tag{14}$$

Therefore, for any values of the model parameter vectors $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{A}$ and $\mathbf{\Phi}$ 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 $\mathbf{y_n}$. In deriving the proof, we have assumed that the discrete choice sub-model has a multinomial logit kernel. In cases where the 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. Analogous proofs can be derived for these cases following the method described here for the multinomial logit kernel.

Figure 2: A hypothetical model of bicycle ownership



3.2 Monte Carlo Experiment I

We validate the result using synthetic data generated using 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 aren't specific to any one dataset. For the purpose of the experiment, we construct a hypothetical model of bicycle ownership, as illustrated in Figure 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:

$$u_{n1} = \beta_{11}x_{n1} + \beta_{13} + \gamma_{11}x^*_{n1} + \gamma_{12}x^*_{n2} + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim GEV(0,1,0) \tag{15}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim GEV(0,1,0) \tag{16}$$

$$x^*_{n1} = \alpha_{11}x_{n1} + \alpha_{12}x_{n2} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{17}$$

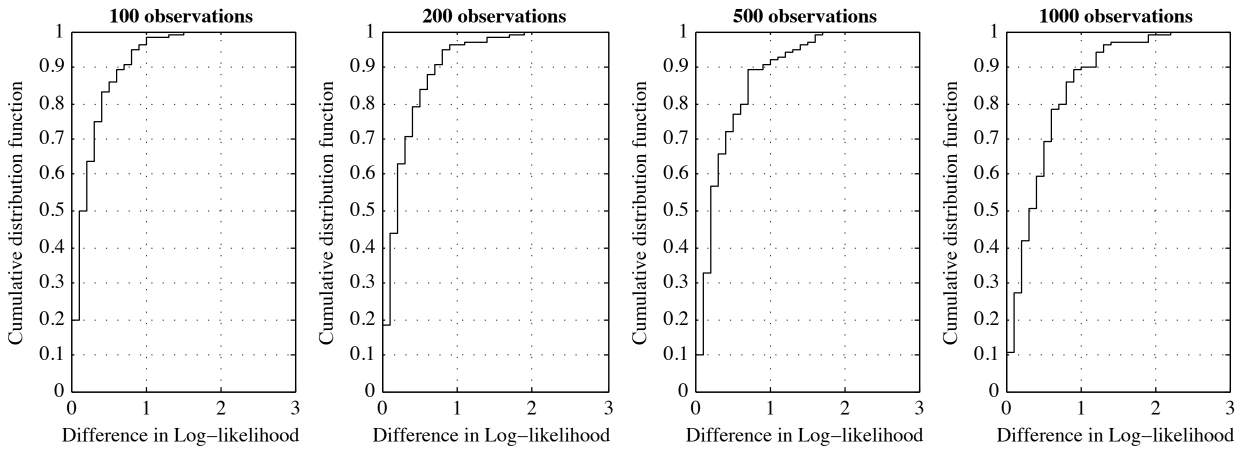
$$x^*_{n2} = \alpha_{21}x_{n1} + \alpha_{23} + \nu_{n2}, \quad \nu_{n2} \sim N(0, \phi_{22}) \tag{18}$$

$$i_{n1} = x^*_{n1} + \eta_{n1}, \quad \eta_{n1} \sim N(0,1) \tag{19}$$

$$i_{n2} = x^*_{n2} + \eta_{n2}, \quad \eta_{n2} \sim N(0,1) \tag{20}$$

Consistent with practices in the literature, we've selected a multinomial logit kernel for the discrete choice sub-model, we've assumed that the covariance matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are diagonal matrices, and we've set the location and scale for each of the latent variables through the measurement equations. Synthetic datasets are generated

Figure 3: A plot of the cumulative distribution function over 100 datasets of the difference in log-likelihood at convergence between the ICLV model given by equations (15)-(20) and the mixed logit model given by equations (21) and (22)



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 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.

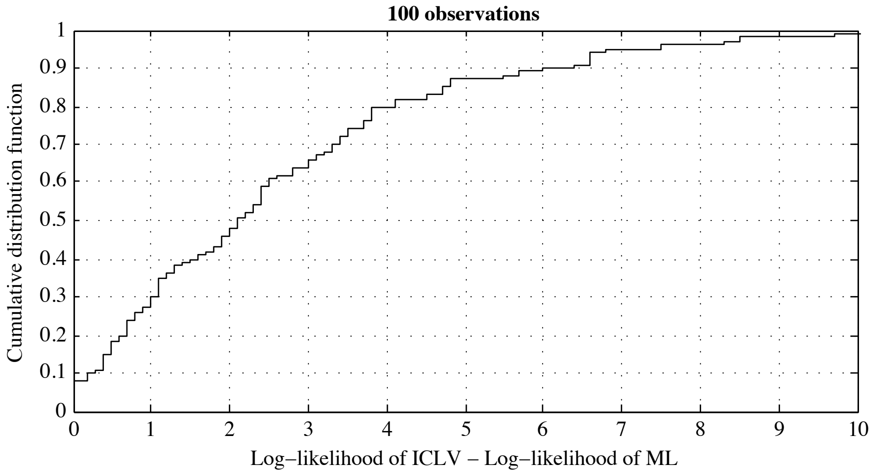
Substituting equations (17) and (18) in equations (15) and (16), we get the following form for the reduced form mixed logit model without latent variables:

$$u_{n1} = \tau_{11}x_{n1} + \tau_{12}x_{n2} + \tau_{13} + \omega_{n1} + \varepsilon_{n1}, \quad \omega_{n1} \sim N(0, \zeta_{11});\ \varepsilon_{n1} \sim GEV(0,1,0) \tag{21}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim GEV(0,1,0) \tag{22}$$

For each of the 400 datasets, we estimate an ICLV model with the specification given by equations (15)-(20) and a mixed logit model with the specification given by equations (21) and (22). All models are estimated using maximum simulated likelihood estimation with the software package Python Biogeme (Bierlaire, 2003). Figure 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 datasets with 100, 200, 500 and 1000 observations. Even though the data was generated according to the ICLV model specification, the mixed logit model can be seen to fit the data just as well. We fail to reject the hypothesis that the log-likelihood at convergence for the reduced form ML model is different than that for the ICLV model at a statistical significance of 10% for each of the four sets of

Figure 4: A plot of the cumulative distribution function over 100 datasets of the difference in log-likelihood at convergence between the ICLV model given by equations (15)-(20) and the sub-optimal mixed logit (ML) model given by equations (23) and (24)



100 datasets corresponding to 100, 200, 500 and 1000 observations. Moreover, the log-likelihood at convergence for the reduced form ML model is within one point of the ICLV model for 98, 96, 92 and 90 of the 100 datasets corresponding to 100, 200, 500 and 1000 observations, respectively. It isn't identical because of simulation noise. Maximum simulated likelihood is consistent if the number of draws T rises faster than the square root of the sample size $\sqrt{N}$. Since the number of draws is held constant across the datasets, simulation noise can be seen to increase with sample size. Therefore, we can conclude that, despite their greater complexity, ICLV models do not result in an improvement in fit with respect to the choice data over simpler choice models without latent variables.

3.3 Monte Carlo Experiment II

Despite the results presented in Sections 3.1 and 3.2, it isn't hard to find studies in the literature that have compared ICLV models to choice models without latent variables and concluded that the former provide a better fit to the choice data. The erroneous 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:

$$u_{n1} = \tau_{12} x_{n2} + \tau_{13} + \omega_{n1} + \varepsilon_{n1}, \quad \omega_{n1} \sim N(0, \zeta_{11});\ \varepsilon_{n1} \sim GEV(0,1,0) \qquad (23)$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim GEV(0,1,0) \qquad (24)$$

Figure 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 equations (15)-(20) and the mixed logit model given by equations (23) and (24) for the 100 datasets generated in Section 3.1 corresponding to 100 observations each. As is apparent, the mixed logit model fits much worse than the ICLV model. 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.

Table 1: Estimation results for the ICLV model given by equations (15)-(20), the reduced form mixed logit model given by equations (21) and (22), and the sub-optimal mixed logit model given by equations (23) and (24)

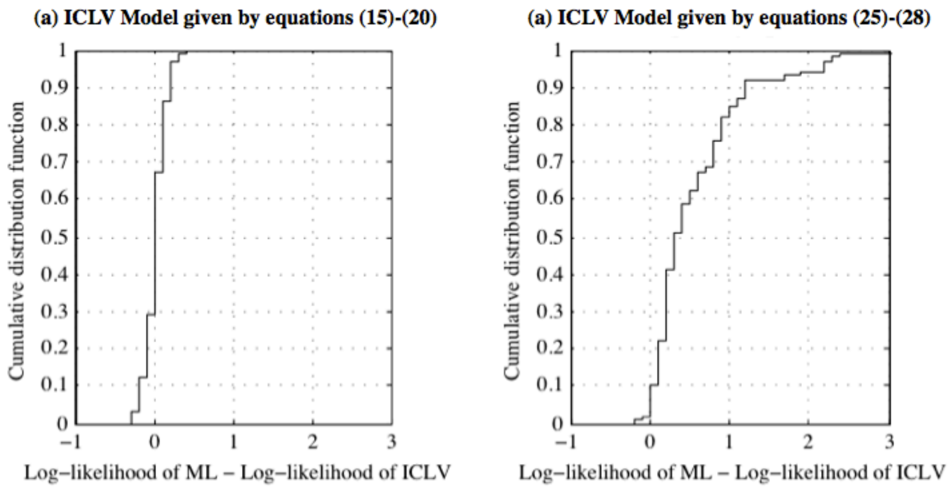
Parameter	**True value**	**ICLV**		**Reduced form mixed logit**		**Sub-optimal mixed logit**	
		est.	**p-val.**	**est.**	**p-val.**	**est.**	**p-val.**
Utility specification of recycling							
Constant	-3.00	-3.20	0.02	-0.14	0.80	-0.46	0.21
Attitude towards environment	0.90	0.93	0.01	-	-	-	-
Susceptibility 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**

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. Therefore, it has been argued, these variables were retained in the ICLV model but omitted from the choice model without latent variables. We 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 entirely 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 $\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21} \rightarrow 0$).

To show that this can happen, we synthesize a separate dataset with 100 observations using the ICLV model specification given by equations (15)-(20) 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 3.2). We use this dataset to recover parameter estimates for the original ICLV model specification, the reduced form mixed logit model given by equations (21) and (22), and the mixed logit model given by equations (23) and (24). Note that each of the three parameters associated with age in the ICLV model are 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

Figure 5: A plot of the cumulative distribution function of the difference in log-likelihood at convergence for datasets with 100 observations each between **(a)** the ICLV model given by equations (15)-(20) and the reduced form mixed logit (ML) model given by equations (21) and (22); and **(b)** the ICLV model given by equations (25)-(28) and the reduced form mixed logit (ML) model given by equations (21) and (22)



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.

3.4 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. 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:

$$u_{n1} = \beta_{13} + \gamma_{11} x^*_{n1} + \varepsilon_{n1}, \quad \varepsilon_{n1} \sim GEV(0,1,0) \tag{25}$$

$$u_{n2} = \varepsilon_{n2}, \quad \varepsilon_{n2} \sim GEV(0,1,0) \tag{26}$$

$$x^*_{n1} = \alpha_{11} x_{n1} + \alpha_{12} x_{n2} + \alpha_{13} + \nu_{n1}, \quad \nu_{n1} \sim N(0, \phi_{11}) \tag{27}$$

$$i_{n1} = x^*_{n1} + \eta_{n1}, \quad \eta_{n1} \sim N(0,1) \tag{28}$$

The reader should check that substituting equation (27) in equations (25) and (26) yields the same reduced form mixed logit model as that given by equations (21) and (22), even though the true underlying ICLV model specification is different in the two cases.

The degrees of freedom available to the choice sub-model of the ICLV model given by equations (25)-(28) are two (β_{13} and γ_{11}) whereas the degrees of freedom available to 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. Figure 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 equations (25)-(28) and the mixed logit model given by equations (21) and (22) for the 100 datasets generated in Section

3.2 corresponding to 100 observations each. 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 equations (15)-(20) and the mixed logit model given by equations (21) and (22) for the same 100 datasets. The plot on the right shows that the ICLV model given by equations (25)-(28) fits the data 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. The difference in fit 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 equation (25) or (26), thereby increasing the degrees of freedom of 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, and can usually be resolved by identifying the appropriate variables and including them in the utility specification. 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 variable $\mathbf{y_n}$ and the vector of indicators $\mathbf{i_n}$, as opposed to just the choice variable $\mathbf{y_n}$ in the case of the reduced form mixed logit model. This can be construed as an alternative explanation to the one provided here. However, if the choice sub-model of the ICLV model has at least as many degrees of freedom as the reduced form mixed logit model, then the ICLV model should fit the data as well as the reduced form mixed logit model.

3.5 Summary

In general, as long as the latent variables are represented as parametric random variables, the analyst can substitute the structural equation of the latent variables 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. If the primary objective is to predict outcomes to the choice indicators with greater accuracy, the analyst might be better served by employing nonparametric representations for the latent variables, such as the discrete formulation used by LCCMs (see, for example, Kamakura and Russell, 1989). In some cases though, even when the latent variables are represented as parametric random variables, it could be argued that the structure imposed by an ICLV model may lead the analyst to a reduced form choice model specification that may never have been evaluated had the analyst not estimated the original choice model with latent variables. Consider, for example, the ICLV model of bicycle route choice behavior presented in Bhat et al. (2014). The discrete choice sub-model has a probit kernel, and the utility of a bicycle route is specified as a linear in parameters function of seventeen attributes of the route (travel time, terrain, car traffic, bike facilities, etc.), twelve of which are interacted with one of two latent variables (attitudes towards bicycling and concern for safety). Each of the latent variables themselves is specified as a linear in parameters function of a different set of five observable demographic variables (age, gender, etc.). The utility specification for the equivalent reduced form multinomial probit choice model without latent variables is linear in seventeen parameters identifying the main effects associated with each of the route attributes, sixty parameters identifying the joint effect of these attributes and different demographic characteristics, and twelve random parameter error components. It would 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 have chanced upon this particular reduced form model specification. However, such examples are rare, and most ICLV models that have been reported in the literature have employed relatively parsimonious representations where the reduced form choice model is usually equally parsimonious. In these cases, it's hard to make the argument that the same reduced form specification could not have been arrived at in the absence of latent variables, and 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.

4. Identification 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 as proxies for these latent behavioral constructs, but without the ability to say exactly what these variables might be proxies for. For example, a number of studies on travel mode choice behavior have frequently used sociodemographic variables related to life cycle, sex, occupation and age as independent explanatory variables (for an early review of the literature on travel mode choice behavior, the reader is referred to Barff et al., 1982). But what are these variables proxies for? Are women less likely to use public transport than men because of a greater concern for safety or a greater need for flexibility? Do younger individuals bicycle more because they have fewer responsibilities and can therefore afford to, or they care more about the environment than older generations? In cases where tastes vary with unobservable variables, or purely randomly, heterogeneity in the decision-making process has typically been captured though additional interactions between observable variables and the stochastic component, resulting in such popular model forms as the mixed logit or the multinomial probit. For example, studies on travel mode choice behavior have repeatedly demonstrated that the value of travel time may vary considerably within a population, even after accounting for systematic differences in the characteristics of the individuals that constitute the population. But again, the question could be asked, why do these tastes vary? Can differences in sensitivities to travel times and costs be ascribed to differences in attitudes towards the transportation system? Or are they reflective of differences in perceptions of the transportation system? Or some other entirely separate reason or combination of reasons? By incorporating additional data through the use of measurement indicators, the ICLV model can help identify additional parameters associated with the latent explanatory variables, and lend structure and meaning to the reasons underlying differences in behavior.

To see that this is the case, we substitute equation (2) in equation (1) to get the following reduced form for the utility specification:

$$\mathbf{u_n} = (\mathbf{B} + \mathbf{\Gamma A})\mathbf{x_n} + \boldsymbol{\varepsilon}_\mathbf{n} \quad (29)$$

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 equation (3), these effects can be explicitly linked to specific behavioral constructs.

Take, for example, the hypothetical model of bicycle ownership first presented in Section 3.1, given by equations (15)-(20). Equations (17) and (18) may be substituted in equations (15) and (16) to get the following utility specification for the reduced form model:

$$\begin{aligned} u_{n1} = {} & (\beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21})x_{n1} + (\gamma_{11}\alpha_{12})x_{n2} + (\beta_{13} + \gamma_{11}\alpha_{13} + \gamma_{12}\alpha_{23}) \\ & + (\gamma_{11}\nu_{n1} + \gamma_{12}\nu_{n2} + \varepsilon_{n1}) \end{aligned} \quad (30)$$

$$u_{n2} = \varepsilon_{n2} \quad (31)$$

As we did in Section 3.2, define the parameters τ_{11}, τ_{12} and τ_{13} as follows:

$$\tau_{11} = \beta_{11} + \gamma_{11}\alpha_{11} + \gamma_{12}\alpha_{21} \quad (32)$$

$$\tau_{12} = \gamma_{11}\alpha_{12} \quad (33)$$

$$\tau_{13} = \beta_{13} + \gamma_{11}\alpha_{13} + \gamma_{12}\alpha_{23} \quad (34)$$

, where τ_{11} and τ_{12} capture the influence of age and gender on the utility of bicycle ownership, respectively, and τ_{13} is the alternative-specific constant. Equation (32) 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, equation (33) states that the influence of gender on the utility of bicycle ownership may be attributed entirely to its influence on attitudes towards the environment ($\gamma_{11}\alpha_{12}$). And finally, equation (34) 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's important to recognize that for the ICLV model, the two latent variables are 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.

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.

5. Consistency of Parameter Estimates

An estimator for a parameter is said to be consistent if with increasing sample size it converges to the true value of the parameter being estimated. In practice, the model parameters for the ICLV model are estimated from maximizing the following likelihood function over all decision-makers:

$$\begin{aligned} L(\mathbf{B},\boldsymbol{\Gamma},\mathbf{D},\boldsymbol{\Psi},\mathbf{A},\boldsymbol{\Phi}|\mathbf{y_n},\mathbf{i_n};\mathbf{x_n}) &= \int_{\mathbf{x}^*} f_y(\mathbf{y_n}|\mathbf{x_n},\mathbf{x_n^*};\mathbf{B},\boldsymbol{\Gamma}) f_i(\mathbf{i_n}|\mathbf{x_n},\mathbf{x_n^*};\mathbf{D},\boldsymbol{\Psi}) f_{x^*}(\mathbf{x_n^*}|\mathbf{x_n};\mathbf{A},\boldsymbol{\Phi}) d\mathbf{x_n^*} \\ &\approx \frac{1}{T}\sum_{t=1}^{T} f_y(\mathbf{y_n}|\mathbf{x_n},\mathbf{x_{nt}^*};\mathbf{B},\boldsymbol{\Gamma}) f_i(\mathbf{i_n}|\mathbf{x_n},\mathbf{x_{nt}^*};\mathbf{D},\boldsymbol{\Psi}) \end{aligned} \tag{35}$$

Similarly, the model parameters for the reduced form choice model without latent variables are estimated from maximizing the following simulated likelihood function over all decision-makers:

$$\begin{aligned} L(\mathbf{T},\mathbf{Z}|\mathbf{y_n};\mathbf{x_n}) &= \int_{\boldsymbol{\omega}} f_y(\mathbf{y_n}|\mathbf{x_n},\boldsymbol{\omega}_\mathbf{n};\mathbf{T}) f_\omega(\boldsymbol{\omega}_\mathbf{n}|\mathbf{Z}) d\boldsymbol{\omega}_\mathbf{n} \\ &\approx \frac{1}{T}\sum_{t=1}^{T} f_y(\mathbf{y_n}|\mathbf{x_n},\boldsymbol{\omega}_\mathbf{nt};\mathbf{T}) \end{aligned} \tag{36}$$

Since both sets of parameter estimates are obtained from maximizing functions that do not have a closed form solution that are usually simulated, the asymptotic properties of simulation-based estimators apply to both. As Train (2009) shows, for maximum simulated likelihood estimators, if the number of draws T rises faster than the square root of the sample size $\sqrt{N}$, then maximum simulated likelihood is consistent. Therefore, parameter estimates for both the ICLV model and the reduced form choice model should be consistent.

Table 2: Sample mean of parameter estimates

Parameter	True value	Model	Number of Observations			
			100	200	500	1000
Age (τ_{11})	0.38	*ICLV*	2.19	0.40	0.38	0.38
		Mixed Logit	0.37	0.38	0.38	0.38
Gender (τ_{12})	0.90	*ICLV*	6.56	1.15	0.89	1.04
		Mixed Logit	0.89	0.92	0.91	0.89
Constant (τ_{13})	-0.15	*ICLV*	-1.81	-0.29	-0.12	-0.17
		Mixed Logit	-0.12	-0.17	-0.13	-0.15
Bias (b)	**-**	***ICLV***	**10.01**	**0.86**	**0.42**	**0.34**
		Mixed Logit	**0.96**	**0.75**	**0.46**	**0.31**

As before, we validate the result using synthetic data. For the one hundred datasets corresponding to 100, 200, 500 and 1000 observations, generated in Section 3.2 using the ICLV model specification given by equations (15)-(20), Table 2 enumerates the sample mean of the estimates for the three parameters τ_{11}, τ_{12} and τ_{13} as recovered by the ICLV model specification and the corresponding reduced form mixed logit model given by equations (21) and (22). 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 equations (32)-(34) and the estimates for $\mathbf{A}, \mathbf{B}$ and $\mathbf{\Gamma}$. Since the reduced form mixed logit model cannot estimate $\mathbf{A}, \mathbf{B}$ and $\mathbf{\Gamma}$, the only way to compare estimates from the two models is to limit our attention to estimates for $\mathbf{T}$.

Let $\hat{\tau}_{jk,pqr}$ be the estimate for τ_{jk} recovered by model p from dataset q corresponding to number of observations r. Then the sample mean $\bar{\tau}_{jk,pr}$ for model p and number of observations r is given by:

$$\bar{\tau}_{jk,pr} = \frac{1}{Q}\sum_{q=1}^{Q} \hat{\tau}_{jk,pqr} \tag{37}$$

, where $Q = 100$ is the number of datasets for any number of observations r. The sample mean of the parameter estimates is a measure of consistency: estimates are consistent if the sample mean of the estimates converge to their true values with increasing number of observations. For estimates recovered using any combination of dataset and model form, we use the $\mathbf{L_1}$ distance, also known as the taxicab distance, between the estimates and the true values to quantify bias. Denoting bias by b, this translates into the following mathematical expression:

$$b_{pqr} = \sum_{j=1}^{J}\sum_{k=1}^{K}\left|\hat{\tau}_{jk,pqr} - \tau_{jk}\right| \tag{38}$$

Table 2 also lists the sample mean of the bias, calculated analogously from equation (37). Though the estimates from both model forms can be seen to converge to their true values, as shown by the decline in bias, estimates from the reduced form mixed logit model converge faster because it has fewer degrees of freedom. 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 for seven of the one hundred datasets corresponding to 100 observations, indicating that 100 observations are likely not enough to support estimation of the ICLV model specification. In contrast, the reduced form mixed logit model had no trouble recovering the true parameter values and the bias is comparable to that for more observations. At 200, 500 and 1000 observations, the difference in bias 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 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.

6. Efficiency of Parameter Estimates

It is widely believed that the additional information available to ICLV models through the incorporation of measurement indicators leads to an improvement in efficiency. 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 efficiency of parameter estimates increases with sample size, it increases 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 we are aware of that offers some evidence of an increase in efficiency, but their findings are not conclusive.

Since a reduced form model without latent variables cannot estimate $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Gamma}$, as was the case in Section 5, we limit our attention to estimates for $\mathbf{T}$. We show specifically that ICLV models may produce more efficient estimates for $\mathbf{T}$ 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 the observable data and provides a mechanism for identifying the unknown model parameters. Using notation not specific to any particular model form, the sample covariance matrix may be expressed as a function of the unknown model parameters, denoted $\boldsymbol{\theta}$, as follows:

$$\sigma_{ij} = f(\boldsymbol{\theta}) \tag{39}$$

, where σ_{ij} is the $(i, j)^{th}$ element of the sample covariance matrix $\boldsymbol{\Sigma}$. 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 (39), 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 $\boldsymbol{\Sigma}$ 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 more efficient estimates for some of these parameters. 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, an assumption that holds true for, among other distributions, the multivariate normal and logistic distributions.

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 observable, and not the absolute levels themselves, we introduce the $\left((\mathrm{J}-1)\times 1\right)$ vector of the differences in utilities, denoted $\Delta\mathbf{u_n}$, where Δ is the linear operator that transforms the J utilities into $(\mathrm{J}-1)$ utility differences taken with respect to the $\mathrm{J^{th}}$ alternative. Δ is a $\left((\mathrm{J}-1)\times\mathrm{J}\right)$ matrix that consists of a $\left((\mathrm{J}-1)\times(\mathrm{J}-1)\right)$ identity matrix with a column vector of -1's appended as the $\mathrm{J^{th}}$ 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 potentially more efficient parameter estimates.

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:

$$\begin{aligned}\mathrm{var}(\Delta\mathbf{u_n}) &= \Delta(\mathbf{B}+\mathbf{\Gamma A})\mathrm{E}[\mathbf{x_n x_n^T}](\mathbf{B}+\mathbf{\Gamma A})^{\mathrm{T}}\Delta^{\mathrm{T}} + \Delta\mathbf{\Gamma\Phi\Gamma}^{\mathrm{T}}\Delta^{\mathrm{T}} + \mathrm{k}\Delta\Delta^{\mathrm{T}} \\ &= \Delta\mathbf{T}\mathrm{E}[\mathbf{x_n x_n^T}]\mathbf{T}^{\mathrm{T}}\Delta^{\mathrm{T}} + \Delta\mathbf{Z}\Delta^{\mathrm{T}} + \mathrm{k}\Delta\Delta^{\mathrm{T}}\end{aligned} \quad (40)$$

$$\mathrm{cov}(\Delta\mathbf{u_n},\mathbf{x_n}) = \Delta(\mathbf{B}+\mathbf{\Gamma A})\mathrm{E}[\mathbf{x_n x_n^T}] = \Delta\mathbf{T}\mathrm{E}[\mathbf{x_n x_n^T}] \quad (41)$$

, where k is a constant, equal to the variance of a standard Gumbel distributed random variable. For a choice model without latent variables, equations (40) and (41) 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 $\mathbf{i_n}$:

$$\mathrm{var}(\mathbf{i_n}) = \mathbf{DA}\mathrm{E}[\mathbf{x_n x_n^T}]\mathbf{A}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}} + \mathbf{D\Phi D}^{\mathrm{T}} + \mathbf{\Psi} \quad (42)$$

$$\mathrm{cov}(\mathbf{i_n},\mathbf{x_n}) = \mathbf{DA}\mathrm{E}[\mathbf{x_n x_n^T}] \quad (43)$$

$$\mathrm{cov}(\Delta\mathbf{u_n},\mathbf{i_n}) = \Delta(\mathbf{B}+\mathbf{\Gamma A})\mathrm{E}[\mathbf{x_n x_n^T}]\mathbf{A}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}} + \Delta\mathbf{\Gamma\Phi D}^{\mathrm{T}} \quad (44)$$

The reader's attention is directed specifically towards equation (44), 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 equations (40)-(44), in some cases equation (44) 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 without latent variables will be equally efficient. But if the latter is true, estimates from the ICLV model may be more efficient.

Table 3: Sample standard error of parameter estimates for the ICLV model given by equations (15)-(20) and the reduced form mixed logit model given by equations (21)-(22)

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

Consider, for the sake of illustration, the two ICLV model specifications given by equations (15)-(20) and equations (25)-(28). For the latter model specification, the reader should verify that $\mathbf{A}$ can be completely identified from equation (43), $\mathbf{B}$ and $\mathbf{\Gamma}$ can be completely identified from equation (41), $\mathbf{\Phi}$ can be completely identified by equation (42), $\mathbf{D}$ and $\mathbf{\Psi}$ are constant matrices that don't need to be estimated, equation (40) helps to set the scale for the utility differences, and equation (44) 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 equation (43), $\mathbf{\Phi}$ can be completely identified by equation (42), $\mathbf{D}$ and $\mathbf{\Psi}$ are constant matrices that don't need to be estimated, and equation (40) helps to set the scale for the utility differences. However, equation (41) alone is no longer sufficient to identify $\mathbf{B}$ and $\mathbf{\Gamma}$, and equation (44) is needed as well to ensure that the model specification can be identified. As a result, estimates from the ICLV model given by equations (15)-(20) and the corresponding reduced form choice model without latent variables given by equations (21)-(22) should be equally efficient. But estimates from the ICLV model given by equations (25)-(28) should be more efficient than the corresponding reduced form choice model without latent variables given by equations (21)-(22).

As before, we validate the result using Monte Carlo experiments. For the one hundred datasets corresponding to 100, 200, 500 and 1000 observations, generated in Section 3.2 using the ICLV model specification given by equations (15)-(20), Table 3 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 equations (32)-(34) and the estimates for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Gamma}$, and directly by the corresponding reduced form mixed logit model given by equations (21) and (22). Given that $\hat{\tau}_{jk,pqr}$ is the estimate for τ_{jk} recovered by model p from dataset q corresponding to observations r, the sample standard error $SE(\hat{\tau}_{jk,pr})$ for model p and observations r may be calculated as:

$$SE(\hat{\tau}_{jk,pr}) = \sqrt{\frac{1}{Q-1}\sum_{q=1}^{Q}\left(\hat{\tau}_{jk,pqr} - \bar{\tau}_{jk,pr}\right)^2} \tag{45}$$

The sample standard error of the parameter estimates is a measure of efficiency: estimates with lower standard errors are more efficient than estimates with higher standard errors. 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

Table 4: Sample standard error of parameter estimates for the ICLV model given by equations (25)-(28) and the reduced form mixed logit model given by equations (21)-(22)

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

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 equally efficient.

Contrast these results with those for the analogous datasets generated using the ICLV model specification given by equations (25)-(28), shown in Table 4. 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 is apparent from the table, 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 estimates from the ICLV model will be at least as efficient as estimates from the reduced form choice model without latent variables, and in some cases, depending upon the model structure, they might actually be more efficient.

7. 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; and 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 remained unclear. This study was born from the discovery that any ICLV model can be reduced to a choice model without latent variables that fits the 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 statistical properties of ICLV models vis-à-vis more traditional choice models without latent variables. In terms of both goodness of fit and the consistency of parameter estimates, we found that choice models without latent variables are no worse than ICLV models. However, the ICLV model allows for the identification of structural relationships between observable variables that could not be identified by a choice model without latent variables, and parameter estimates from the ICLV model might potentially be more efficient than equivalent estimates from a comparable choice model without latent variables.

The focus of this paper was on the statistical benefits of ICLV models. Much work still remains to be done to establish the policy benefits of the framework (cf. Chorus and Kroesen, 2014). Unfortunately, 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. If an improvement in efficiency, and corresponding improvements in the precision of policy outputs such as willingness to pay measures and demand forecasts, are the only policy benefits over choice models without latent variables, then estimating an ICLV model hardly seems worth the effort. Studies need to show either that 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, or that 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. If used appropriately, the ICLV model could prove to be a very valuable 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 thoughtful with regards to the design of the model framework and more creative in illustrating its value to practitioners and policy-makers.

Acknowledgements

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, Chandra Bhat, Michel Bierlaire, David Brownstone, David Bunch, Ricardo Daziano, Anthony Marley and Joffre Swait, among others. Of course, any mistakes and errors are the fault of the authors alone.

References

Abou-Zeid, M., (2009), "**Measuring and modeling activity and travel well-being**," *PhD dissertation*, Massachusetts Institute of Technology.

Abou-Zeid, M., and 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,* Vol. 12, pp. 65-78.

Azjen, I. (1991), "**The theory of planned behavior**," *Organizational Behavior and Human Decision*, Vol. 50, No. 2, pp. 179-211.

Bamberg, S., Schmidt, P. (2001), "**Theory-driven evaluation of an intervention to reduce the private car-use**," *Journal of Applied Social Psychology*, Vol. 31, No. 6, pp. 1300–1329.

Barff, R., MacKay, D., and Olshavsky, R. W. (1982), "**A selective review of travel-mode choice models**," *Journal of Consumer Research*, pp. 370-380.

Ben-Akiva, M., and Boccara, B. (1995), "**Discrete choice models with latent choice sets**", *International Journal of Research in Marketing*, Vol. 12, No. 1, pp. 9-24.

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

Bhat, C. R., & Dubey, S. K. (2014), "**A new estimation approach to integrate latent psychological constructs in choice modeling**," *Transportation Research Part B: Methodological*, *67*, pp. 68-85.

Bhat, C. R., Dubey, S. K., and Nagel, K. (2014), "**Introducing Non-Normality of Latent Psychological Constructs in Choice Modeling with an Application to Bicyclist Route Choice**."

Bierlaire, M. (2003), "**BIOGEME: A free package for the estimation of discrete choice models**," *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Chorus, C. G., and Kroesen, M. (2014), "**On the (im-)possibility of deriving transport policy implications from hybrid choice models**," *Transport Policy*, *36*, pp. 217-222.

Daziano, R. A., & Bolduc, D. (2013), "**Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model**," *Transportmetrica A: Transport Science*, *9*(1), pp. 74-106.

Daziano, R. A. (2015), "**Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model**," *Transportation Research Part B: Methodological*, *76*, pp. 1-26.

Gärling, T., Fujii, S., Gärling, A., and Jakobsson, C. (2003), "**Moderating effects of social value orientation on determinants of pro-environmental behavior intention**," *Journal of Environmental Psychology*, Vol. 23, pp. 1-9.

Homer, P. M., and Kahle, L. R. (1988), "**A structural equation test of the value-attitude-behavior hierarchy**," *Journal of Personality and Social Psychology*, Vol. 54, No. 4, pp. 638-646.

Kamakura, W. A., and Russell, G. J. (1989), "**A probabilistic choice model for market segmentation and elasticity structure**," *Journal of Marketing Research*, Vol. 26, No. 3, pp. 379–390.

Kirk, J. L. (2010), "**Sustainable environments and pro-environmental behavior**," *Master's thesis*, University of Nebraska, Lincoln.

McFadden, D. L. (1986), "**The choice theory approach to marketing research**," *Marketing Science*, Vol. 5, No. 4, pp. 275–297.

Muthén, L. K., and Muthén, B. O. (2011), "**Mplus User's Guide**," Sixth Edition, Los Angeles, CA: Muthén & Muthén.

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.

Prato, C. G., Bekhor, S., and Pronello, C. (2012), "**Latent variables and route choice behavior**," *Transportation*, Vol. 39, No. 2, pp. 299-319.

Raveau, S., Daziano, R. A., Yáñez, M. F., Bolduc, D., and Ortúzar, J. de D. (2010), "**Sequential and simultaneous estimation of hybrid discrete choice models: Some new findings**," *Transportation Research Record: Journal of the Transportation Research Board*, No. 2156, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 131–139.

Steinberg, L., and Monahan, K. C. (2007), "**Age differences in resistance to peer influence**," *Developmental psychology*, *43*(6), pp. 1531-1543.

Train, K. E. (2009), "**Discrete choice models with simulation**," Cambridge University Press, Cambridge.

Triandis, H. C. (1977), "**Interpersonal Behaviour**," Monterey, CA: Brooks/Cole.

Williams, H. C. W. L., and Ortúzar, J. de D. (1982), "**Behavioural theories of dispersion and the misspecification of travel demand models**," *Transportation Research Part B: Methodological,* Vol. 16, No. 3, pp. 167–219.