Technical Report 102

A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling

Chandra R. Bhat
Subodh K. Dubey
Center for Transportation Research

September 2014

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

1. Report No. D-STOP/2016/102	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling		5. Report Date September 2014
		6. Performing Organization Code
7. Author(s) Chandra R. Bhat, Subodh K. Dubey		8. Performing Organization Report No. Report 102
9. Performing Organization Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701		13. Type of Report and Period Covered
		14. Sponsoring Agency Code

15. Supplementary Notes
Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

16. Abstract

In the current paper, we propose a new multinomial probit-based model formulation for integrated choice and latent variable (ICLV) models, which, as we show in the paper, has several important advantages relative to the traditional logit kernel-based ICLV formulation. Combining this MNP-based ICLV model formulation with Bhat's maximum approximate composite marginal likelihood (MACML) inference approach resolves the specification and estimation challenges that are typically encountered with the traditional ICLV formulation estimated using simulation approaches. Our proposed approach can provide very substantial computational time advantages, because the dimensionality of integration in the log-likelihood function is independent of the number of latent variables. Further, our proposed approach easily accommodates ordinal indicators for the latent variables, as well as combinations of ordinal and continuous response indicators. The approach can be extended in a relatively straightforward fashion to also include nominal indicator variables. A simulation exercise in the virtual context of travel mode choice shows that the MACML inference approach is very effective at recovering parameters. The time for convergence is of the order of 30 minutes to 80 minutes for sample sizes ranging from 500 observations to 2000 observations, in contrast to much longer times for convergence experienced in typical ICLV model estimations.

17. Key Words Multinomial probit, ICLV models, MACML estimation approach		18. Distribution Statement No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161	
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 44	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

.

Table of Contents

List of Illustrations

Chapter 1. Introduction

Economic choice modeling has been the mainstay of human behavioral modeling in many fields, including geography, urban planning, marketing, sociology, and transportation. The typical paradigm is based on a latent construct representing the value or utility that an individual decision-maker assigns to each of many available and mutually exclusive alternatives. The choice of an alternative is assumed to be the result of that alternative's utility being higher than its competitors in the perception space of the decision-maker. This utility itself is typically mapped to observed characteristics of the decision-maker (such as the socio-demographics of an individual in work mode choice modeling) and observed characteristics of the alternatives (such as travel time and travel costs by alternative modes in work mode choice modeling). To acknowledge that there may be unobserved characteristics of decision-makers (such as attitudes and lifestyle preferences) that are likely to impact choice, one of three approaches has been used in the literature. The first approach allows the intrinsic preference for alternatives as well as the sensitivities to alternative attributes to vary across decision-makers, using discrete (non-parametric) or continuous (parametric) random distributions to capture sensitivity variations (or taste heterogeneity). Early examples include the studies by Revelt and Train (1996), Bhat (1997), and Bhat (1998), and there have now been many applications of this approach, using latent multinomial logit and mixed logit formulations. A problem with this approach, though, is that some of the attitudes may be correlated with explanatory variables. Thus, an individual who is environmentally-conscious (say an unobserved variable) may locate herself or himself near transit stations, generating a correlation between the unobserved variable and a transit travel time variable used as an explanatory variable. Such correlations lead to inconsistent estimation. Besides, this method treats unobserved psychological preliminaries of choice (i.e., attitudes and preferences) as being contained in a "black box" to be integrated out. The second approach uses indicators of attitudes directly as explanatory variables in choice models. Such a technique has been used by Koppelman and Hauser (1978), Bhat et al. (1993), and many other subsequent studies. But this approach assumes that the indicators of attitudes directly represent the underlying attitudes that actually impact choice, which may not be the case. Rather, the indicators may be proxies of attitudes that are captured with some measurement error. Ignoring measurement error will, in general, lead to inconsistent estimation (see Ashok et al., 2002). Further, the attitude indicators may be correlated with other unobserved individual-specific factors that influence choice, rendering the estimation potentially inconsistent. In addition, the lack of a structural model to relate the attitudes to observed explanatory variables implies that the estimated model cannot be used in forecasting mode. The third approach is to undertake a factor analysis of the indicators to develop latent variables, typically using a multiple indicator multiple cause (MIMIC) model in which the latent variables are explained by a combination of observable indicators and observed (individual and alternative-specific) covariates. Essentially, factor analysis has the purpose of reducing the high number of correlated attitudinal indicators to a more manageable and relatively orthogonal set of latent variables, which are subsequently used as "error-free" explanatory variables (along with other covariates) in the choice model of interest. But such an approach, like the second

approach discussed earlier, is, in general, econometrically inconsistent. This is because latent variables specific to individual alternatives (such as comfort level of traveling on a bus in a mode choice model), or latent variables interacted with variables that vary across alternatives (such as perceptions of security that may interact with the travel time on the mode), lead to heteroscedasticity across the errors of the alternatives in the choice model, and latent variables applicable to a subset of alternatives (such as the sociable nature of the individual that may affect the utility ascribed to all transit modes) generate correlation patterns across the errors of the alternatives. Further, if the latent variables are interacted with individual-specific observed variables (such as the comfort level of traveling on the bus affecting bus utility through its interaction with the travel time on the bus), the result is also heterogeneity across individuals in the entire covariance matrix of alternatives (this is an issue that does not seem to have been acknowledged in the previous literature). Such a complex covariance matrix structure across alternatives and across individuals necessitates the explicit consideration of stochasticity in the latent variables.

A rapidly growing field of study that integrates latent psychological constructs such as attitudes and preferences within traditional choice models takes the form of a hybrid model that is commonly referred to as the Integrated choice and latent variable (ICLV) model (see Ben-Akiva et al., 2002 and Bolduc et al., 2005). In this approach, the objective is to gain a deeper understanding into the decision process of individuals by combining traditionally used "hard" covariates with "soft" psychometric measures associated with individual attitudes and perceptions. In this way, there is recognition that latent individual-specific variables (attitudes and perceptions) may be just as important as observed covariates in shaping choice and that their inclusion is likely not only to shed more light on the actual decision process but also potentially enhance the predictive ability of the model (Temme et al., 2008, Bolduc and Alvarez-Daziano, 2010). A typical ICLV model includes a latent variable structural equation model that relates latent constructs of attitudes and perceptions to observed covariates. Further, the latent constructs (or variables) themselves are viewed as being manifested through the attitudinal and perception indicator variables in a latent measurement equation model, which recognizes the presence of measurement error in capturing the intrinsic latent constructs. In the event that one of more of the indicators are not observed on a continuous scale, but observed on an ordinal or nominal scale, the measurement equation also serves the role of mapping the continuous latent constructs to the ordinal or nominal scale of the observed attitudinal indicator variables. Finally, the "soft" latent variables and the "hard" observed variables are used together to explain choice in a random utility maximizing choice model set-up.

While the number of applications of ICLV models has been on the rise in recent years (see, for example, Johansson et al., 2006, Bolduc et al., 2005, Temme et al., 2008, Daly et al., 2012, and Alvarez-Daziano and Bolduc, 2013), the use of such models is severely hampered by (1) the restrictive specifications used in application, (2) the difficulties encountered in estimation, and (3) the amount of time it takes to estimate these models (typically of the order of a day for one specification run). Thus, earlier applications of the ICLV model typically use an independent and identically distributed Gumbel error term for the stochastic component of the utility of alternatives, imposing a priori the notion that, net of the latent attitudinal factors and observed covariates, there is no remaining

correlation across the utilities of alternatives.[1] Similarly, the error correlations in the latent variables are almost always ignored within the latent variable structural equation model, as has also been pointed out by Vij and Walker (2014). Such correlations in the latent variables may arise because of common underlying unobserved individual values that are precursors to attitude formation and that may impact multiple attitude variables at once (see Temme et al., 2008). Also, the estimation method of choice for ICLV models has been the maximum simulated likelihood approach, similar to those developed for the traditional mixed logit model with random coefficients or error components (see Bhat, 2001). However, while these simulation techniques work quite well for the traditional mixed logit, their use in ICLV models has been problematic because the integrand in ICLV models is itself a mixture of two probabilities (the probability of choice conditional on explanatory and latent variables, and the probability of the latent variable conditional on explanatory variables), which has to be integrated over the distribution of the latent variables conditional on explanatory variables. On the other hand, the integrand in the mixed logit model is simply the multinomial probability that has to be integrated over the distribution of the unobservables, which is far easier and less involved. As a result, it is quite routine to encounter convergence problems in ICLV models. For example, Alvarez-Daziano and Bolduc (2013) indicate that, unless the second derivatives of the logarithm of the likelihood function of ICLV models is analytically coded and provided, it is difficult to obtain convergence in ICLV models using simulation techniques. In this regard, they note that most software that allow for custom likelihood provision but use approximations to the Hessian during optimization do not guarantee convergence in ICLV models. This is particularly the case with many latent variables or constructs, since the number of latent variables has a direct bearing on the dimensionality of the integral that needs to be evaluated in the log-likelihood function of ICLV models. The consequence has been that most ICLV models in the literature have gravitated toward the use of a limited number of latent constructs, rather than exploring a fuller set of possible latent variables. Also, while Alvarez-Daziano and Bolduc (2013) present a Bayesian Markov Chain Monte Carlo (MCMC) simulation approach to estimating the ICLV model, this remains cumbersome, requires extensive simulation related to the Metropolis Hastings-within-Gibbs algorithm needed to generate an instance from the otherwise not-explicitly-characterizable posterior distribution of the ICLV model, and poses convergence assessment problems as the number of latent variables (equivalent to the number of dimensions of integration within the classic MSL) increases (see Franzese et al., 2010 for a discussion of this issue). The MCMC method also becomes very challenging when there are several ordinal indicators of the latent variables in an ICLV model, which is why one would assume Alvarez-Daziano and Bolduc (2013) chose to work with all continuous indicators of latent variables in their study. Finally, the Bayesian MCMC methods typically parallel the computation intensity of classical MSL methods and so do not offer any substantial order-of-time computational advantages.

In the current paper, we propose a different model formulation for the ICLV model, based on a multivariate probit (MNP) kernel that alleviates the specification and

---

[1] While a more general covariance structure can be incorporated through the use of normally-mixed error-components or random coefficients (or a combination), this adds more to the integration dimension and makes an already very difficult MSL estimation problem more difficult, which is certainly one important reason why earlier applications of the logit-kernel ICLV have not introduced such general covariance structures.

estimation challenges discussed above. To our knowledge, this is the first study to use a probit kernel within a general ICLV setting. In this context, it is quite remarkable that earlier and general ICLV formulations have not considered using an MNP kernel. We believe that this is because of a fixation with the mixed logit model, which may be easier to estimate than the MNP model in a traditional choice model when there are a number of alternatives and few random coefficients or error components to integrate over. But the situation changes with the ICLV model, and, in general, the MNP kernel is much more convenient to use. As we show in the rest of this paper, doing so has many advantages, especially when combined with our proposal to estimate the resulting model using Bhat's maximum approximate composite marginal likelihood (MACML) inference approach. In particular, in our approach, the dimensionality of integration in the composite marginal likelihood (CML) function that needs to be maximized to obtain a consistent estimator (under standard regularity conditions) for the ICLV model parameters is independent of the number of latent variables and the number of ordinal indicator variables, and is only of the order of the number of alternatives in the choice model. Given that the number of alternatives in most applications of ICLV models is small, our formulation has a distinct advantage over the traditional ICLV formulation. Further, the use of our analytic approximation in the MACML approach to evaluate the multivariate cumulative normal distribution (MVNCD) function in the CML function corresponding to our MNP-based ICLV model simplifies the estimation procedure even further. Importantly, regardless of the number of ordinal indicators, or the number of latent variables, or the number of random coefficients in the choice model, or the covariance structures assumed in the ICLV model set-up, the proposed MACML procedure for estimating ICLV models requires the maximization of a function that has no more than bivariate normal cumulative distribution functions to be evaluated. This is quite remarkable. Additionally, our procedure does away with the convergence problems associated with the MSLE estimation. This is because of the smoothness of our analytically approximated log-likelihood surface, which leads to well-behaved surfaces for the gradient and hessian functions. In turn, this allows the use of widely available optimization software in which the Hessian is approximated numerically. In addition, our approach very easily handles the case of ordinal indicators for the latent variables, as well as combinations of ordinal and continuous response indicators, rather than, as in almost all earlier studies, assuming these indicators to all be continuous (as in Alvarez-Daziano and Bolduc, 2013) or all be ordinal (Daly et al., 2012). In this regard, we develop a blue print, complete with appropriate matrix notation, for the formulation, estimation, and software coding of ICLV models with a combination of two different kinds of response indicators (the approach is extendible even to nominal indicators, though, for focus and presentation ease, we confine attention in this paper to ordinal and continuous variables). Finally, we provide some computational time statistics for estimating our MNP-based ICLV model using the MACML procedure, and show how our approach substantially reduces the time for estimating ICLV models. The order of magnitude reduction in computation time can then be used by analysts to explore a wide array of observed and latent variable specifications, rather than examining just a few specifications and settling quickly on one that may not be the best.

The remainder of this paper is structured as follows. In the next section, we formulate our MNP-based ICLV model. In Section 3, we discuss identification considerations and the

estimation procedure. In Section 4, we develop the experimental design to generate a simulated mode choice data set that is then used in Section 5 to examine the performance of the proposed estimation procedure in terms of recovering parameters and evaluating the finite-sample behavior of the proposed estimator. Our use of a simulated data set rather than real data allows us to assess the performance of our estimation approach for different sample sizes. Finally, Section 6 summarizes the key findings of the paper and identifies directions for further research.

Chapter 2. Model Formulation

There are three components to the model: (1) the latent variable structural equation model, (2) the latent variable measurement equation model, and (3) the choice model. These components are discussed in turn below. In the following presentation, for ease in notation, we will consider a cross-sectional model. However, extension to the case of a panel model with multiple choice instances from the same individual is quite straightforward. Also, we will use the index l for latent variables ($l=1,2,…L$), and the index i for alternatives ($i=1,2,…,I$). As appropriate and convenient, we will suppress the index q for individuals ($q=1,2,…,Q$) in parts of the presentation.

2.1 Latent Variable Structural Equation Model

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

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

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

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

2.2 Latent Variable Measurement Equation Model

For the latent variable measurement equation model, let there be H continuous variables $(y_1, y_2, ..., y_H)$ with an associated index h $(h = 1, 2, ..., H)$. Let $y_h = \delta_h + \boldsymbol{d}_h' \boldsymbol{z}^* + \xi_h$ in the usual linear regression fashion, where δ_h is a scalar constant, $\boldsymbol{d}_h$ is an $(L \times 1)$ vector of latent variable loadings on the hth continuous indicator variable, and ξ_h is a normally distributed measurement error term. Stack the H continuous variables into a $(H \times 1)$-vector $\boldsymbol{y}$, the H constants δ_h into a $(H \times 1)$ vector $\boldsymbol{\delta}$, and the H error terms into another

$(H \times 1)$ vector $\xi = (\xi_1, \xi_2, ..., \xi_H)$. Also, let Σ_y be the covariance matrix of ξ. And define the $(H \times L)$ matrix of latent variable loadings $d = (d_1, d_2, ..., d_H)'$. Then, one may write, in matrix form, the following measurement equation for the continuous indicator variables:

$$y = \delta + dz^* + \xi \qquad (3)$$

Similar to the continuous variables, let there also be G ordinal indicator variables, and let g be the index for the ordinal variables $(g = 1, 2, ..., G)$. Let the index for the ordinal outcome category for the gth ordinal variable be represented by j_g. For notational ease only, assume that the number of ordinal categories is the same across the ordinal indicator variables, so that $j_g \in \{1, 2, ..., J\}$. Let y_g^* be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the gth ordinal indicator variable, and let the individual under consideration choose the n_g^{th} ordinal outcome category for the gth ordinal indicator variable. Then, in the usual ordered response formulation, we may write: $y_g^* = \widetilde{\delta}_g + \widetilde{d}_g' z^* + \widetilde{\xi}_g$, $\psi_{g,n_g-1} < y_g^* < \psi_{g,n_g}$, where δ_g is a scalar constant, $\widetilde{d}_g$ is an $(L \times 1)$ vector of latent variable loadings on the underlying variable for the gth indicator variable, and ξ_g is a standard normally distributed measurement error term (the normalization on the error term is needed for identification, as in the usual ordered-response model; see McKelvey and Zavoina, 1975). Note also that, for each ordinal indicator variable, $\psi_{g,0} < \psi_{g,1} < \psi_{g,2} ... < \psi_{g,N_g-1} < \psi_{N_g}$; $\psi_{g,0} = -\infty, \psi_{g,1} = 0$, and $\psi_{g,J} = +\infty$. For later use, let $\psi_g = (\psi_{g,2}, \psi_{g,3}, ... \psi_{g,J-1})'$, and $\psi = (\psi_1', \psi_2', ... \psi_G')'$. Stack the G underlying continuous variables y_g^* into a $(G \times 1)$ vector y^* and the G constants $\widetilde{\delta}_g$ into a $(G \times 1)$ vector $\widetilde{\delta}$. Also, define the $(G \times L)$ matrix of latent variable loadings $\widetilde{d} = (\widetilde{d}_1, \widetilde{d}_2, ..., \widetilde{d}_G)'$, and let Σ_{y^*} be the correlation matrix of $\widetilde{\xi} = (\widetilde{\xi}_1, \widetilde{\xi}_2, ..., \widetilde{\xi}_G)$. Stack the lower thresholds $\psi_{g,n_g-1} (g = 1, 2, ..., G)$ into a $(G \times 1)$ vector ψ_{low} and the upper thresholds $\psi_{g,n_g} (g = 1, 2, ..., G)$ into another vector ψ_{up}. Then, in matrix form, the measurement equation for the ordinal indicators may be written as:

$$y^* = \widetilde{\delta} + \widetilde{d}z^* + \widetilde{\xi}, \quad \psi_{low} < y^* < \psi_{up}. \qquad (4)$$

Define $\breve{y} = \left(y', \left[y^* \right]' \right)'$, $\breve{\delta} = (\delta', \widetilde{\delta}')'$, $\breve{d} = (d', \widetilde{d}')'$, and $\breve{\xi} = (\xi', \widetilde{\xi}')'$. Then, the continuous parts of Equations (3) and (4) may be combined into a single equation as:

$$\breve{y} = \breve{\delta} + \breve{d}z^* + \breve{\xi}, \text{ with } \quad E(\breve{y}) = \begin{bmatrix} \delta + dz^* \\ \widetilde{\delta} + \widetilde{d}z^* \end{bmatrix}, \text{ and } \text{Var}(\breve{\xi}) = \breve{\Sigma} = \begin{bmatrix} \Sigma_y & \Sigma_{yy^*} \\ \Sigma_{yy^*}' & \Sigma_{y^*} \end{bmatrix} \qquad (5)$$

2.3 Choice Model

Assume a typical random utility-maximizing model, and let i be the index for alternatives (i = 1, 2, 3, …, I). Note that some alternatives may not be available for some individuals, but the modification to allow this is quite trivial. So, for presentation convenience, we will assume that all alternatives are available to all individuals. We also will assume fixed rather than random coefficients on the exogenous variables, though the consideration of normally distributed coefficients does not increase the computational complexity in our set-up (because of the normal kernel error term in the utilities, and the conjugate additive nature of the normal distribution).[2] The utility for alternative i is then written as:

$$U_i = \boldsymbol{\beta}' \boldsymbol{x}_i + \boldsymbol{\gamma}_i' (\boldsymbol{\varphi}_i \boldsymbol{z}^*) + \varepsilon_i ,\tag{6}$$

where $\boldsymbol{x}_i$ is a (D×1)-column vector of exogenous attributes. $\boldsymbol{\beta}$ is a (D×1)-column vector of corresponding coefficients, $\boldsymbol{\varphi}_i$ is a $(N_i \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i, $\boldsymbol{\gamma}_i$ is a $(N_i \times 1)$-column vector of coefficients capturing the effects of latent variables and its interaction effects with other exogenous variables, and ε_i is a normal error term. The notation above is very general. Thus, if each of the latent variables impacts the utility of alternative i purely through a constant shift in the utility function, $\boldsymbol{\varphi}_i$ will be an identity matrix of size L , and each element of $\boldsymbol{\gamma}_i$ will capture the effect of a latent variable on the constant specific to alternative i. Alternatively, if the first latent variable is the only one relevant for the utility of alternative i, and it affects the utility of alternative i through both a constant shift as well as an exogenous variable, then N_i=2, and $\boldsymbol{\varphi}_i$ will be a $(2 \times L)$-matrix, with the first row having a '1' in the first column and '0' entries elsewhere, and the second row having the exogenous variable value in the first column and '0' entries elsewhere. A whole range of other latent variable specifications may also be considered based on appropriately configuring the matrix $\boldsymbol{\varphi}_i$.

To proceed further, let the variance-covariance matrix of the vertically stacked vector of errors $\boldsymbol{\varepsilon}$ $[= (\varepsilon_1, \varepsilon_2, ..., \varepsilon_I)']$ be $\boldsymbol{\Lambda}$. The choice model above may be written in a more compact form by defining the following vectors and matrices: $\boldsymbol{U} = (U_1, U_2, ..., U_I)'$ ($I \times 1$ vector), $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, ..., \boldsymbol{x_I})' (I \times D$ matrix),   and $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', ..., \boldsymbol{\varphi}_I')'$ $\left(\sum_{i=1}^{I} N_i \times L \right)$ matrix. Also, define the $\left(I \times \sum_{i=1}^{I} N_i \right)$ matrix $\boldsymbol{\gamma}$, which is initially filled with all zero values. Then, position the $(1 \times N_1)$ row vector $\boldsymbol{\gamma_1}'$ in the first row to occupy columns 1 to N_1 , position the $(1 \times N_2)$ row vector $\boldsymbol{\gamma_2}'$ in the second row to occupy

---

[2] One can also consider a skew-normal distribution for the coefficients that can accommodate asymmetric and non-normal (but unimodal) distributions. This will still enable the application of our proposed procedure, because the cumulative distribution of the skew-normal takes the form of the multivariate cumulative normal distribution (see Bhat and Sidharthan, 2012). Alternatively, one can accommodate non-normality in our set-up by considering a scale mixture of normals.

columns N_1+1 to $N_1 + N_2$, and so on until the $(1 \times N_I)$ row vector γ'_I is appropriately positioned. Then, in matrix form, Equation (6) may be written as:

$$U = x\beta + (\gamma\varphi) \, z^* + \varepsilon = x\beta + \lambda z^* + \varepsilon, \text{ with } \lambda = \gamma\varphi. \tag{7}$$

Consider now that the individual under consideration chooses alternative m. Under the utility maximization paradigm, $U_i - U_m$ must be less than zero for all $i \neq m$, since the individual chose alternative m. Let $u^*_{im} = U_i - U_m (i \neq m)$, and stack the latent utility differentials into a vector $u^* = \left[\left(u^*_{1m}, u^*_{2m}, ..., u^*_{Im} \right)'; i \neq m \right]$.

In the context of the choice model formulation above, several important identification issues need to be addressed (in addition to the usual identification consideration that one of the alternatives has to be used as the base for each nominal variable when introducing alternative-specific constants and variables that do not vary across the I alternatives). First, only the covariance matrix of the error differences is estimable. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\Lambda}$ of $\varsigma = (\varsigma_2, \varsigma_3, ..., \varsigma_I)$, where $\varsigma_i = \varepsilon_i - \varepsilon_1$ ($i \neq 1$), are estimable. However, the condition that $u^* < 0_{I-1}$ takes the difference against the alternative m that is chosen. Thus, during estimation, the covariance matrix $\tilde{\Lambda}$ (of the error differences taken with respect to alternative m is desired). Since m will vary across individuals, $\tilde{\Lambda}$ will also vary across households. But all the $\tilde{\Lambda}$ matrices must originate in the same covariance matrix Λ for the original error term vector ε. To achieve this consistency, Λ is constructed from $\breve{\Lambda}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. Second, an additional scale normalization needs to be imposed on $\breve{\Lambda}$. For this, we normalize the first element of $\breve{\Lambda}$ to the value of one. Third, in MNP models, identification is tenuous when only individual-specific covariates are used (see Keane, 1992 and Munkin and Trivedi, 2008). In particular, exclusion restrictions are needed in the form of at least one individual characteristic being excluded from each alternative's utility in addition to being excluded from a base alternative (but appearing in some other utilities). But these exclusion restrictions are not needed when there are alternative-specific variables.

Chapter 3. Model System Identification and Estimation

For convenience, Table 1 provides a list of all matrices and their dimensions. Also, as in all earlier studies of ICLV models, we assume that the error vectors $\boldsymbol{\eta}$, $\breve{\boldsymbol{\xi}}$, and $\boldsymbol{\varepsilon}$ are independent of each other. Let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vech}(\boldsymbol{\Gamma}), \breve{\boldsymbol{\delta}}, \text{Vech}(\breve{\boldsymbol{d}}), \boldsymbol{\psi}, \text{Vech}(\breve{\boldsymbol{\Sigma}}), \boldsymbol{\beta}, \text{Vech}(\boldsymbol{\gamma}), \text{Vech}(\breve{\boldsymbol{\Lambda}})]$, where $\text{Vech}(\boldsymbol{\alpha})$, $\text{Vech}(\breve{\boldsymbol{d}})$, and $\text{Vech}(\boldsymbol{\gamma})$ represent vectors of the elements of the $\boldsymbol{\alpha}$, $\breve{\boldsymbol{d}}$, and $\boldsymbol{\gamma}$, respectively, to be estimated, and $\text{Vech}(\boldsymbol{\Gamma})$ represents the vector of the non-zero upper triangle elements of $\boldsymbol{\Gamma}$ (and similarly for other covariance matrices). The data for estimation include, for each individual, (1) the $\boldsymbol{x}, \boldsymbol{\varphi}$ and $\boldsymbol{w}$ covariate matrices, (2) The $(H \times 1)$-vector of continuous indicator variables, (3) the $n_g{}^{th}$ outcome category for each of the g ordinal indicator variables, and (4) the observed choice outcome m (note that a particular empirical context may not have any continuous indicator variable, or may not have any ordinal indicator variable, but we will assume the presence of a combination of the two to reflect the general case).

To develop the reduced form equations, replace the right side of Equation (1) for z^* in Equations (5) and (7) to obtain the following system:

$$\breve{y} = \breve{\delta} + \breve{d}z^* + \breve{\xi} = \breve{\delta} + \breve{d}(\alpha w + \eta) + \breve{\xi} = \breve{\delta} + \breve{d}\alpha w + \breve{d}\eta + \breve{\xi} \tag{8}$$

$$\boldsymbol{U} = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda} z^* + \boldsymbol{\varepsilon} = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda}(\alpha w + \eta) + \boldsymbol{\varepsilon} = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\alpha w + \boldsymbol{\lambda}\eta + \boldsymbol{\varepsilon} \tag{9}$$

Now, consider the $[(H + G + I) \times 1)]$ vector $\boldsymbol{YU} = [\breve{y}', \boldsymbol{U}']'$. Define

$$\boldsymbol{B} = \begin{bmatrix} \breve{\delta} + \breve{d}\alpha w \\ \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\alpha w \end{bmatrix} \text{ and } \boldsymbol{\Omega} = \begin{bmatrix} \breve{d}\boldsymbol{\Gamma}\breve{d}' + \breve{\Sigma} & \breve{d}\boldsymbol{\Gamma}\boldsymbol{\lambda}' \\ \boldsymbol{\lambda}\boldsymbol{\Gamma}\breve{d}' & \boldsymbol{\lambda}\boldsymbol{\Gamma}\boldsymbol{\lambda}' + \boldsymbol{\Lambda} \end{bmatrix} \tag{10}$$

Then $\boldsymbol{YU} \sim MVN_{H+G+I}(\boldsymbol{B}, \boldsymbol{\Omega})$. $\tag{11}$

Table 1: Matrix notation, description, and dimension

Equation	Notation	Represents…	Dimension
Structural Equation	z^*	Vector of latent variables	$L \times 1$
	α	Matrix of exogenous variable loadings on z^*	$L \times \widetilde{D}$
	w	Vector of exogenous variables affecting z^*	$\widetilde{D} \times 1$
	η	Vector of errors in structural equation	$L \times 1$
Measurement Equation	Γ	Correlation matrix of error vector η in latent variable structural equation	$L \times L$
	$\breve{y}$	Vector of observed latent measurement equation dependent variables	$(G+H) \times 1$
	$\breve{\delta}$	Intercept vector in measurement equation	$(G+H) \times 1$
	$\breve{d}$	Vector of coefficients representing the effect of latent variables on observed/latent indicators	$(G+H) \times 1$
	$\breve{\xi}$	Vector of error in measurement equation	$(G+H) \times 1$
	$\breve{\Sigma}$	Covariance matrix of $\breve{\xi}$ (assumed diagonal for identification)	$(G+H) \times (G+H)$
Choice Model	U	Vector of alternative utilities	$I \times 1$
	β	Vector of exogenous variable effects on U	$D \times 1$
	x	Vector of exogenous variables in choice model	$I \times D$
	γ	Matrix of coefficients capturing effects of latent variables and their interactions with exogenous variables	$I \times \left[\sum_{i=1}^{I} N_i \right]$
	φ	Matrix of variables interacting with latent variables	$\left[\sum_{i=1}^{I} N_i \right] \times L$
	ε	Utility error vector	$I \times 1$
	Λ	Covariance matrix of ε	$I \times I$

All parameters to be estimated in the $\boldsymbol{B}$ vector and $\boldsymbol{\Omega}$ matrix are identifiable by ensuring that $\boldsymbol{\Gamma}$ is a correlation matrix, and $\breve{\boldsymbol{\Sigma}}$ is diagonal with the elements corresponding to the ordinal variables being normalized to 1. To see this, a helpful way is to first invoke Stapleton's (1978) identification conditions for the MIMIC model of Equation (8) (that includes the latent variable structural and measurement equations). That is, we ignore the information contained about the latent variables (and the parameters of the structural model of Equation (1)) through the utilities in the choice model of Equation (6). Then, following the excellent exposition in Stapleton (which we do not reproduce here to conserve on space), one can show that, under the conditions that (a) $\boldsymbol{\Gamma}$ is a correlation matrix, (b) $\breve{\boldsymbol{\Sigma}}$ is diagonal with the elements corresponding to the ordinal variables being normalized to 1, and (c) for each latent variable, there is at least one indicator variable that loads only on that latent variable and no other latent variable (that is, there is at least one factor complexity one indicator variable for each latent variable), the elements of the parameter vectors $\breve{\boldsymbol{\delta}}$ and $\breve{\boldsymbol{d}}$, and the elements of the matrices $\boldsymbol{\alpha}$, $\boldsymbol{\Gamma}$ and $\breve{\boldsymbol{\Sigma}}$, are all (over) identified. Next, we proceed to the choice model component. The covariance matrix of the reduced form of the choice model, ignoring the covariance between the MIMIC model and the choice model, is $\boldsymbol{\lambda}\boldsymbol{\Gamma}\boldsymbol{\lambda}' + \boldsymbol{\Lambda}$. With a general specification of $\boldsymbol{\Lambda}$ (but $\boldsymbol{\Lambda}$ should adhere to the conditions discussed in Section 2.3), and the identification of $\boldsymbol{\Gamma}$ from earlier, the choice model can provide estimates of all the elements of $\boldsymbol{\Lambda}$, but then none of the coefficients of the matrix $\boldsymbol{\gamma}$ embedded in $\boldsymbol{\lambda}$ are identified from this covariance matrix. However, $\boldsymbol{\lambda}$ appears both in the mean element $\boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\boldsymbol{\alpha}\boldsymbol{w}$ of the choice model component in Equation (10) as well as in the covariance term $\boldsymbol{\lambda}\boldsymbol{\Gamma}\breve{\boldsymbol{d}}'$ between the errors of the indicator equation and the choice equation. To examine identification of the $\boldsymbol{\beta}$ vector and the elements of the $\boldsymbol{\lambda}$ matrix, it is instructive to consider the simple but challenging identification case when the latent variables appear in the choice model without any interactions with other exogenous variables (if all latent variables appear only as interactions and not as direct shifters of utility, then there will be effectively no common exogenous variable effects through the direct $\boldsymbol{x}$ effect and through the indirect (via the latent variables) $\boldsymbol{w}$ effect, and so identification of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is immediate through the mean $\boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\lambda}\boldsymbol{\alpha}\boldsymbol{w}$). In the no interactions case, consider the specific case when each latent variable impacts the utility of each alternative (this corresponds to the situation when each $\boldsymbol{\varphi}_i$ matrix is an identity matrix of size L, and $\boldsymbol{\lambda} = \boldsymbol{\gamma}\boldsymbol{\varphi} = \tilde{\boldsymbol{\gamma}},$ where $\tilde{\boldsymbol{\gamma}}$ is of size $I \times L$ with each row comprising the coefficients representing the effects of the latent variable vector $\boldsymbol{z}^*$ on the utility of each alternative). In this case, if there are no common variables in each row of the $\boldsymbol{x}$ matrix and the $\boldsymbol{w}$ vector, then the $\boldsymbol{\beta}$ vector and the elements of the $\tilde{\boldsymbol{\gamma}}$ matrix are identifiable from the mean element $\boldsymbol{x}\boldsymbol{\beta} + \tilde{\boldsymbol{\gamma}}\boldsymbol{\alpha}\boldsymbol{w}$ in the choice model (since $\boldsymbol{\alpha}$ has already been identified from earlier). This is the most common way that identification has been achieved in most earlier ICLV studies, which use relatively "weak" structural equations and employ only individual-specific socioeconomic variables as explanatory factors of the latent variables. Alternatively, one may include common elements (including alternative-specific attributes in the choice model and those same variables in the structural model), but appropriate restrictions on the coefficients of the $\tilde{\boldsymbol{\gamma}}$ matrix (such as specifying no effects of the latent variables on one of the alternatives, and specifying generic coefficients

across all alternatives for those variables that are common, as we do in the simulation study in Equation (21)) will be adequate for identification from the mean element $x\beta + \widetilde{\gamma}\,\alpha w$ in the choice model. Of course, if the elements of $\lambda = \widetilde{\gamma}$ can be identified purely from the covariance matrix element $\lambda\Gamma\breve{d}'$ of Ω, then β can always be estimated even if all variables are common between each row of the x matrix and the w vector. However, general and convenient sufficiency rules for identification are not yet available in this general case of the ICLV model, and it is common place to resort to the more convenient case of ignoring the covariance term $\lambda\Gamma\breve{d}'$ when imposing sufficient (but not necessary) restrictions for identification. In this context, the issue of parameter identification in ICLV models remains an open research area, and few studies have discussed considerations related to the situation of overlapping explanatory variables in the structural equation and in the choice model. While we have provided sufficiency conditions based on the mean element $x\beta + \widetilde{\gamma}\,\alpha w$ in the choice model, these are likely to be restrictive and unnecessary because of identification possibilities through the covariance term $\lambda\Gamma\breve{d}'$. We leave a detailed examination of this issue for future research (the reader is referred to Vij and Walker, 2014, for a recent treatment of identification in ICLV models, though they do so by compartmentalizing the ICLV model into independent components).

To estimate the model, we need to develop the distribution of the vector $Yu = \left(\breve{y}', u^{*'}\right)' = \left(y', y^{*'}, u^{*'}\right)'$. To do so, define a matrix $\mathbf{M}$ of size $[G+H+I-1]\times[G+H+I]$. Fill this matrix up with values of zero. Then, insert an identity matrix of size $G+H$ into the first $G+H$ rows and $G+H$ columns of the matrix $\mathbf{M}$. Next, consider the last $(I-1)$ rows and last I columns, and insert an identity matrix of size $(I-1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Then, we can write $Yu \sim MVN_{H+G+I-1}(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\Omega}})$, where $\widetilde{\boldsymbol{B}} = \mathbf{M}\boldsymbol{B}$ and $\widetilde{\boldsymbol{\Omega}} = \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'.$ Next, partition the vector $\widetilde{\boldsymbol{B}}$ into components that correspond to the mean of the vectors y, y^*, and u^*, and the matrix $\widetilde{\boldsymbol{\Omega}}$ into the variances of y, y^*, and u^* and their covariances:

$$\widetilde{\boldsymbol{B}} = \begin{bmatrix} \widetilde{\boldsymbol{B}}_y \\ \widetilde{\boldsymbol{B}}_{y^*} \\ \widetilde{\boldsymbol{B}}_{u^*} \end{bmatrix} \text{ and } \widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}}_y & \widetilde{\boldsymbol{\Omega}}_{yy^*} & \widetilde{\boldsymbol{\Omega}}_{yu^*} \\ \widetilde{\boldsymbol{\Omega}}'_{yy^*} & \widetilde{\boldsymbol{\Omega}}_{y^*} & \widetilde{\boldsymbol{\Omega}}_{y^*u^*} \\ \widetilde{\boldsymbol{\Omega}}'_{yu^*} & \widetilde{\boldsymbol{\Omega}}'_{y^*u^*} & \widetilde{\boldsymbol{\Omega}}_{u^*} \end{bmatrix} \tag{12}$$

Define $\widetilde{u} = \left(y^{*'}, u^{*'}\right)'$, so that $Yu = (y', \widetilde{u}')'$. Re-partition $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{\Omega}}$ in a different way such that:

$$\widetilde{\boldsymbol{B}} = \begin{bmatrix} \widetilde{\boldsymbol{B}}_y \\ \widetilde{\boldsymbol{B}}_{\widetilde{u}} \end{bmatrix}, \text{ where } \widetilde{\boldsymbol{B}}_{\widetilde{u}} = \begin{bmatrix} \widetilde{\boldsymbol{B}}_{y^*} \\ \widetilde{\boldsymbol{B}}_{u^*} \end{bmatrix}, \text{ and} \tag{13}$$

$$\widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}}_y & \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}} \\ \widetilde{\boldsymbol{\Omega}}'_{y\widetilde{u}} & \widetilde{\boldsymbol{\Omega}}_{\widetilde{u}} \end{bmatrix}, \text{ where } \widetilde{\boldsymbol{\Omega}}_{\widetilde{u}} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}}_{y^*} & \widetilde{\boldsymbol{\Omega}}_{y^* u^*} \\ \widetilde{\boldsymbol{\Omega}}'_{y^* u^*} & \widetilde{\boldsymbol{\Omega}}_{u^*} \end{bmatrix} \text{ and } \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}}_{yy^*} & \widetilde{\boldsymbol{\Omega}}_{yu^*} \end{bmatrix}$$

The conditional distribution of $\widetilde{u}$, given y, is multivariate normal with mean $\boldsymbol{\ddot{B}}_{\widetilde{u}} = \widetilde{\boldsymbol{B}}_{\widetilde{u}} + \widetilde{\boldsymbol{\Omega}}'_{y\widetilde{u}} \widetilde{\boldsymbol{\Omega}}_y^{-1}(\boldsymbol{y} - \widetilde{\boldsymbol{B}}_y)$ and variance $\boldsymbol{\ddot{\Omega}}_{\widetilde{u}} = \widetilde{\boldsymbol{\Omega}}_{\widetilde{u}} - \widetilde{\boldsymbol{\Omega}}'_{y\widetilde{u}} \widetilde{\boldsymbol{\Omega}}_y^{-1} \widetilde{\boldsymbol{\Omega}}_{y\widetilde{u}}$. Next, supplement the threshold vectors defined earlier as follows: $\widetilde{\boldsymbol{\psi}}_{low} = \left[\boldsymbol{\psi}'_{low}, (-\infty_{I-1})'\right]'$, and

$\widetilde{\boldsymbol{\psi}}_{up} = \left[\boldsymbol{\psi}'_{up}, (\boldsymbol{0}_{I-1})'\right]'$, where $-\infty_{I-1}$ is a $(I-1)\times 1$-column vector of negative infinities, and $\boldsymbol{0}_{I-1}$ is another $(I-1)\times 1$-column vector of zeros. Then the likelihood function may be written as:

$$L(\boldsymbol{\theta}) = f_H(\boldsymbol{y} - \widetilde{\boldsymbol{B}}_y \mid \widetilde{\boldsymbol{\Omega}}_y) \times \Pr\left[\widetilde{\boldsymbol{\psi}}_{low} \leq \widetilde{u} \leq \widetilde{\boldsymbol{\psi}}_{up}\right], \tag{14}$$

$$= f_H(\boldsymbol{y} - \widetilde{\boldsymbol{B}}_y \mid \widetilde{\boldsymbol{\Omega}}_y) \times \int_{D_{\widetilde{u}}} f_{G+I-1}(\widetilde{u} \mid \boldsymbol{\ddot{B}}_{\widetilde{u}}, \boldsymbol{\ddot{\Omega}}_{\widetilde{u}}) du,$$

where the integration domain $D_{\widetilde{u}} = \{\widetilde{u} : \widetilde{\boldsymbol{\psi}}_{low} \leq \widetilde{u} \leq \widetilde{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\widetilde{u}$ vector determined by the observed ordinal indicator outcomes, and the range $(-\infty_{I-1}, 0)$ for the utility differences taken with respect to the utility of the observed choice outcome variable. $f_{G+I-1}(.)$ is the multivariate normal density function of dimension $G + I - 1$. The likelihood function for a sample of Q individuals is obtained as the product of the individual-level likelihood functions. If all the indicator variables are ordinal, then there is no $\boldsymbol{y}$ vector and the first term in the likelihood equation above drops out. On the other hand, if all the indicator variables are continuous, then there is no $\boldsymbol{y}^*$ vector in the $\widetilde{u}$ vector, and the dimension of integration drops to $I-1$.

The above likelihood function involves the evaluation of a $G + I - 1$ dimensional integral for each individual. As can be noticed, the dimensionality of integration does not increase with an increase in the number of latent variables in the model. This has been a major restrictive challenge in the typical way that ICLV models have been formulated and estimated, but not when we change to a probit kernel as we do here in this paper. Indeed, the independence of the integral dimensionality from the number of latent variables is remarkable, and should substantially enhance the applicability of the ICLV model. However, the dimensionality of the integration is still $G + I - 1$, which can be computationally expensive if there are several ordinal variables, or if the choice variable has a number of alternatives. So, the Maximum Approximate Composite Marginal Likelihood (MACML) approach of Bhat (2011), in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is used in this paper.[3]

---

[3] Note, however, that the MACML inference approach is very general and its use requires appropriate customization to the problem at hand. Thus, the MACML estimation we propose here is very different from the application of the MACML method for the simple cross-sectional multinomial probit model in Bhat (2011).

3.1 The MACML Estimation Approach

The MACML approach, similar to the parent CML approach (see Varin *et al.*, 2011, Lindsay *et al.*, 2011, Bhat, 2011, and Yi *et al.*, 2011 for recent reviews of CML approaches), maximizes a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods. The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), may be explained in a simple manner as follows. In the multi-dimensional model, instead of developing the likelihood function for the entire set of the choice outcome and observed ordinal indicators at once, as in Equation (14), one may compound (multiply) the probabilities of each pair of the choice outcome with an ordinal indicator, as well as the probabilities of each pair of ordinal indicators. The CML estimator (in this instance, the pairwise CML estimator) is then the one that maximizes the compounded probability of all pairwise events. The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004, Yi *et al.*, 2011). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191, Xu and Reid, 2011), the CML estimator is consistent and asymptotically normally distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood; for a formal proof, see Yi *et al.*, 2011 and Xu and Reid, 2011).

In the context of the proposed model, consider the following (pairwise) composite marginal likelihood function:

$$L_{CML}(\boldsymbol{\theta}) = f_H(\boldsymbol{y} - \widetilde{\boldsymbol{B}}_y \mid \widetilde{\boldsymbol{\Omega}}_y) \times \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr(j_g = n_g, j_{g'} = n_{g'}) \right) \times \left(\prod_{g=1}^{G} \Pr(j_g = n_g, i = m,) \right)$$

(15)

where i is an index for the individual's choice for the choice variable. In the above CML approach to estimating the ICLV model, the MVNCD function appearing in the CML function is of dimension equal to two for the second component in the equation above (corresponding to each pair of observed ordinal indicators), and equal to I for the probabilities corresponding to the third component in the equation above (corresponding to each pair of the choice outcome and an ordinal indicator outcome). In the MACML approach, we estimate the third set of components of the CML function involving the choice outcome (that is, those components that have I dimensions of integration) with an *analytic approximation* method rather than a simulation method. This combination of the CML with an analytic approximation for the MVNCD function is effective because it involves only univariate and bivariate cumulative normal distribution function evaluations. The MVNCD approximation method used here is based on linearization with binary variables (see Bhat, 2011). As has been demonstrated by Bhat and Sidharthan (2011), the MACML method has the virtue of computational robustness in that the approximate CML surface is smoother and easier to maximize than traditional simulation-based likelihood surfaces.[4]

---

[4] We should point out here that it is the combination of the probit based-kernel and the MACML approach that facilitates things in our approach. Specifically, the use of the probit based-kernel makes the reduced form utility

To explicitly write out the CML function in terms of the standard and bivariate standard normal density and cumulative distribution function, define $\boldsymbol{\omega}_{\Delta}$ as the diagonal matrix of standard deviations of matrix $\boldsymbol{\Delta}$, $\phi_R(.;\boldsymbol{\Delta}^{**})$ for the multivariate standard normal density function of dimension R and correlation matrix $\boldsymbol{\Delta}^*$ ($\boldsymbol{\Delta}^* = \boldsymbol{\omega}_{\Delta}^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_{\Delta}^{-1}$), and $\Phi_E(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal cumulative distribution function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$. Let $\boldsymbol{S}_g$ be a $I \times (G+I-1)$ selection matrix constructed as follows. To begin with, fill this matrix with values of zero for all elements. Then, position an element of '1' in the first row and the gth column. Also, position an identity matrix of size $I-1$ in the last $I-1$ rows and last $I-1$ columns. Let $\vartheta_{g,up} = \dfrac{\left[\widetilde{\boldsymbol{\psi}}_{up}\right]_g - \left[\ddot{\boldsymbol{B}}_{\widetilde{u}}\right]_g}{\sqrt{\left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{gg}}}$,

$\vartheta_{g,low} = \dfrac{\left[\widetilde{\boldsymbol{\psi}}_{low}\right]_g - \left[\ddot{\boldsymbol{B}}_{\widetilde{u}}\right]_g}{\sqrt{\left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{gg}}}$, $\qquad v_{gg'} = \dfrac{\left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{gg'}}{\sqrt{\left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{gg} * \left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{g'g'}}}$, $\qquad \breve{\boldsymbol{\psi}}_{g,low} = \boldsymbol{S}_g\left(\left[\boldsymbol{\psi}_{low}', (\boldsymbol{0}_{I-1})', \right]\right)'$,

$\breve{\boldsymbol{\psi}}_{g,up} = \boldsymbol{S}_g\widetilde{\boldsymbol{\psi}}_{up}$, $\breve{\boldsymbol{B}}_{g\widetilde{u}} = \boldsymbol{S}_g\ddot{\boldsymbol{B}}_{\widetilde{u}}$, and $\breve{\boldsymbol{\Omega}}_{g\widetilde{u}} = \boldsymbol{S}_g\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\boldsymbol{S}_g'$, where $\left[\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}\right]_{gg}$ represents the gg^{th} element of the matrix $\ddot{\boldsymbol{\Omega}}_{\widetilde{u}}$. Then, the CML function to be maximized is:

$$
\begin{aligned}
L_{CML}(\boldsymbol{\theta}) = &\left(\prod_{h=1}^{H}\omega_{\widetilde{\boldsymbol{\Omega}}_y}\right)^{-1}\phi_H\left(\left[\omega_{\widetilde{\boldsymbol{\Omega}}_y}\right]^{-1}\left[\boldsymbol{y} - \widetilde{\boldsymbol{B}}_y\right]; \widetilde{\boldsymbol{\Omega}}_y^*\right) \times \\
&\left(\prod_{g=1}^{G-1}\prod_{g'=g+1}^{G}\left[\begin{array}{l}\Phi_2(\vartheta_{g,up}, \vartheta_{g',up}, v_{gg'}) - \Phi_2(\vartheta_{g,up}, \vartheta_{g',low}, v_{gg'}) \\ -\Phi_2(\vartheta_{g,low}, \vartheta_{g',up}, v_{gg'}) + \Phi_2(\vartheta_{g,low}, \vartheta_{g',low}, v_{gg'})\end{array}\right]\right) \times \\
&\left(\prod_{g=1}^{G}\Phi_I\left[\omega_{\breve{\boldsymbol{\Omega}}_{g\widetilde{u}}}^{-1}\left\{\breve{\boldsymbol{\psi}}_{g,up} - \breve{\boldsymbol{B}}_{g\widetilde{u}}\right\}; \breve{\boldsymbol{\Omega}}_{g\widetilde{u}}^*\right] - \Phi_I\left[\omega_{\breve{\boldsymbol{\Omega}}_{g\widetilde{u}}}^{-1}\left\{\breve{\boldsymbol{\psi}}_{g,low} - \breve{\boldsymbol{B}}_{g\widetilde{u}}\right\}; \breve{\boldsymbol{\Omega}}_{g\widetilde{u}}^*\right]\right)
\end{aligned} \qquad (16)
$$

In the above expression, $[\vartheta]_g$ represents the gth element of the column vector ϑ, and similarly for other vectors. The reader will note that the expression in Equation (16) involves an MVNCD function evaluation that is utmost of dimension I. That is the dimensionality is purely a function of the number of alternatives in the choice model, irrespective of the number of latent variables or the number of ordinal indicators involved. Further, the MVNCD function of dimension I is itself approximated with the analytic approach in Bhat (2011), so that only univariate and bivariate normal cumulative distributions need to be evaluated. Write the resulting equivalent of Equation (16) as

---

functions in the choice model normally distributed, and develops a multivariate normally distributed form for the overall reduced forms in Equations (8) and (9). This is critical, because the CML approach cannot be used on Equations (8) and (9) if there were a mixing of a Gumbel error term with a normal error term as in the logit kernel-based ICLV model formulation (the CML as applied here is predicated on the fact that marginal distributions of subsets of multivariate normally distributed variables are also multivariate normally distributed). Second, the CML inference approach simplifies the function to be maximized (with respect to parameters), which reduces the dimensionality of integration from what would be if a maximum likelihood inference approach is used (see Equations (14) and (15)). Finally, the analytic approximation of the MACML comes in handy to evaluate the CML function of Equation (15).

$L_{MACML,q}(\boldsymbol{\theta})$, after introducing the index q for individuals. The MACML estimator is then obtained by maximizing the following function: $\log L_{MACML}(\boldsymbol{\theta}) = \sum_{q=1}^{Q} \log L_{MACML,q}(\boldsymbol{\theta})$. [5]

The covariance matrix of the parameters $\boldsymbol{\theta}$ may be estimated by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005).

$$V_{MACML}(\boldsymbol{\theta}) = \left[G(\boldsymbol{\theta})\right]^{-1} = [H(\boldsymbol{\theta})]^{-1}[J(\boldsymbol{\theta})][H(\boldsymbol{\theta})]^{-1},$$

where $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ can be estimated in a straightforward manner at the MACML estimate $\hat{\boldsymbol{\theta}}_{MACML}$ as follows:

$$\hat{H}(\hat{\boldsymbol{\theta}}) = -\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]_{\hat{\theta}_{MACML}}, \text{ and}$$

$$\hat{J}(\hat{\boldsymbol{\theta}}) = \sum_{q=1}^{Q} \left[\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)\right]_{\hat{\theta}_{MACML}}.$$

(17)

The code for the MACML estimation of the ICLV model is available at http://www.caee.utexas.edu/prof/bhat/CODES.htm.

3.2 Ensuring the Positive-Definiteness of Matrices

The $\widetilde{\boldsymbol{\Omega}}_y$ and $\bar{\boldsymbol{\Omega}}_{\tilde{u}}$ covariance matrix in the CML function need to be positive definite. This can be assured by ensuring that the covariance matrix $\widetilde{\boldsymbol{\Omega}}$ is positive definite, which itself requires that $\boldsymbol{\Omega}$ be positive definite. From Equation (10), $\boldsymbol{\Omega}$ will be positive definite if the matrices $\boldsymbol{\Gamma}$, $\breve{\boldsymbol{\Sigma}}$, and $\breve{\boldsymbol{\Lambda}}$ are positive definite. The simplest way to ensure the positive-definiteness of these matrices is to use a Cholesky-decomposition and parameterize the CML function in terms of the Cholesky parameters (rather than the original covariance matrices). Also, the matrix $\boldsymbol{\Gamma}$ is a correlation matrix, which can be maintained by writing each diagonal element (say the aa^{th} element) of the lower triangular Cholesky matrix of $\boldsymbol{\Gamma}$ as $\sqrt{1 - \sum_{j=1}^{a-1} d_{aj}^2}$, where the d_{aj} elements are the Cholesky factors that are to be estimated. In addition, note that the top diagonal element of $\breve{\boldsymbol{\Lambda}}$ has to be normalized to one (as discussed earlier in Section 2.3), which implies that the first

---

[5] In the case of repeated choice data from the same individual, the set of indicator variables are typically collected only once for an individual, which implies that the latent variables are fixed for an individual across the repeated choices (such as in a stated preference experiment). The third component of the composite marginal likelihood function for individual q in Equation (16) then is a product across all combinations of ordinal outcomes and choice occasions of the joint probability of each ordinal outcome and the outcome at each choice occasion. Thus, the difference from Equation (16) is that there is an additional outer product over T in the third component of Equation (16), where T is the number of choice occasions from the individual. In addition, there will be another fourth component in the CML function for repeated choice occasions that corresponds to the product of the likelihoods of each possible pairing of outcomes in the choice occasions. But the MACML estimation of this repeated choice model still entails the evaluation of only univariate and bivariate normal cumulative distributions.

element of the Cholesky matrix of $\breve{\Lambda}$ is fixed to the value of one. Finally, the matrix $\breve{\Sigma}$ is diagonal, and hence the Cholesky matrix of $\breve{\Sigma}$ is also diagonal (and comprises standard deviations of $\breve{\xi}$). The diagonals corresponding to ordinal variables in the Cholesky matrix of $\breve{\Sigma}$ are fixed to one for identification.

Chapter 4. Simulation Study

The simulation exercise undertaken in this section examines the ability of the MACML estimator to recover parameters from finite samples in an ICLV model of travel mode choice. The use of a simulated travel mode choice exercise is valuable because the true parameters underlying the data generating process (DGP) are set by the analyst, and the analyst can evaluate the behavior of the MACML estimator for different finite sample sizes. This is important to do when a new estimator is being proposed. Also, the framing of the simulation in the context of mode choice is purely for ease in interpretation and understanding; the results from the simulation exercise should be applicable to any other empirical context.

In the simulation experiments, we consider three modal choice alternatives in a weekday intercity travel context: Drive, air, and bus.

4.1 Experimental Design

In the latent variable structural equation model of Equation (2), consider five latent variables as follows: (1) Flexibility of travel by air (z_1^*), (2) flexibility of travel by bus (z_2^*), (3) comfort of travel by air (z_3^*), (4) comfort of travel by bus (z_4^*), and (5) environmental consciousness (z_5^*). Of the five variables above, the first four are qualitative attributes specific to two modes (air and bus), while the last variable is an individual-specific qualitative attribute (that does not vary across modes). Also, consider six variables in the observed covariate vector w to explain the latent variables: (1) Frequency of air service in the weekday (w_1), (2) frequency of bus service in the weekday (w_2), (3) travel time by air (w_3), (4) travel time by bus (w_4), (5) Income of traveler (w_5), and (6) Educational status of traveler (w_6). Then, we write Equation (2) as:

$$\begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \\ z_5^* \end{bmatrix} = \begin{bmatrix} 0.5 & 0.0 & 0.6 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & 0.0 & -0.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} \quad (18)$$

The α matrix indicates the observed covariates influencing each latent variable. Thus, for example, the first row of the α matrix indicates that the "flexibility of travel by air" (z_1^*) is affected by the frequency of air service in the weekday (w_1) and the travel time by air (w_3). The second row of the α matrix indicates that the "flexibility of travel by bus" (z_2^*) is affected by the frequency of bus service in the weekday (w_2) and the travel

time by bus (w_4). The effect of the frequency of service by a mode on the perception of flexibility for that mode is specified to be 0.5 for both the air and bus modes. Similarly, the effect of travel time by a mode on the perception of comfort for that mode is specified to be 0.6 for both the air and bus modes. The same is true for the effect of frequency of service on a mode on the comfort level of travel on that mode, with this parameter fixed to 0.3 for both the air and bus modes. The comfort perception for the bus mode (z_4^*) is negatively influenced by the income earnings of the individual (w_5) (see the "-0.4" entry in the fourth column and fourth row of the $\boldsymbol{\alpha}$ matrix), and environmental consciousness (z_5^*) is positively influenced by education status (w_6) (see the "0.8" entry in the final column and final row of the $\boldsymbol{\alpha}$ matrix). The parameters to be estimated in the $\boldsymbol{\alpha}$ matrix may be stacked up in a vector $\text{Vech}(\boldsymbol{\alpha}) = [\bar{\alpha}_1 = 0.5, \bar{\alpha}_2 = 0.6, \bar{\alpha}_3 = 0.5, \bar{\alpha}_4 = 0.6, \ \bar{\alpha}_5 = 0.3, \bar{\alpha}_6 = 0.3, \bar{\alpha}_7 = -0.4, \text{ and } \bar{\alpha}_8 = 0.8]$. The correlation matrix $\boldsymbol{\Gamma}$ of $\boldsymbol{\eta}$ is specified as follows:

$$\text{Var}(\boldsymbol{\eta}) = \boldsymbol{\Gamma} = \begin{bmatrix} 1.00 & 0.00 & 0.60 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.60 & 0.00 \\ 0.60 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.60 & 0.00 & 1.00 & 0.48 \\ 0.00 & 0.00 & 0.00 & 0.48 & 1.00 \end{bmatrix}$$

$$= \mathbf{L}_\Gamma \mathbf{L}_\Gamma' = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.6 & 0.0 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.6 & 0.8 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.8 & 0.6 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \end{bmatrix}$$

(19)

This generates a correlation in the unobserved factors impacting perceptions of flexibility and comfort for each of the air and bus modes. For ease, we maintain the same correlation coefficients between these two perceptions for each mode (as reflected by the value of 0.6 in the $\boldsymbol{\Gamma}$ matrix in the first two rows). We also specify a correlation coefficient of 0.48 for the perceptions of comfort on the bus mode and environmental consciousness, to reflect the notion that those who are environmentally conscious may be more likely to view the bus mode as being comfortable than those less environmentally conscious. Thus, there are three parameters in the $\boldsymbol{\Gamma}$ matrix. As indicated earlier, to maintain positive definiteness, we work with the Cholesky decomposition elements. Then, there are three Cholesky matrix elements to be estimated in $\mathbf{L}_\Gamma$ ($l_{\Gamma 1} = 0.6$, $l_{\Gamma 2} = 0.6$, and $l_{\Gamma 3} = 0.6$), corresponding to the non-diagonal elements in the matrix (note that the diagonal elements are simply a function of the non-diagonal elements and are not estimated directly, because $\boldsymbol{\Gamma}$ is a correlation matrix with unit diagonals; see Section 3.2). Collectively, the three elements to be estimated in $\boldsymbol{\Gamma}$, vertically stacked into a column vector, will be referred to as l_Γ.

In the latent variable measurement model of Equation (5), we assume one continuous indicator variable and four ordinal indicator variables: (1) Number of miles of non-motorized travel on a typical day y (the continuous indicator variable), (2) Ease of use of the air mode (y_1^*), (3) Ease of use of the bus mode (y_2^*), (4) level of relaxation on the air mode (y_3^*), and (5) level of relaxation on the rail mode (y_4^*). In the simulation experiments, we set the elements of the $\breve{\delta}$ vector to the value of '1' for the continuous variable and '-1' for the remaining four ordinal variables. We assume that the number of miles of non-motorized travel on a typical day (y) is an indicator for environmental consciousness (z_5^*). The ease of use of the air mode (y_1^*) is a reflection of flexibility of travel by air (z_1^*), while the ease of use of the bus mode (y_2^*) is a reflection of flexibility of travel by bus (z_2^*). Similarly, the level of relaxation on each mode $(y_3^*$ and $y_4^*)$ is considered as an indicator variable for comfort of travel by that mode $(z_3^*$ and $z_4^*)$, respectively. Then, we write Equation (5) as:

$$
\begin{bmatrix} y \\ y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 0.3 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.6 & 0.0 \end{bmatrix} \begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \\ z_5^* \end{bmatrix} + \begin{bmatrix} \tilde{\xi} \\ \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} \tag{20}
$$

The $\breve{\delta}$ vector to be estimated has five elements: $\breve{\delta} = (\tilde{\delta} = 1, \delta_1 = -1, \delta_2 = -1, \delta_3 = -1, \delta_4 = -1)$, and the $\breve{d}$ matrix elements to be estimated also has five elements: $\breve{d} = (\tilde{d} = 0.2, d_1 = 0.3, d_2 = 0.4, d_3 = 0.5, d_4 = 0.6)$. The four ordinal variables are measured on a three point scale, so that $\psi_g = \psi_{g,2}$ for each ordinal variable g $(g=1,2,3,4)$, and $\psi = (\psi_1', \psi_2', \psi_3', \psi_4')'$. We set each of these thresholds to a value of 1.5. Thus, there are a total of four threshold parameters to estimate (with true parameter values of 1.5) across all the ordinal variables.[6]

The covariance matrix $\breve{\Sigma}$ of $\breve{\xi}$ has to be diagonal for identification, with the elements corresponding to the ordinal variables being normalized to 1. So, the only element of $\breve{\Sigma}$

---

[6] In our simulation experiment, we specify as many indicators as the number of latent variables. In typical models, the number of indicators will exceed the number of latent variables. However, the intent here is to show that our estimation procedure can easily handle many latent variables, as opposed to most earlier ICLV models that specify one or two latent variables (see Daly *et al.*, 2012 or Bolduc *et al.* 2005 or Alvarez-Daziano and Bolduc, 2013). But, in many empirical contexts, latent variables are likely to be alternative-specific, which implies more than the traditional one or two latent variables. So, our emphasis in the simulation design was on incorporating five latent variables. On the other hand, having multiple indicators for each latent variable does not affect the dimensionality of integration either in the logit kernel-based ICLV model or our new model approach, though it increases the number of parameters to be estimated. Thus, given that we already have 38 parameters to estimate in the design, we have used one indicator for each latent variable to focus the simulation effort on showing how our approach is easily able to accommodate the realistic mode choice case of a high number of latent variables.

to be estimated is the first element ($=\text{Var}\,(\widetilde{\xi})$). We set this value as 1, and estimate the standard error $l_{\widetilde{\xi}}$ (for consistency with the Cholesky matrix elements in other covariance matrices) that also takes a true value of 1.

Finally, we specify the choice model (Equation 7) as follows:

$$
\begin{bmatrix} U_{car} \\ U_{air} \\ U_{bus} \end{bmatrix} = \begin{bmatrix} 0 & 0 & TT_{car} & TC_{car} \\ 1 & 0 & TT_{air} & TC_{air} \\ 0 & 1 & TT_{bus} & TC_{bus} \end{bmatrix} \begin{bmatrix} \beta_{ASC,air} = 0.5 \\ \beta_{ASC,bus} = -1.0 \\ \beta_{TT} = -1.0 \\ \beta_{TC} = -0.8 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 & -0.5 \\ 0.0 & 0.2 & 0.0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} z_1^* \\ z_2^* \\ z_3^* \\ z_4^* \\ z_5^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}
$$

(21)

In the above equation, the parameters to be estimated include the elements of the $\boldsymbol{\beta}$ vector ($\beta_{ASC,air}=0.5$, $\beta_{ASC,bus}=-1.0$, $\beta_{TT}=-1.0$, and $\beta_{TC}=-0.8$) and the elements of the γ matrix stacked up in a vector $\text{Vech}(\gamma)=[\overline{\gamma}_1=0.5, \overline{\gamma}_2=0.5, \overline{\gamma}_3=-0.5, \overline{\gamma}_4=0.2, \overline{\gamma}_5=0.2,\ \text{and}\ \overline{\gamma}_6=0.3].$ [7]

Next, we specify the covariance matrix $\boldsymbol{\Lambda}$. for the error vector $\boldsymbol{\varepsilon}$ as

$$
\boldsymbol{\Lambda} = \begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.60 \\ 0.00 & 0.60 & 1.36 \end{bmatrix} = \mathbf{L}_{\boldsymbol{\Lambda}}\mathbf{L}'_{\boldsymbol{\Lambda}} = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix}\begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.6 \\ 0.0 & 0.0 & 1.0 \end{bmatrix},
$$

(22)

There are two Cholesky matrix elements to be estimated in $\mathbf{L}_{\boldsymbol{\Lambda}}$ ($l_{\Lambda 1}=0.6$, $l_{\Lambda 2}=1.0$). Collectively, these two elements, vertically stacked into a column vector, will be referred to as $l_{\boldsymbol{\Lambda}}$.

To complete the simulation design, we draw values for the elements of the vector w (i.e., for w_1, w_2, w_3, w_4, w_5, and w_6), and for the travel time and travel cost variables for each mode, from independent standard continuous uniform distributions. We consider different samples sizes to assess the accuracy and appropriateness of the asymptotic properties of the MACML estimator for finite sample sizes. In particular, we construct

---

[7] In the general notation of Equation (7),

$$
\gamma = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & -0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.2 & 0.2 & 0.3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varphi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.
$$

$\boldsymbol{\varphi}$ is fixed and pre-specified based on the discussion in Section 2.3 and Equation (21), and the elements of the γ matrix are to be estimated.

synthetic simulated data samples of 500, 1000, and 2000 by drawing the corresponding number of realizations of the exogenous variables. Once drawn, the exogenous variables are held fixed for the rest of the simulation exercise. Next, for each data sample, the mean $\boldsymbol{B}$ and covariance matrix $\boldsymbol{\Omega}$ for the vector $\boldsymbol{YU}$ for each observation are computed based on Equation (10). Then, for each observation, we draw a realization of $\boldsymbol{YU}$ from its multivariate distribution. The first five elements of the realization of the $\boldsymbol{YU}$ vector for each observation correspond to y, y_1^*, y_2^*, y_3^*, and y_4^*. The value for y is retained as is, and constitutes the continuous indicator value for each observation. The values for y_1^*, y_2^*, y_3^*, and y_4^* are compared to the threshold values of 0 and 1.5 and, based on this comparison, ordinal indicator variables are assigned for each of the four ordinal variables for each observation. Next, the last three elements of $\boldsymbol{YU}$ correspond to U_{car}, U_{air}, and U_{bus}. The alternative with the highest utility is selected and designated as the chosen alternative for each observation. With this, a complete data set for each sample size (of 500, 1000, 2000) is generated from which to estimate the following 38 parameters: $\overline{\alpha}_1, \overline{\alpha}_2, \overline{\alpha}_3, \overline{\alpha}_4, \overline{\alpha}_5, \overline{\alpha}_6, \overline{\alpha}_7, \overline{\alpha}_8,\ l_{\Gamma 1}, l_{\Gamma 2}, l_{\Gamma 3},\ \widetilde{\delta}, \delta_1, \delta_2, \delta_3, \delta_4,\ \widetilde{d}, d_1, d_2, d_3, d_4, \psi_1, \psi_2, \psi_3, \psi_4,$ $l_{\widetilde{\xi}}, \beta_{ASC,air}, \beta_{ASC,bus}, \beta_{TT}, \beta_{TC}, \overline{\gamma}_1, \overline{\gamma}_2, \overline{\gamma}_3, \overline{\gamma}_4, \overline{\gamma}_5, \overline{\gamma}_6\ l_{\Lambda 1},$ and $l_{\Lambda 2}$.

For each sample size, the above data generation process is undertaken 50 times with different realizations of the $\boldsymbol{YU}$ vector to generate 50 different data sets. After each data generation, we checked to ensure that there were adequate observations that "chose" each ordinal outcome for the ordinal indicator variables and "chose" each alternative for the choice variable. The estimator is then applied to each data set to estimate data specific values for the 38 parameters. A single random permutation is generated for each individual (the random permutation varies across individuals, but is the same across iterations for a given individual) to decompose the multivariate normal cumulative distribution (MVNCD) function into a product sequence of marginal and conditional probabilities (see Section 2.1 of Bhat, 2011).[8] The estimator is applied to each dataset 10 times with different permutations to obtain the approximation error. Thus, we run 500 estimations for each sample size (50 data sets × 10 runs with different permutations per data set). Given the three sample sizes, there are a total of 500×3=1500 estimations undertaken.

4.2 Performance Evaluation

The performance of the MACML inference approach in estimating the parameters of the proposed model and the corresponding standard errors is evaluated, for each sample size, as follows:

(1) Estimate the MACML parameters for each data set and for each of 10 independent sets of permutations. Estimate the standard errors (s.e.) using the Godambe (sandwich) estimator.

---

[8] Technically, the MVNCD approximation should improve with a higher number of permutations in the MACML approach. However, when we investigated the effect of different numbers of random permutations per individual, we noticed little difference in the estimation results between using a single permutation and higher numbers of permutations, and hence we settled with a single permutation per individual.

(2) For each data set *s*, compute the mean estimate for each model parameter across the 10 random permutations used. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage (finite sample) bias** (APB) of the estimator as:

$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$$

(3) Compute the standard deviation of the MED values across the 50 datasets, and label this as the **finite sample standard error or FSEE** (essentially, this is the empirical standard error). Calculate the FSEE as a percentage of the mean estimate.

(4) For each data set, compute the mean standard error for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 50 data sets and label this as **the asymptotic standard error or ASE** (essentially this is the standard error of the distribution of the estimator as the sample size gets large). Compute the ASE as a percentage of the mean estimate.

(5) Next, to evaluate the accuracy of the asymptotic standard error formula as computed using the MACML inference approach for the finite sample size used, compute the relative efficiency of the estimator as:

$$\text{Relative efficiency} = \frac{ASE}{FSEE} \tag{23}$$

Relative efficiency values in the range of 0.75-1.25 indicate that the ASE, as computed using the Godambe matrix in the MACML method, does provide a good approximation of the FSSE. In general, the relative efficiency values should be less than 1, since we expect the asymptotic standard error to be less than the FSSE. But, because we are using only a limited number of data sets to compute the FSSE, values higher than one can also occur. The more important point is to examine the closeness between the ASE and FSEE, as captured by the 0.75-1.25 range for the relative efficiency value.

(6) Compute the standard deviation of the parameter values around the MED parameter value for each data set, and take the mean of this standard deviation value across the data sets; label this as the **approximation error (APERR)**.

Chapter 5. Simulation Results

Tables 2, 3, and 4 provide the simulation results for the 500, 1000, and 2000 observation cases, respectively. The tables provide the true value of the parameters (second column), followed by the parameter estimate results and the standard error estimate results.

Several observations may be made from the tables. *First*, the MACML procedure is able to recover the parameters remarkably well even with only 500 observations, with the APB value having a mean (across parameters) value of 5.10% (see the final row of Table 2 labeled "Overall mean value across parameters" under the column "Absolute Percentage Bias (APB)"). The individual parameter APB values range from 0.3% to 14%, though even the seemingly large 14% APB is rather deceiving since the true estimate for this parameter (the $\tilde{d}$ parameter) is 0.20 and the mean estimate of the parameter is 0.228. Thus, the absolute finite sample bias is only 0.028, but gets inflated in percentage because of the small magnitude of the true value for the parameter. The APB values also, in general, reduce steadily (but rather marginally) with an increase in sample size. The mean APB value reduces to 4.807% (with a range of 0.133% to 15.5%) with 1000 observations (see Table 3) and further to 4.158% (with a range of 0% to 10.4%) with 2000 observations (see Table 4).[9] *Second*, across the different sample sizes, the mean APB values for the $\breve{d}$ $[=(\tilde{d}, d_1, d_2, d_3, d_4)]$ vector elements and for the γ matrix elements $(\text{Vech}(\gamma) = [\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3, \bar{\gamma}_4, \bar{\gamma}_5, \text{ and } \bar{\gamma}_6])$ are consistently higher than the overall mean APB. In particular, the mean APB values for the $\breve{d}$ vector elements are 8.413%, 8.860%, and 4.813% for the 500, 1000, and 2000 observations cases, and the corresponding mean APB values for the γ matrix elements are 7.939%, 6.035%, and 4.950%. This suggests that the coefficients on the latent variable vector z^* in the measurement equation (Equation 5) and in the choice model (Equation 7) are somewhat more difficult to recover than other parameters. This is not surprising, since these elements enter into the covariance matrix Ω in a non-linear fashion (see Equation 10), and Ω itself enters into the composite likelihood function (Equation 15) in a complex manner. *Third*, the finite sample standard errors (FSSE) clearly decrease as the sample size increases. As a percentage of the mean estimate, the FSSE indicates an average value (across all parameters) of 47.7% for the case of 500 observations, 35.7% for 1000 observations, and 22.4% for 2000 observations. The same trend is observed for the asymptotic standard error (ASE), with the ASE (as a percentage of the mean estimate) having a mean value (across all parameters) of 52.2% for 500 observations, 37.8% for 1000 observations, and 23.2% for 2000 observations. In general, these results indicate good empirical efficiency of the MACML estimator,

---

[9] Note that the CML estimator score functions are unbiased and the CML estimator has the asymptotic properties of being consistent and normally distributed under usual regularity conditions (see Bhat, 2011, Xu and Reid, 2011, and Bhat, 2014). Thus, it is not very surprising that the APB does not change too much with an increase in the number of observations (after all, we are computing the mean value of parameters across 500 realizations for each sample size). Indeed, this is a distinct advantage of the CML for the ICLV relative to MSL and other simulation techniques. That is, the CML reduces the dimensionality while retaining the consistency property of the estimator. Then, because we can evaluate the reduced-dimension CML function (three-dimensional in our simulation experiment) very accurately using the analytic approximation for the MVNCD function, we are able to recover parameters very well, as reflected in the small APB values regardless of sample size.

especially when the sample size is of the order of 1000 or more. But, the FSSE and ASE values are particularly high for the γ matrix elements, with the values (as a percentage of the mean estimate) being 86.7% (500 observations), 69.5% (1000 observations), and 44.3% (2000 observations) for the FSEE, and 90.1% (500 observations), 72.2% (1000 observations), and 43.3% (2000 observations) for the ASE. This suggests some caution in estimating models with small sample sizes when there are many latent variables. *Fourth*, the FSEE and the ASE values are close to one another regardless of sample size, with the relative efficiency (RE) value between 0.75-1.25 for all parameters in the case of sample sizes of 1000 and 2000, and the value between 0.75-1.25 for all but one parameter in the case of a sample size of 500. Overall, across all parameters, the average relative efficiency is in the range of 1.06-1.10 for all sample sizes, indicating that the asymptotic formula is performing well in estimating the finite sample standard error even for a sample size of the order of 500. Finally, the last columns of the tables present the approximation error (APERR) for each of the parameters, because of the use of different permutations. These entries indicate that the APERR reduces with sample size, as expected. However, even for the case of 500 observations, the average APERR is only 0.030 and the maximum is only 0.078. More importantly, the approximation error (as a percentage of the FSEE or the ASE), averaged across all the parameters, is only of the order of 12% of the sampling error. This statistic decreases to 8% when the sample size increases to 1000 or 2000. The implication is that even a single permutation (per observation) of the approximation approach used to evaluate the MVNCD function provides adequate precision, in the sense that the convergent values are about the same for a given data set regardless of the permutation used for the decomposition of the multivariate probability expression.

The convergence time for the proposed approach has a median value of about 40 minutes for the case of 500 observations, one hour for the case of 1000 observations, and 80 minutes for the case of 2000 observations, all based on scaling to a desktop computer with an Intel Core™ i7 860@2.80GHz processor and 8GB of RAM. The time to compute the covariance matrix of the parameters was of the same order as the time for convergence. However, a more thorough analysis of computational times is warranted using a single machine. Further, it would be beneficial in a future study to run a rigorous exercise to compare, with a traditional MSL approach, the accuracy and econometric efficiency of our MACML estimator in finite samples, as well as computational times, even if with a much simpler model specification than that used in the current paper. Indeed, our main purpose in this study was to propose and apply the new method for a realistic specification rather than a "toy" specification. The important point is that completing 1500 estimations of ICLV models with (a) the very general specifications for the covariance matrices adopted here, (b) the presence of five latent variables, and (c) the presence of five indicators (four of which are ordinal) is literally infeasible with the traditional logit kernel specification and the maximum simulated likelihood (MSL) estimation approach, at least with the computer hardware that we had at our disposal for this research (which is also the kind of computer hardware typically available to most analysts). The typical approach is notorious for very long estimation times (if convergence is achieved at all), with durations of 15 hours or more not at all uncommon. The order of magnitude reduction in computation time resulting from the use of our approach can then be used by analysts to explore a wide array of observed and latent variable specifications, rather than examining just a few specifications and settling quickly on one that may not be the best.

Table 2: Simulation results for the 500 observations case with 50 datasets (based on a total of 50×10 runs/dataset=500 runs)

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
$\overline{\alpha}_1$	0.5	0.482	0.018	3.600	0.284	0.277	0.978	0.0350
$\overline{\alpha}_2$	0.6	0.588	0.012	2.000	0.203	0.225	1.105	0.0388
$\overline{\alpha}_3$	0.5	0.511	0.011	2.200	0.290	0.334	1.155	0.0298
$\overline{\alpha}_4$	0.6	0.622	0.022	3.667	0.303	0.324	1.070	0.0354
$\overline{\alpha}_5$	0.3	0.317	0.017	5.667	0.272	0.300	1.104	0.0237
$\overline{\alpha}_6$	0.3	0.314	0.014	4.667	0.133	0.161	1.207	0.0176
$\overline{\alpha}_7$	-0.4	-0.413	0.013	3.250	0.203	0.238	1.173	0.0246
$\overline{\alpha}_8$	0.8	0.843	0.043	5.375	0.351	0.385	1.097	0.0201
$l_{\Gamma 1}$	0.6	0.573	0.027	4.500	0.266	0.296	1.114	0.0400
$l_{\Gamma 2}$	0.6	0.537	0.063	10.500	0.227	0.249	1.099	0.0282
$l_{\Gamma 3}$	0.6	0.576	0.024	4.000	0.256	0.316	1.235	0.0205
$\widetilde{\delta}$	1.0	1.004	0.004	0.400	0.056	0.066	1.193	0.0023
δ_1	-1.0	-1.029	0.029	2.900	0.231	0.255	1.104	0.0122
δ_2	-1.0	-1.060	0.060	6.000	0.239	0.274	1.148	0.0512
δ_3	-1.0	-1.027	0.027	2.700	0.262	0.279	1.065	0.0262
δ_4	-1.0	-1.024	0.024	2.400	0.174	0.200	1.150	0.0250
$\widetilde{d}$	0.2	0.228	0.028	14.000	0.156	0.175	1.120	0.0047
d_1	0.3	0.330	0.030	10.000	0.163	0.184	1.131	0.0249
d_2	0.4	0.428	0.028	7.000	0.311	0.375	1.206	0.0780
d_3	0.5	0.512	0.012	2.400	0.342	0.395	1.156	0.0422

Table 2: (Continued) Simulation results for the 500 observation case with 50 datasets (based on a total of 50×10 runs/dataset=500 runs)

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
d_4	0.6	0.652	0.052	8.667	0.211	0.268	1.268	0.0440
ψ_1	1.5	1.479	0.021	1.400	0.136	0.153	1.126	0.0125
ψ_2	1.5	1.572	0.072	4.800	0.283	0.296	1.044	0.0678
ψ_3	1.5	1.538	0.038	2.533	0.315	0.325	1.031	0.0345
ψ_4	1.5	1.511	0.011	0.733	0.217	0.243	1.117	0.0316
$l_{\xi 1}$	1.0	0.967	0.033	3.300	0.099	0.088	0.882	0.0012
$\beta_{ASC,air}$	0.5	0.498	0.002	0.400	0.378	0.428	1.133	0.0174
$\beta_{ASC,bus}$	-1.0	-0.997	0.003	0.300	0.476	0.569	1.195	0.0275
β_{TT}	-1.0	-1.071	0.071	7.100	0.250	0.256	1.023	0.0298
β_{TC}	-0.8	-0.869	0.069	8.625	0.196	0.209	1.066	0.0241
$\bar{\gamma}_1$	0.5	0.540	0.040	8.000	0.214	0.258	1.204	0.0482
$\bar{\gamma}_2$	0.5	0.530	0.030	6.000	0.483	0.493	1.021	0.0475
$\bar{\gamma}_3$	-0.5	-0.496	0.004	0.800	0.212	0.257	1.209	0.0216
$\bar{\gamma}_4$	0.2	0.221	0.021	10.500	0.244	0.287	1.174	0.0233
$\bar{\gamma}_5$	0.2	0.222	0.022	11.000	0.327	0.269	0.823	0.0385
$\bar{\gamma}_6$	0.3	0.334	0.034	11.333	0.297	0.326	1.095	0.0152
$l_{\Lambda 1}$	0.6	0.621	0.021	3.500	0.459	0.460	1.002	0.0397
$l_{\Lambda 2}$	1.0	0.924	0.076	7.600	0.430	0.434	1.011	0.0353
Overall mean value across parameters		0.030	5.100		0.262	0.288	1.106	0.0300

Table 3: Simulation results for the 1000 observations case with 50 datasets (based on a total of 50×10 runs/dataset=500 runs)

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
$\bar{\alpha}_1$	0.5	0.496	0.004	0.800	0.136	0.161	1.184	0.0209
$\bar{\alpha}_2$	0.6	0.574	0.026	4.333	0.158	0.193	1.222	0.0243
$\bar{\alpha}_3$	0.5	0.454	0.046	9.200	0.180	0.208	1.156	0.0152
$\bar{\alpha}_4$	0.6	0.537	0.063	10.500	0.176	0.194	1.097	0.0171
$\bar{\alpha}_5$	0.3	0.311	0.011	3.667	0.211	0.222	1.047	0.0118
$\bar{\alpha}_6$	0.3	0.319	0.019	6.333	0.189	0.206	1.090	0.0118
$\bar{\alpha}_7$	-0.4	-0.409	0.009	2.250	0.179	0.176	0.983	0.0161
$\bar{\alpha}_8$	0.8	0.824	0.024	3.000	0.279	0.292	1.047	0.0142
$l_{\Gamma 1}$	0.6	0.585	0.015	2.500	0.198	0.184	0.929	0.0187
$l_{\Gamma 2}$	0.6	0.606	0.006	1.000	0.154	0.134	0.870	0.0111
$l_{\Gamma 3}$	0.6	0.550	0.050	8.333	0.217	0.191	0.876	0.0139
$\tilde{\delta}$	1.0	0.988	0.012	1.200	0.047	0.044	0.957	0.0015
δ_1	-1.0	-1.021	0.021	2.100	0.149	0.120	0.805	0.0069
δ_2	-1.0	-1.056	0.056	5.600	0.189	0.191	1.011	0.0230
δ_3	-1.0	-1.020	0.020	2.000	0.131	0.138	1.062	0.0141
δ_4	-1.0	-1.029	0.029	2.900	0.174	0.194	1.115	0.0131
$\tilde{d}$	0.2	0.231	0.031	15.500	0.079	0.090	1.139	0.0020
d_1	0.3	0.320	0.020	6.667	0.066	0.079	1.197	0.0171
d_2	0.4	0.424	0.024	6.000	0.248	0.255	1.028	0.0344
d_3	0.5	0.544	0.044	8.800	0.192	0.234	1.219	0.0251

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
d_4	0.6	0.644	0.044	7.333	0.226	0.269	1.190	0.0237
ψ_1	1.5	1.453	0.047	3.133	0.090	0.098	1.089	0.0084
ψ_2	1.5	1.529	0.029	1.933	0.226	0.219	0.969	0.0287
ψ_3	1.5	1.491	0.009	0.600	0.127	0.137	1.079	0.0165
ψ_4	1.5	1.498	0.002	0.133	0.169	0.200	1.183	0.0156
$l_{\xi 1}$	1.0	0.983	0.017	1.700	0.031	0.035	1.129	0.0006
$\beta_{ASC,air}$	0.5	0.474	0.026	5.200	0.279	0.326	1.168	0.0119
$\beta_{ASC,bus}$	-1.0	-1.036	0.036	3.600	0.316	0.306	0.968	0.0195
β_{TT}	-1.0	-1.064	0.064	6.400	0.182	0.193	1.060	0.0201
β_{TC}	-0.8	-0.855	0.055	6.875	0.147	0.158	1.082	0.0158
$\bar{\gamma}_1$	0.5	0.491	0.009	1.800	0.205	0.221	1.078	0.0364
$\bar{\gamma}_2$	0.5	0.522	0.022	4.473	0.311	0.337	1.084	0.0242
$\bar{\gamma}_3$	-0.5	-0.521	0.021	4.271	0.227	0.234	1.026	0.0124
$\bar{\gamma}_4$	0.2	0.220	0.020	10.000	0.164	0.188	1.146	0.0121
$\bar{\gamma}_5$	0.2	0.220	0.020	10.000	0.305	0.299	0.980	0.0102
$\bar{\gamma}_6$	0.3	0.317	0.017	5.667	0.187	0.183	0.979	0.0089
$l_{\Lambda 1}$	0.6	0.593	0.007	1.167	0.376	0.364	0.968	0.0207
$l_{\Lambda 2}$	1.0	0.943	0.057	5.700	0.270	0.311	1.152	0.0198
Overall mean value across parameters		0.027	4.807	0.189	0.200	1.062	0.0163

Table 4: Simulation results for the 2000 observations case with 50 datasets (based on a total of 50×10 runs/dataset=500 runs)

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
$\bar{\alpha}_1$	0.5	0.4	0.019	3.800	0.098	0.103	1.051	0.0082
$\bar{\alpha}_2$	0.6	0.563	0.037	6.167	0.135	0.141	1.044	0.0099
$\bar{\alpha}_3$	0.5	0.448	0.052	10.400	0.100	0.118	1.180	0.0011
$\bar{\alpha}_4$	0.6	0.562	0.038	6.333	0.101	0.113	1.108	0.0128
$\bar{\alpha}_5$	0.3	0.290	0.010	3.333	0.068	0.054	0.806	0.0051
$\bar{\alpha}_6$	0.3	0.278	0.022	7.333	0.075	0.076	1.013	0.0058
$\bar{\alpha}_7$	-0.4	-0.408	0.008	2.000	0.124	0.136	1.097	0.0096
$\bar{\alpha}_8$	0.8	0.863	0.063	7.875	0.216	0.212	0.981	0.0072
$l_{\Gamma 1}$	0.6	0.606	0.006	1.000	0.069	0.051	0.750	0.0139
$l_{\Gamma 2}$	0.6	0.623	0.023	3.833	0.096	0.089	0.927	0.0086
$l_{\Gamma 3}$	0.6	0.541	0.059	9.833	0.158	0.170	1.076	0.0130
$\tilde{\delta}$	1.0	0.999	0.001	0.100	0.033	0.032	0.970	0.0011
δ_1	-1.0	-1.081	0.081	8.100	0.098	0.101	1.031	0.0031
δ_2	-1.0	-1.024	0.024	2.400	0.106	0.116	1.094	0.0166
δ_3	-1.0	-1.037	0.037	3.700	0.092	0.105	1.130	0.0108
δ_4	-1.0	-1.020	0.020	2.000	0.114	0.132	1.158	0.0077
$\tilde{d}$	0.2	0.200	0.000	0.000	0.052	0.058	1.115	0.0013
d_1	0.3	0.321	0.021	7.000	0.061	0.066	1.082	0.0070
d_2	0.4	0.422	0.022	5.500	0.124	0.125	1.008	0.0277
d_3	0.5	0.532	0.032	6.400	0.125	0.141	1.128	0.0174

Table 4: (Continued) Simulation results for the 2000 observations case with 50 datasets (based on a total of 50×10 runs/dataset=500 runs)

Parameters	True value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Err. (FSSE)	Asymptotic St. Err. (ASE)	Relative Efficiency	Approximation Error (APERR)
d_4	0.6	0.631	0.031	5.167	0.135	0.145	1.149	0.0153
ψ_1	1.5	1.461	0.039	2.600	0.056	0.066	1.179	0.0034
ψ_2	1.5	1.493	0.007	0.467	0.109	0.129	1.183	0.0218
ψ_3	1.5	1.503	0.003	0.200	0.113	0.122	1.080	0.0134
ψ_4	1.5	1.483	0.017	1.133	0.120	0.141	1.175	0.0098
$l_{\xi 1}$	1.0	0.997	0.003	0.300	0.016	0.018	1.125	0.0003
$\beta_{ASC,air}$	0.5	0.527	0.027	5.400	0.180	0.203	1.128	0.0086
$\beta_{ASC,bus}$	-1.0	-1.026	0.026	2.600	0.223	0.217	0.973	0.0157
β_{TT}	-1.0	-1.030	0.030	3.000	0.129	0.149	1.155	0.0141
β_{TC}	-0.8	-0.821	0.021	2.625	0.108	0.121	1.120	0.0115
$\bar{\gamma}_1$	0.5	0.481	0.019	3.800	0.141	0.168	1.191	0.0144
$\bar{\gamma}_2$	0.5	0.474	0.026	5.200	0.226	0.230	1.013	0.0198
$\bar{\gamma}_3$	-0.5	-0.501	0.001	0.200	0.165	0.170	1.030	0.0100
$\bar{\gamma}_4$	0.2	0.219	0.019	9.500	0.086	0.092	1.070	0.0098
$\bar{\gamma}_5$	0.2	0.216	0.016	8.000	0.160	0.127	0.794	0.0059
$\bar{\gamma}_6$	0.3	0.309	0.009	3.000	0.132	0.132	1.000	0.0077
$l_{\Lambda 1}$	0.6	0.609	0.009	1.500	0.247	0.246	0.992	0.0183
$l_{\Lambda 2}$	1.0	0.938	0.062	6.200	0.265	0.311	1.174	0.0094
Overall mean value across parameters		0.025	4.158	0.122	0.130	1.060	0.0105

Chapter 6. Conclusions

Integrated choice and latent variable (ICLV) models are increasingly being considered in many fields as a means to gain a deeper understanding into the decision process of individuals as well as to potentially improve predictive ability. However, consideration and actual use of these models have been two different things in the literature. In particular, the use of ICLV models has been severely hampered by the difficulties encountered in usual maximum simulated likelihood (MSL) estimation as well as the amount of time to estimate these models (typically of the order of a day for one specification run, and even that for rather restrictive specifications). The reason for the estimation problems and computation time issues is that, in the traditional way of doing things, the integrand in ICLV models is itself a mixture of two probabilities (the probability of choice conditional on explanatory and latent variables, and the probability of the latent variable conditional on explanatory variables), which has to be integrated over the distribution of the latent variables conditional on explanatory variables.

In the current paper, we have proposed a different model formulation for the ICLV model, based on a multivariate probit (MNP) kernel for the choice component. To our knowledge, this is the first study to use a probit kernel within a general ICLV setting. As we show in the paper, combining this MNP-based choice model formulation with Bhat's maximum approximate composite marginal likelihood (MACML) inference approach immediately alleviates the specification and estimation challenges discussed above, and provides substantial computational time advantages. In particular, the dimensionality of integration in the log-likelihood function is independent of the number of latent variables, and we are able to specify quite general covariance structures (up to certain identification limits) in the error terms involved in the ICLV set-up. Further, our proposed approach easily accommodates ordinal indicators for the latent variables, as well as combinations of ordinal and continuous response indicators. The approach can be extended in a relatively straightforward fashion to also include nominal indicator variables.

The paper designed a simulation experiment in a virtual context of travel mode choice, and undertook a simulation exercise to evaluate the ability of the MACML approach to recover model parameters. The simulation results show that, irrespective of the sample size used in estimation (the sample sizes tested were 500, 1000, and 2000 observations), the MACML estimator recovers the parameters of the model remarkably well. The MACML estimator is also quite efficient in the overall, though the results indicate the need for relatively large sample sizes. This is needed to pin down the effects of the latent variables in the choice model. Additional investigation of efficiency consideration in the context of the proposed MACML estimation of ICLV models is an important direction for future research. For example, an extensive simulation study to compare the MACML estimation of probit-based ICLV models (based on Equation (15)) with the MSL estimation of probit-based ICLV models (based on Equation (16)) would be helpful, to examine the impact of different settings (such as the number of ordinal versus continuous indicators, the number of latent variables, general versus independent covariance matrix specifications for the latent variables and/or choice alternative utilities, the number of choice alternatives, and varying sample sizes) on not only efficiency considerations, but also estimation convergence and the accuracy of parameter recovery. But, for all sample

sizes, the asymptotic formula (based on the inverse of the Godambe information matrix) in our proposed MACML approach is performing well in estimating the finite sample standard errors. Also, the approximation error due to the use of a single permutation (per observation) in the analytic approximation for the MVNCD function evaluation is so small that it is a non-issue.

There were no convergence issues in our proposed approach even though we have not yet coded the Hessian of the analytically approximated CML function that is maximized (rather, we currently use a numerical procedure to obtain the Hessian, and then use the numerical Hessian and the analytic gradient in the computation of the Godambe covariance matrix of Equation 17). Additionally, the computational time is much lesser than for traditional logit-based kernels and associated MSL estimation procedures. Importantly, it is quite remarkable that this paper is the first study to use a probit choice kernel within a general ICLV setting. This is far more convenient for the usual applications for the ICLV models where the number of alternatives is few, and the number of latent variables can be many. Further, interactions of latent variables with individual demographic and other observed variables is easy to accommodate.

In closing, it is hoped that our new ICLV formulation and associated inference approach will unshackle researchers and practitioners from the constraints imposed by the traditional ICLV formulation, and open the door for the extensive use of "soft" psychometric measures (along with traditionally used "hard" covariates) in discrete choice modeling. Future research efforts should consider the performance of our inference approach with normal and non-normal random coefficients, interactions of latent variables with observed variables, panel settings with varying numbers of observations per individual, multiple indicators per latent variable, and increasing number of alternatives in the choice model. However, the results in this paper paint a very encouraging picture for the use of the MACML method for the quick, accurate, and practical estimation of ICLV models with flexible and rich stochastic specifications.

References

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

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.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H. (ed.) *In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*, Pergamon, Elsevier Science, 431-470.

Bhat, C.R., 1997. Endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science* 31(1), 34-48.

Bhat, C.R., 1998. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A* 32(7), 495-507.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B* 35(7), 677-693.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7), 923-939.

Bhat, C.R., 2014. The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends® in Econometrics*, forthcoming.

Bhat, C.R., Sidharthan, R., 2011. A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B* 45(7), 940-953.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

Bhat, C.R., Schofer, J.L., Koppelman, F.S., Bautch, R.C., 1993. Driver recruitability for advanced traveler information system experiments. *Transportation Research Part C* 1(4), 265-274.

Bolduc, D., Alvarez-Daziano, R.A., 2010. On estimation of hybrid choice models. In: Hess, S., Daly, A. (eds.) *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference*. Emerald Group Publishing, 259-287.

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

Cox, D.R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3), 729–737.

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.

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

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208-1211.

Johansson, M.V., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A* 40(6), 507-525.

Keane, M., 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10(2), 193-200.

Koppelman, F.S., Hauser, J.R., 1978. Destination choice behavior for non-grocery-shopping trips. *Transportation Research Record: Journal of the Transportation Research Board* 673, 157-165.

Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Lindsay, B.G., Yi, G.Y. and Sun, J., 2011. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* 21(1), 71-105.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4(1), 103-120.

Molenberghs, G., Verbeke, G., 2005. *Models for Discrete Longitudinal Data*. Springer Series in Statistics. Springer Science + Business Media, Inc., New York.

Munkin, M.K., Trivedi, P.K., 2008. Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* 143(2), 334-348.

Revelt, D., Train, K., 1996. Incentives for appliance efficiency in a competitive energy environment. ACEEE Summer Study Proceedings Paper, Panel 3, Paper 15, August.

Stapleton, D.C., 1978. Analyzing political participation data with a MIMIC Model. *Sociological Methodology* 9, 52-74.

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

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5-42.

Vij, A., Walker, J., 2014. Hybrid choice models: The identification problem. In: Hess, S., Daly, A. (eds.) *Handbook of Choice Modelling*, Edward Elgar Publishing, *forthcoming*.

Xu, X., Reid, N., 2011. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141(9), 3047-3054.

Yi, G.Y., Zeng, L., Cook, R.J., 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *The Canadian Journal of Statistics* 39(1), 34-51.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3), 335-356.