On estimation of Hybrid Choice Models

by

Denis Bolduc[1], Ricardo Alvarez-Daziano
Département d'économique
Université Laval.
Pavillon J.-A.-DeSève, QC G1V 0A6,
(418) 656-5427; Fax : (418) 656-2707
denis.bolduc@ecn.ulaval.ca ([1]corresponding author)
ricardo.alvarez-daziano.1@ulaval.ca

March 25, 2010

Abstract

Within the continuous search for flexible models capable of dealing with different practical and realistic situations, discrete choice modeling has developed especially quickly: the simple but restrictive Multinomial Logit model has evolved into the powerful Logit Mixture model. This search for flexibility has continuously faced the problem of a more involved and extremely demanding estimation process. On the other hand, in the last few years the flexibility search has been extended to the next level, and discrete choice modeling now aims to explicitly incorporate psychological factors affecting decision making, with the goal of enhancing the behavioral representation of the choice process. Hybrid Choice Models (HCM) expand on standard choice models by including attitudes, opinions and perceptions as psychometric latent variables, which become observable through a group of measurement relationships or indicators.

In this paper we describe the classical estimation techniques for a simulated maximum likelihood solution for a general Hybrid Choice Model, allowing for interactions among the latent variables and for different distributions for the indicator variables. To show that joint estimation of HCM is feasible, we apply the classical estimation techniques we developed to data of stated personal vehicle choices made by Canadian consumers when faced with technological innovations.

We then go beyond classical methods, and we introduce Hierarchical Bayes. Specifically, we present the details of the HCM Gibbs sampling sequence of full conditional posterior distributions for a simple case, considering both a Probit and a Logit Mixture discrete choice kernel. We then carry out a Monte Carlo experiment to test how the HCM Gibbs sampler works in practice. To our knowledge, this is the first practical application of HCM Bayesian estimation.

We show that although HCM joint estimation requires the evaluation of complex multidimensional integrals, simulated maximum likelihood can be successfully implemented and offers an unbiased, consistent and smooth estimator of the true probabilities. The HCM framework not only proves to be capable of introducing latent variables, but also makes it possible to tackle the problem of measurement errors in variables in a very natural way. Finally, we also show that working with Bayesian methods has the potential to break down the complexity of classical estimation.

1 Introduction

Within the continuous search for flexible models capable of dealing with different practical and realistic situations, discrete choice modeling has developed especially quickly: the simple but restrictive Multinomial Logit model has evolved into the powerful Logit Mixture model. In the last few years this flexibility search has been extended to the next level, and discrete choice modeling now aims to explicitly incorporate psychological factors affecting decision making, with the goal of enhancing the behavioral representation of the choice process.

Hybrid Choice Models (HCM) is a new generation of discrete choice models with improved explanatory power that integrate discrete choice and latent variables models taking into account the impact of *attitudes* and *perceptions* on the decision process. The use of HCM will permit us to adequately predict individual preferences and to assess the impact of unobserved factors involved in the behavioral representation of the choice process.

In this paper we describe HCM estimation techniques and we analyze the practical implementation of HCM in order to include perceptions and attitudes in a standard discrete choice setting. In section 2, we discuss how HCM integrate latent variables into a standard discrete choice setting. Section 3 describes the technical details regarding the maximum simulated likelihood implementation of a general Hybrid Choice Model with a Logit Mixture kernel for the discrete choice sub-model. We expand the method presented in Bolduc et al. (2005) to a more general case, through simultaneity among the latent variables and through incorporating not only continuous but also discrete indicator variables. In section 4, we present an example of the feasibility of HCM classical estimation using empirical data on private vehicle choice. We present the results of each partial model that configures the hybrid model setting. Section 5 introduces HCM Bayesian estimation. In this section we carry out a Monte Carlo experiment to test how the HCM Gibbs sampler works in practice. This section connects to the rest of the paper by adding to the discussion on estimation of HCM. While in section 3 and 4 we provide the elements of classical HCM estimation, in section 5 we introduce the idea that by using a Bayesian approach we could potentially break down the complexity of classical techniques. In section 6, we present the main conclusions of our work identifying guidelines for future research.

2 Hybrid Choice Modeling

2.1 Standard Discrete Choice Modeling

In discrete choice modeling, the most common approach is based on *random utility maximization* theory or RUM (McFadden, 1974), which introduces the concept of individual

choice behavior being intrinsically probabilistic. According to this theory, each individual has a utility function associated with each of the alternatives. This individual function can be divided into a *systematic part*, which considers the effect of the explanatory variables, and a random part that takes into account all the effects not included in the systematic part of the utility function. In other words, choices are modeled using a *structural equation* (1) – the utility function – representing the individual preferences, where the explanatory variables are the alternative attributes and individual characteristics. The observed choice corresponds to the alternative which maximizes the individual utility function, a process represented by a *measurement equation* (2). Because the utility function has a random nature, the output of the model actually corresponds to the choice probability of individual n choosing alternative i. The set of equations describing the standard discrete choice setting is given by:

$$U_{in} = X_{in}\beta + \upsilon_{in} \tag{1}$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where U_{in} corresponds to the utility of alternative i as perceived by individual n; X_{in} is a row vector of attributes of alternative i and socioeconomic characteristics of individual n; β is a column vector of unknown parameters; υ_{in} is an error term; y_{in} corresponds to an indicator of whether alternative i is chosen by individual n or not; and C_n is the individual set of available alternatives.

Different choice models can be derived depending on the assumptions considered for the distribution of the random error term (Ben-Akiva and Lerman, 1985). So far the workhorses in this area have been the Multinomial Logit model (McFadden, 1974) and the Nested Logit model (Ben-Akiva, 1973). Both offer closed forms for choice probabilities but rely on restrictive simplifying assumptions. To gain generality, more flexible models have been incorporated in practice. Indeed, one powerful modeling alternative is the Logit Mixture model or MMNL (Bolduc and Ben-Akiva, 1991 and Brownstone and Train, 1999), which can approximate any random utility maximization model (McFadden and Train, 2000). The main idea of this kind of model is to consider more than one random component, allowing for the presence of a more flexible covariance structure. The estimation implies the evaluation of integrals without a closed form solution, although it is possible to use computer aided simulation techniques to approximate it.

In this way it can be seen that the development of discrete choice modeling has evolved quickly and that powerful models can be used. However, under the standard random utility approach (McFadden, 1974) discrete choice models represent the decision process as an obscure black box, where attitudes, perceptions and knowledge are neglected. According to 2002 Nobel Laureate Daniel Kahneman, there still remains a significant difference between economist modelers that develop practical models of decision-making and behavioral scientists which focus on in-depth understanding of agent behavior. Both

have fundamental interests in behavior but each work with different assumptions and tools. McFadden (1986) points out the need to bridge these worlds by incorporating attitudes in choice models. In his 2000 Nobel lecture, McFadden emphasized the need to incorporate attitudinal constructs in conventional economics models of decision making.

2.2 Latent Variables and Discrete Choice: The Hybrid Choice Model

In the last few years the flexibility search we described in the previous section has been extended to the next level, and discrete choice modeling now aims to explicitly incorporate psychological factors affecting decision making (Ben-Akiva et al., 2002), with the goal of enhancing the behavioral representation of the choice process. The *Hybrid Choice Modeling* approach embraces this improved representation. In fact, Hybrid Choice Models (HCM) expand on standard choice models by considering the following extensions (Walker and Ben-Akiva, 2002): heterogeneity through flexible error structures (such as the use of a MMNL Kernel formulation), the combination of revealed (RP) and stated preference (SP) data, the presence of latent classes explaining market segments, and the integration of latent (unobserved) constructs according to an *Integrated Choice and Latent Variable* (ICLV) model. It is the ICLV model inside the HCM conceptual framework which permits the inclusion of attitudes, opinions and perceptions as psychometric latent variables in such a way that understanding of consumer behavior is improved while the model gains in predictive power. Since it is methodologically trivial to add in a latent class model and to consider mixed RP/SP data, our HCM analysis will focus on both the ICLV model and the consideration of flexible disturbances.

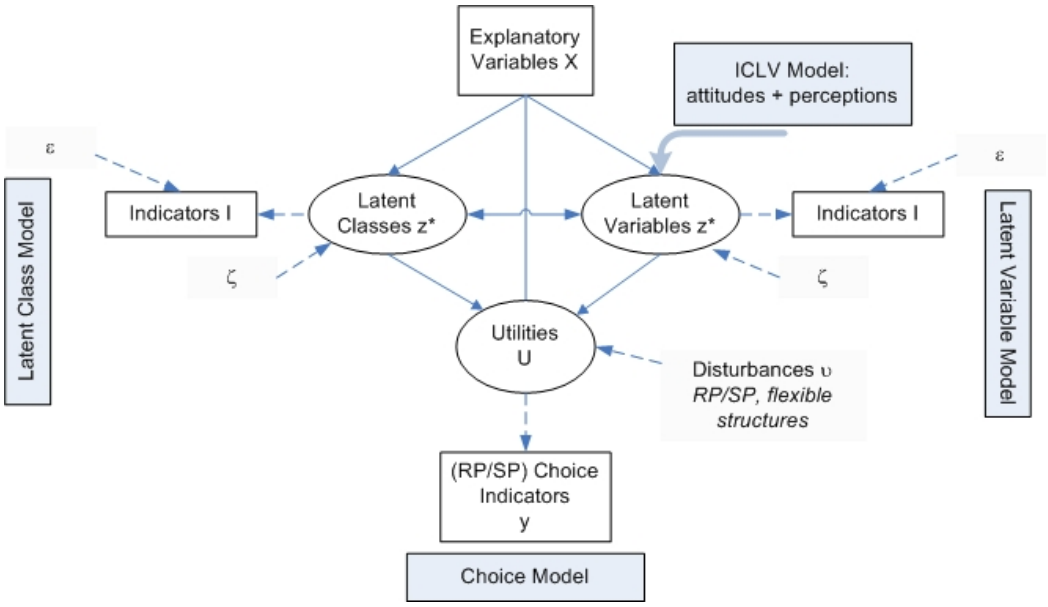

FIGURE 1 Hybrid Choice Model

The ICLV setting is an improved representation model of the choice process that involves dealing with a choice model formulation that contains unobserved psychometric variables (*perceptions* and *attitudes*) among the explanatory variables incorporated as *latent variables* (Bolduc at al., 2005). On the one hand, *perception variables* measure the cognitive individual capacity to represent and evaluate the attributes of different alternatives. Perceptions are relevant because the choice process depends on how attribute levels are perceived by the individual beliefs of a specific consumer. On the other hand, *attitude variables* measure the individual evaluation of importance assigned to the features of different alternatives. Attitudes are related to individual heterogeneity (*taste variations*) and reflect individual tastes, needs, values, goals, and capabilities that develop over time and are affected by experience and external factors, such as the socioeconomic characteristics of the decision-maker (Walker, 2002).

Econometrically, the latent variable model is composed of a set of *structural equations*, which describe the latent variables in terms of observable exogenous variables, and a group of measurement relationships (*measurement model*) linking latent variables to *indicators* (Jöreskog and Sörbom, 1984). Since the latent variables are unobserved, they are normally linked to answers to questions of a survey: the indicators. These indicators can be continuous, binary or categorical variables expressed by responses to attitudinal and perceptual survey questions. Note that under the RUM framework, the standard choice model is a latent model itself. The utility function is a latent construct that measures the individual level of satisfaction conditional on each alternative (*choice model structural equation*). While the utility function is unobservable, revealed or stated choice serves as an indicator of the underlying choice process.

The latent variable model can be integrated into the standard choice model setting, obtaining the HCM group of structural and measurement equations which may be written as follows:

<u>*Structural equations*</u>

$$z_n^* = \Pi z_n^* + Bw_n + \zeta_n = (I_L - \Pi)^{-1} Bw_n + (I_L - \Pi)^{-1}\zeta_n, \quad \zeta_n \sim N(0, \Psi) \tag{3}$$

$$U_n = X_n\beta + \Gamma z_n^* + \upsilon_n \tag{4}$$

<u>*Measurement equations*</u>

$$I_n = \alpha + \Lambda z_n^* + \varepsilon_n, \quad \varepsilon_n \sim N(0, \Theta) \tag{5}$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where z_n^* is a $(L \times 1)$ vector of latent variables; we introduce the $(L \times L)$ matrix Π allowing the eventual presence of simultaneity or interactions among the latent variables – we assume that $(I_L - \Pi)$ is invertible, where I_L represents the identity matrix of size L; w_n is a $(M \times 1)$ vector of explanatory variables affecting the latent variables;

B is a $(L \times M)$ matrix of unknown parameters used to describe the global effect of $(I_L - \Pi)^{-1} B w_n$ on the latent variables; and Ψ is a $(L \times L)$ variance covariance matrix which describes the relationship among the latent variables through the error term. The choice model in equation (4) is written in vector form where we assume that there are J alternatives. Therefore, U_n is a $(J \times 1)$ vector of utilities; υ_n is a $(J \times 1)$ vector of error terms associated with the utility terms. X_n is a $(J \times K)$ matrix with X_{in} designating the ith row. β is a $(K \times 1)$ vector of unknown parameters. Γ is a $(J \times L)$ matrix of unknown parameters associated with the latent variables present in the utility function, with Γ_i designating the ith row of matrix Γ.

In the set of measurement equations, I_n corresponds to a $(R \times 1)$ vector of indicators of latent variables associated with individual n; α is a $(R \times 1)$ vector of constants and Λ is a $(R \times L)$ matrix of unknown parameters that relate the latent variables to the indicators. The term ε_n is a $(R \times 1)$ vector of independent error terms. This implies that Θ is a diagonal matrix with variance terms on the diagonal. Finally, we stack the choice indicators y_{in}'s into a $(J \times 1)$ vector called y_n.

If the latent variables were not present, the choice probability of individual n selecting alternative i would correspond exactly to the standard choice probability $P(y_{in} = 1 \,|X_n, \beta) \equiv P_n(i \,|X_n, \beta)$. In a setting with given values for the latent variables z_n^*, the choice probability would be represented by $P_n(i \,|z_n^*, X_n, \theta)$ where θ contains all the unknown parameters in the choice model of equation (4). Since latent variables are not actually observed, the choice probability is obtained by integrating the latter expression over the whole space of z_n^*:

$$P_n(i|X_n, w_n, \theta, B, \Pi, \Psi) = \int_{z_n^*} P_n(i \,|z_n^*, X_n, \theta) g(z_n^*|w_n, B, \Pi, \Psi) dz_n^*, \tag{7}$$

which is an integral of dimension equal to the number of latent variables in z_n^* and where $g(z_n^*|w_n, B, \Pi, \Psi)$ is the density of z_n^* defined in equation (3).

Indicators are introduced in order to characterize the unobserved latent variables, and econometrically they permit identification of the parameters of the latent variables. Indicators also provide efficiency in estimating the choice model with latent variables, because they add information content. The variables y_n and I_n are assumed to be correlated only via the presence of the latent variables z_n^* in equations (4) and (5). Given our assumptions, the joint probability $P(y_{in} = 1, I_n) \equiv P_n(i, I)$ of observing y_n and I_n may thus be written as:

$$P_n(i, I|X_n, w_n, \delta) = \int_{z_n^*} P_n(i \,|z_n^*, X_n, \theta) f(I_n|z_n^*, \Lambda, \Theta) g(z_n^* \,|w_n, B, \Pi, \Psi) dz_n^*, \tag{8}$$

where $f(I_n|z_n^*, \Lambda, \Theta)$ is the density of I_n defined in equation (5). The term δ designates the full set of parameters to estimate jointly the discrete choice and the latent variable models (i.e. $\delta = \{\theta, B, \Pi, \Psi, \Lambda, \Theta\}$).

Few applications of the hybrid choice model are found in the literature, usually focusing on one latent variable – allowing numerical integration – or using a consistent but not efficient two-stage estimation procedure. While Ben-Akiva and Boccara (1987) develop the idea of hybrid modeling in the context of a more comprehensive travel behavior framework, Walker and Ben-Akiva (2002) and Morikawa, Ben-Akiva and McFadden (2002) extend this development, showing how the model can be estimated and implemented in practice. The contribution of Bolduc, Ben-Akiva, Walker and Michaud (2005) is the first example of the analysis and implementation of a situation characterized by a large number of latent variables and a large number of choices. The work of Ben-Akiva, Bolduc and Park (2008) is a recent application of the Hybrid Choice setting applied to the freight sector. For a personal-vehicle-technology choice context, Bolduc, Boucher and Alvarez-Daziano (2008) analyze the practical use of a large number of indicators.

3 Hybrid Choice Models: Classical Estimation

This section provides the analytical details regarding the maximum simulated likelihood implementation of a general Hybrid Choice Model (HCM) with a Logit Mixture kernel for the discrete choice sub-model. To gain generality and flexibility, we expand the method presented in Bolduc et al. (2005) in two relevant ways. In our model we allow for both the presence of simultaneity among the latent variables (through equation 3) and the incorporation of latent variables with associated indicators that can be not only continuous but also discrete (binary or multinomial).

3.1 Evaluating the joint choice probability

For efficiency reasons, we only focus on a full information solution. Following Walker and Ben-Akiva (2002), HCM classical full information estimation requires the evaluation of the joint probability $P_n(i, I|X_n, w_n, \delta)$ defined in equation (8). This joint probability depends, first, on the discrete choice kernel $P_n(i\,|z_n^*, X_n, \theta)$. In addition, the analytical form of the discrete choice kernel depends on the assumptions regarding the distribution of the random term υ_n defined in equation (4).

Indeed, if υ_n is i.i.d. extreme value type 1 distributed, then conditional on z_n^* the probability of choosing alternative i has the Multinomial Logit (MNL) form, which leads to the following expression:

$$P_n(i, I|X_n, w_n, \delta) = \int_{z^*} \frac{\exp(X_{in}\beta + C_i z_n^*)}{\sum_{j \in C_n} \exp(X_{jn}\beta + C_j z_n^*)} f(I_n|z_n^*, \Lambda, \Theta) g(z_n^*|w_n, B, \Pi, \Psi) dz_n^*. \quad (9)$$

Assuming a Multinomial Logit kernel provides an easier calculation of $P_n(i, I|X_n, w_n, \delta)$ because the choice probability $P_n(i\,|z_n^*, X_n, \theta)$ has a closed form. However, the same

modeling disadvantages found in the standard case still obtain. MNL assumes a restricted covariance structure, with no correlation and no heteroscedasticity.

We can derive a Probit kernel if we make the assumption that the error terms υ_n are multivariate Normal distributed. The Probit kernel solves the problem of restrictive simplifying assumptions of MNL. However, in the Probit case the choice probability no longer has a closed form. In fact, Probit classical estimation has proven to be burdensome in practice.

For classical estimation, a Logit Mixture kernel is the most convenient assumption to model flexible error structures. We will decompose υ_n assuming a Normal distributed factor analytic structure:

$$\upsilon_n = PT\xi_n + \nu_n, \tag{10}$$

where P is a $(J \times F)$ matrix of factor loadings; T is a $(F \times F)$ diagonal matrix that contains factor specific standard deviations($T \in \theta$); ξ is a $(F \times 1)$ of i.i.d. normally distributed factors; and ν is a $(J \times 1)$ vector of independent and identically distributed extreme value type 1 error terms. The logit mixture kernel adds an additional F-dimensional integral to the joint probability $P_n(i, I|X_n, w_n, \delta)$, which now implies solving:

$$P_n(i, I|X_n, w_n, \delta) = \int_{\xi_n} \int_{z_n^*} P_n(i\,|z_n^*, X_n, \theta, \xi_n) f(I_n|z_n^*, \Lambda, \Theta) g(z_n^*\,|w_n, B, \Pi, \Psi) N_\xi(0, I_F) dz_n^* d\xi_n. \tag{11}$$

Since ν is i.i.d. extreme value type 1, note that $P_n(i\,|z_n^*, X_n, \theta, \xi_n)$ has the following MNL form:

$$P_n(i\,|z_n^*, X_n, \theta, \xi_n) = \frac{\exp(X_{in}\beta + C_i z_n^* + P_i T\xi_n)}{\sum_{j \in C_n} \exp(X_{jn}\beta + C_j z_n^* + P_j T\xi_n)}, \tag{12}$$

where P_i denotes row i of P. Assuming that z_n^* and ξ_n are mutually independent, equation (12) can be incorporated directly into equation (11).

Regarding the measurement model and its distribution $f(I_n|z_n^*, \Lambda, \Theta)$, we assume that each equation that links the indicators and the latent variables corresponds to a continuous, a binary, or a multinomial ordered response. A measurement equation r in the continuous case is given by $I_{rn} = I_{rn}^*$ with:

$$I_{rn}^* = \alpha_r + \Lambda_r z_n^* + \varepsilon_{rn}, \quad \varepsilon_{rn} \sim N(0, \theta_r^2). \tag{13}$$

In the binary case, we rather get instead:

$$I_{rn} = \begin{cases} 1 & \text{if } I_{rn}^* \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

while in the multinomial ordered case with Q responses, we obtain:

$$I_{rn} = \begin{cases} 1 & \text{if } \gamma_0 < I^*_{rn} \leq \gamma_1 \\ 2 & \text{if } \gamma_1 < I^*_{rn} \leq \gamma_2 \\ \vdots & \\ Q & \text{if } \gamma_{Q-1} < I^*_{rn} \leq \gamma_Q, \end{cases} \tag{15}$$

where I_{rn} and ε_{rn} are the rth element of I_n and ε_n respectively. θ_r^2 is the rth element on the diagonal of Θ, and Λ_r denotes row r of Λ. In the multinomial cases, the γ_q's are estimated. By convention, γ_0 and γ_Q are fixed to values that represent $-\infty$ and ∞ respectively. We assume that Θ is diagonal, which implies that the indicators are not cross-correlated.

Given our assumptions, the density $f(I_n|z^*_n, \Lambda, \Theta)$ that we denote as $f(I_n)$ to simplify, corresponds to:

$$f(I_n) = \prod_{r=1}^{R} f(I_{rn}). \tag{16}$$

According to the assumptions of equation (13), if measurement equation r is continuous, then

$$f(I_{rn}) = \frac{1}{\theta_r}\phi\left(\frac{I_{rn} - \alpha_r - \Lambda_r z^*_n}{\theta_r}\right), \tag{17}$$

where ϕ denotes the probability density function (pdf) of a standard normal. If the measurement equation r corresponds to a binary response, then

$$f(I_{rn}) = \Phi\left(\frac{\alpha_r + \Lambda_r z^*_n}{\theta_r}\right)^{I_{rn}} \left(1 - \Phi\left(\frac{\alpha_r + \Lambda_r z^*_n}{\theta_r}\right)\right)^{(1-I_{rn})}, \tag{18}$$

where Φ denotes the cumulative distribution function (cdf) of a standard normal. Finally, if measurement equation r corresponds to a multinomial ordered response, then

$$f(I_{rn} = q) = \Phi\left(\frac{\gamma_q - \Lambda_r z^*_n}{\theta_r}\right) - \Phi\left(\frac{\gamma_{q-1} - \Lambda_r z^*_n}{\theta_r}\right) \quad . \tag{19}$$

Additionally, $g(z^*_n|w_n, B, \Pi, \Psi)$ corresponds simply to the multivariate normal distribution $MVN((I_L - \Pi)^{-1}Bw_n, [(I_L - \Pi)^{-1}]\Psi[(I_L - \Pi)^{-1}]')$.

3.2 HCM simulated maximum likelihood solution

Now that we have described each component of the joint probability shown in equation (8), we can write the likelihood equation as:

$$l(\delta) = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i, I|X_n, w_n, \delta)^{y_{in}}, \tag{20}$$

which leads to the following maximum log-likelihood problem:

$$\max_{\delta} l(\delta) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i, I|X_n, w_n, \delta). \tag{21}$$

The evaluation of the joint probability $P_n(i, I|X_n, w_n, \delta)$ is required to find the solution of the problem (21) $\hat{\delta} = argmax\{l(\delta)\}$. The number of latent variables has an impact on the computation of this probability. In fact, for the case of a Logit Mixture kernel, note that equation (11) implies the computation of an integral of dimension $F + L$. In a moderate size model with say $F = 5$ factors and $L = 4$ latent variables, this integral is of dimension 9. Clearly, simulation would be required. In practice, with a large number of latent variables (more than 3), we replace the multidimensional integral with a smooth simulator which has good properties.

Taking advantage of the expectation form of equation (11), we can replace the probability with the following empirical mean:

$$\tilde{P}_n(i, I\,|X_n, w_n, \delta\,) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp(X_{in}\beta + C_i z_n^{*s} + P_i T \xi_n^s)}{\sum\limits_{j \in C_n} \exp(X_{jn}\beta + C_j z_n^{*s} + P_j T \xi_n^s)} f(I_n\,|z_n^{*s}, \Lambda, \Theta), \tag{22}$$

where z_n^{*s} corresponds to a random draw s from the $g(z_n^*|w_n, B, \Pi, \Psi)$ distribution, and ξ_n^s is a random draw s taken over the distribution of ξ. This sum is computed over S draws.

This simulator is known to be unbiased, consistent (as $S \to \infty$) and smooth with respect to the unknown parameters. Replacing $P_n(i, I|X_n, w_n, \delta)$ with $\tilde{P}_n(i, I|X_n, w_n, \delta)$ in the log likelihood leads to a maximum simulated likelihood (MSL) solution. We therefore consider the following objective function – often called the sample average approximation (SAA): $\sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln \tilde{P}_n(i, I\,|X_n, w_n, \delta)$.

In the past few years, a lot of progress has been made regarding MSL estimation. Train (2003) gives an in-depth analysis of the properties of MSL estimators. Recent results, based mainly on the analysis of Logit Mixture models (MMNL) and mostly attributable to Bhat (2001), suggest the use of *Halton* draws. Halton-type sequences are known to produce simulators with a given level of accuracy using fewer draws than when using conventional uniform random draws (Ben-Akiva et al., 2002; Munizaga and Alvarez-Daziano, 2005). Currently, our HCM estimation software (Bolduc and Giroux, 2005) makes use of both Halton sequences and standard pseudo-random numbers.

Simulated maximum likelihood is now well known and has been applied in numerous circumstances. The logit probability kernel present in equation (11) makes the simulated log likelihood fairly well behaved. Asymptotically, meaning as $S \to \infty$ and as $N \to \infty$, the solution becomes identical to a solution arising from maximizing the actual log like-

lihood function: $\sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i, I \,|X_n, w_n, \delta)$.

3.3 Identification discussion

While identification issues are now well understood in the context of traditional discrete choice models (Ben Akiva and Lerman, 1985; Train, 2003), general necessary and sufficient conditions for identification of Hybrid Choice Models have not yet been developed. A sufficient but not necessary technique for HCM identification is a two-step approach, where we apply separate conditional identification rules for the choice model and the latent variable model (Walker and Ben-Akiva, 2002).

On the one hand, in discrete choice models what matters are differences between utilities and not the level of the utilities itself. Therefore and as a general framework, we have an order condition that establishes a limit for the total number of nuisance parameters that can be estimated. This boundary, which is a necessary condition for identification, is equal to the number of potentially different cells in the deviated covariance matrix. The next step is to examine the rank condition, which is more restrictive than the order condition and which is a sufficient condition for identification. This condition states that the number of nuisance parameters that can be estimated is generally equal to the rank of the Jacobian matrix of the vector that contains the different elements of the deviated covariance matrix, minus one term which sets scale. Dansie (1985), Bunch (1991) and Bolduc (1992) study the identification conditions for the case of Multinomial Probit models, which is equivalent to the case of the Logit Mixture model, discussed specifically in Walker (2002) and Ben-Akiva, Bolduc and Walker (2001).

On the other hand, for identification of the parameters of the measurement equation of the latent variable model (equation 5) the constant terms α_r must be set to 0 in the non continuous cases. Additionally –except for the continuous case– the variances θ_r cannot be estimated. In that case they need to be fixed to 1.

4 Classical estimation: real data application

In this section we study empirically the application of the HCM classical estimation method developed in the previous section. This method allows us to estimate jointly the parameters of a discrete choice model and a latent variable model. Using real data about virtual personal vehicle choices made by Canadian consumers, we want to verify that Hybrid Choice is genuinely capable of adapting to practical situations.

4.1 Personal vehicle choice data

We use data from a survey conducted by the EMRG (Energy and Materials Research Group, Simon Fraser University) of virtual personal vehicle choices made by Canadian consumers when faced with technological innovations. Horne (2003) provides all the details regarding the survey, including the design of the questionnaire, the process of conducting the survey, and analysis of the collected data.

Survey participants were first contacted in a telephone interview for personalizing a detailed questionnaire that was then mailed to them. The mailed questionnaire had 5 different parts: *Part 1:* Transportation options, requirements and habits; *Part 2:* Virtual personal vehicle choices made by Canadian consumers when faced with technological innovations (stated preferences experiment); *Part 3:* Transportation mode preferences; *Part 4:* Views on transportation issues; and *Part 5:* Additional information (gender, education, income).

The Stated Preferences (SP) hypothetical personal vehicle choices in Part 2 of the questionnaire considered four vehicle types: *standard gasoline vehicle* (SGV, operating on gasoline or diesel), *alternative fuel vehicle* (AFV, natural-gas vehicle), *hybrid vehicle* (HEV, gasoline-electric), and *hydrogen fuel cell vehicle* (HFC). For each one of these alternative vehicle types, the attributes were defined as: *capital cost* (purchase price), *operating costs* (fuel costs), *fuel available* (percentage of stations selling the proper fuel), *express lane access* (whether or not the vehicle would be granted express lane access), *emissions data* (emissions compared to a standard gasoline vehicle), *power* (power compared to current personal vehicle).

Each participant was asked to make up to four consecutive virtual choices while the vehicle attribute values were modified after each round according to the SP experimental design. The sample has 866 completed surveys (of the total of 1150 individuals, 75% response rate), where each respondent provided up to 4 personal vehicle choices. After a clean up where we kept the individuals who answered the whole attitudinal rating exercise, there remains 1877 usable observations for HCM estimation. The SP design is described in Horne et al. (2005), including the various values assumed by the characteristics of the vehicles that were used as a basis for developing the experimental design.

The first step to build a Hybrid Choice model is to set the latent variables involved. We conduct our analysis focusing on three different relevant questions of the survey:

Transport Policies Support (TPS): Evaluation of 8 different policies (see Table 1 for the definition of each policy) - or government actions that would influence the transportation system - according to degree of support: 5 levels from Strongly Opposed to Strongly Supportive.

Transport Problems Evaluation (TPE): Evaluation of 6 different related to transportation problems (see Table 1 for the definition of each problem) according to degree of seriousness of problem: 5 levels from Not a Problem to Major Problem.

Car Attributes Importance (CAI): Evaluation of 7 different factors or car attributes (see Table 1 for the definition of each attribute) that influence the family's decision to purchase the current personal vehicle. Evaluation made according to degree of importance: 5 levels from Not at All Important to Very Important.

Considering the answers to these questions as indicators we identify two latent variables:

Environmental Concern (EC): related to transportation and its environmental impact.

Appreciation of new car features (ACF): related to car purchase decisions and how important are the characteristics of this new alternative.

The full list of indicators for each question (Transport Policies Support TPS, Transport Problems Evaluation TPE and Car Attributes Importance CAI) and their relations with the transport related latent variables *environmental concern* EC and *appreciation of new car features* ACF is provided in the following table:

	Indicator	Description – evaluated issue	EC	ACF
Transport Policies Support (TPS)	TPS 1	Expanding & Upgrading Roads	✓	
	TPS 2	Road Tolls & Gas Taxes	✓	✓
	TPS 3	Bike Lanes & Speed Controls	✓	
	TPS 4	Emissions Testing & Standards	✓	✓
	TPS 5	HOV & Transit Priorities	✓	
	TPS 6	Improving Transit Service	✓	
	TPS 7	Promoting Compact Communities	✓	
	TPS 8	Encouraging Short Work Weeks	✓	
Transport Problems Evaluation (TPE)	TPE 1	Traffic Congestion	✓	
	TPE 2	Traffic Noise	✓	
	TPE 3	Poor Local Air Quality	✓	✓
	TPE 4	Accidents caused by bad drivers	✓	
	TPE 5	High greenhouse gas emissions	✓	
	TPE 6	Speeding drivers in neighborhoods	✓	
Car Attributes Importance (CAI)	CAI 1	Purchase Price		✓
	CAI 2	Fuel Economy		✓
	CAI 3	Horsepower		✓
	CAI 4	Safety		✓
	CAI 5	Seating Capacity		✓
	CAI 6	Reliability		✓
	CAI 7	Styling		✓

TABLE 1 Indicators and Latent Variables

Our model also includes a third latent variable, the ***latent income variable*** (REV), to account for the measurement error problem in quantifying the income variable. Note that modeling measurement errors as latent variables as well as other practical situations

such as self-selection are cases where the application of the HCM framework naturally fits to solve the related problems (bias) that eventually arise.

The hybrid model setting that we consider is represented in Figure 2, where the complete set of structural and measurement equations is sketched, depicting the relationships between explanatory variables and each partial model. Indeed, we can distinguish the choice model, which is centered on the utility function modeling; the latent variables structural model, which links the latent variables with the characteristics of the traveler and the latent variables measurement model, which links each latent variable with the indicators.

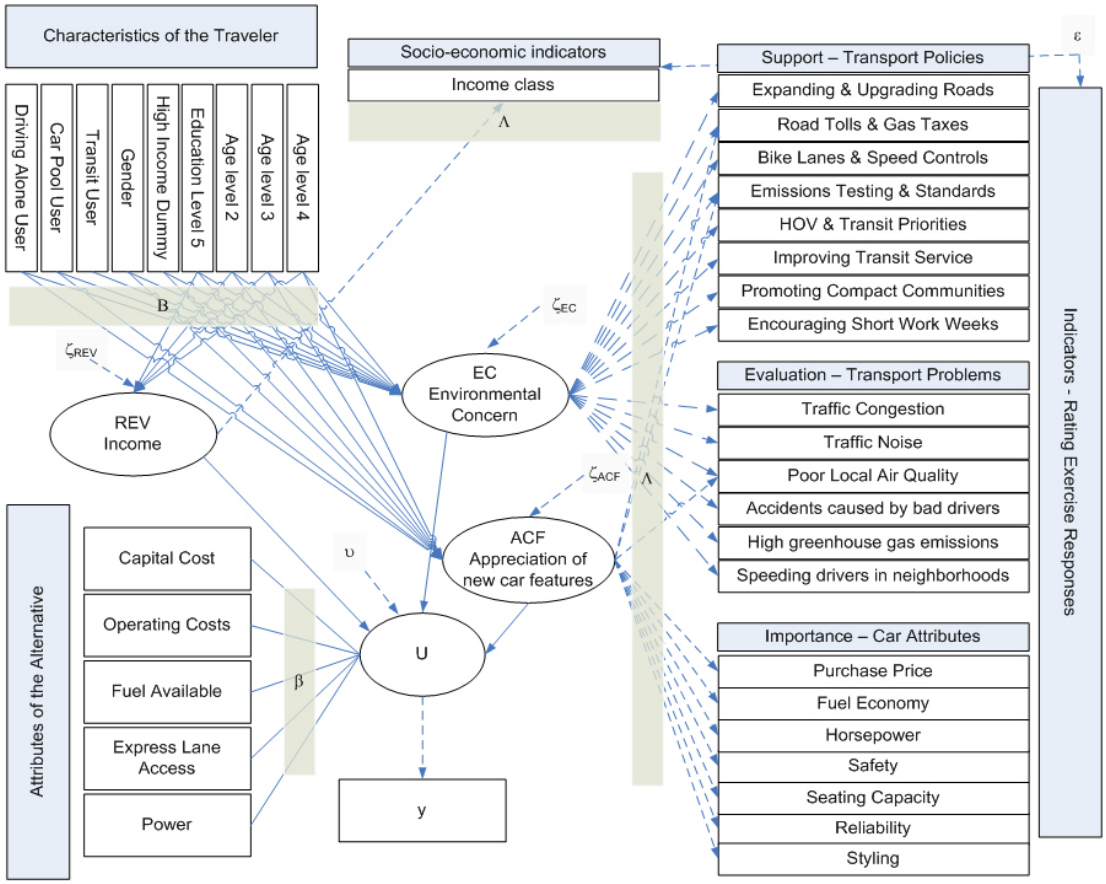

FIGURE 2 Hybrid Choice Model.

We will now present the results of the HCM classical estimation process. We implemented the SML solution presented in Section 3.2 in a custom-coded software written in Fortran (Bolduc and Giroux, 2005). Although the estimation process implies that all the equations are calibrated simultaneously, we will present the results separately for each HCM sub-model, i.e. the car choice model, the latent variable structural model and the latent variable measurement model.

4.2 Car Choice Model

The set of equations for the mode choice model alone are given by:

$$
\begin{aligned}
U_{SGVn} &= V_{SGVn} + \Gamma_{1,2}\mathrm{ACF}_n + \upsilon_{SGVn} && (23)\\
U_{AFVn} &= V_{AFVn} + \Gamma_{2,1}\mathrm{EC}_n + \Gamma_{2,2}\mathrm{ACF}_n + \upsilon_{AFVn} && (24)\\
U_{HEVn} &= V_{HEVn} + \Gamma_{3,1}\mathrm{EC}_n + \Gamma_{3,2}\mathrm{ACF}_n + \upsilon_{HEVn} && (25)\\
U_{HFCn} &= V_{HFCn} + \Gamma_{4,1}\mathrm{EC}_n + \Gamma_{4,2}\mathrm{ACF}_n + \Gamma_{4,3}\mathrm{REV}_n + \upsilon_{HFCn}, && (26)
\end{aligned}
$$

where $V_{in} = X_{in}\beta$ denotes the deterministic part of the utility expression for alternative i and individual n.

	Hybrid Choice Model		**MNL**	
	estimates	**t-stat**	**estimates**	**t-stat**
ASC_AFV	-6.626	-5.162	-4.500	-6.81
ASC_HEV	-4.383	-5.688	-1.380	-2.18
ASC_HFC	-6.403	-8.750	-2.100	-3.26
Capital Cost	-0.943	-4.369	-0.856	-4.07
Operating Cost	-0.849	-3.917	-0.826	-4.18
Fuel available	1.384	7.096	1.360	7.32
Expess lane access	0.162	2.229	0.156	2.29
Power	2.710	3.985	2.700	4.12
Latent Variables				
ACF on SGV	3.160	25.984	-	-
EC on AFV	0.798	2.695	-	-
ACF on AFV	2.810	30.708	-	-
EC on HEV	0.770	3.965	-	-
ACF on HEV	2.810	30.708	-	-
EC on HFC	1.085	5.620	-	-
ACF on HFC	3.054	31.048	-	-
REV on HFC	0.456	3.373	-	-
Number of individuals	1877		1877	
HCM Final global function	-57624.26		-	
MNL log likelihood	-		-1984.55	
Choice Model adj. rho-square	0.236		0.234	
Number of Halton draws	500		-	

TABLE 2 Car Choice Model Results

The deterministic utility contains the experimental attributes capital cost, operating costs, fuel available, express lane access, power as well as alternative specific constants for the *alternative fuel vehicle* AFV, the *hybrid vehicle* HEV, and the *hydrogen fuel cell*

vehicle HFC. The utility specification also contains the effect of the latent variables. The latent variable related to *environmental concern* (EC) was not considered on the *standard gasoline vehicle* SGV. On the basis of several attempts, the *latent income variable* (REV) was included only in the *hydrogen fuel cell vehicle* (HFC) alternative.

Common parameters with the standard Multinomial Logit model have the same sign and magnitude, except for alternative specific constants. Although the rho-square of the Hybrid model is only slightly better than the MNL rho-square, the significance of the latent variable parameters shows a relevant effect on the individual utilities. The latent variables all enter very significantly and positively into the choice model specification. Environmental concern (EC) encourages the choice of green technologies. In fact, EC has the highest effect on the *hydrogen fuel cell vehicle* HFC, followed by the *alternative fuel vehicle* AFV, and then the *hybrid vehicle* HEV. Note that *hydrogen fuel cell vehicle* HFC represents the cleanest engine technology. At the same time, all vehicles show a positive effect of the appreciation of the car features (ACF) on the choice probabilities.

4.3 Structural Model

For each one of the three latent variables (environmental concern EC, appreciation of new car features ACF, and income REV), we assume a linear structural regression equation whose estimation results are shown in Table 3:

	EC		**ACF**		**REV**	
	est	**t-stat**	**est**	**t-stat**	**est**	**t-stat**
Intercept	2.434	9.660	-3.094	-18.495	1.293	7.374
Driving Alone User	-0.020	-0.428	-0.118	-2.244	-	-
Car Pool User	0.097	1.135	0.100	0.938	-	-
Transit User	0.204	2.572	-	-	-	-
Female Dummy	0.258	7.392	0.283	6.474	-	-
High Income Dummy (>80K$)	-0.011	-0.294	-0.001	-0.017	-	-
Education: University	0.064	1.712	0.008	0.170	0.598	9.522
Age level: 26-40 years	0.187	2.279	0.328	3.961	0.592	5.363
Age level: 41-55 years	0.262	3.105	0.621	6.657	0.889	7.952
Age level: 56 years & more	0.332	3.702	0.525	5.278	0.228	1.840

TABLE 3 Structural Model Results

The structural equation links consumer characteristics with the latent variables. For example, we can conclude that *environmental concern* (EC) is more important for public transportation users than for car pool users. We in fact observe a negative parameter for a driving alone user. We also find that women are more worried about environmental issues than men.

High education level has a significant positive effect on the *latent income variable* (REV) while the effect is positive but not significant on both variables *environmental concern* and *appreciation of new car features*. In addition, the effect of age on *environmental concern* increases as the individuals get older. The effect of age on *appreciation of new car features* (ACF) is the highest for people between 41 to 55 years old. Not surprisingly, the effect of age on the *latent income variable* REV is small and not significant for people older than 55 years of age (an effect of people being retired).

Note that we also estimate the elements of the covariance matrix. As the results show (see Table 4), the elements in the diagonal of the Ψ matrix are significantly different from 0 and show the presence of heteroscedasticity. In this version, we assumed that the latent variables are uncorrelated.

	estimates	**t-stat**
Var(EC)	0.553	20.934
Var(ACF)	0.708	23.410
Var(REV)	0.449	3.336

TABLE 4 Latent Variables Variances

4.4 Measurement Model

Finally, the latent variable measurement model links the latent variables with the indicators, and a typical equation for this model has the form:

$$TPS_2 = \alpha_{TPS_2} + \lambda_{EC,TPS_2}\mathrm{EC}_n + \lambda_{AF,TPS_2}\mathrm{ACF}_n + \varepsilon_{TPS_2}. \tag{27}$$

In this example, we can see that the effects on the second indicator related to the Transport Policies Support (TPS) question are measured using a constant and the latent variables *environmental concern* EC and *appreciation of new car features* ACF. We have considered 21 indicators, so it is necessary to specify 21 equations. Their relation with the latent variables is depicted in Figure 2 and the results are displayed in Table 5.

As explained before, this model measures the effect of the latent variables on each indicator. Some interesting conclusions can be seen from the estimated parameters. For example, the effect of *environmental concern* EC on the indicator related to the support of expanding and upgrading roads is negative. This sign reflects the idea that car priority in the context of rising road capacity is negatively perceived because of the negative impact on the environment.

On the one hand, we see that the effect of *environmental concern* EC on the indicator related to the support of applying road tolls and gas taxes is positive, indicating a perceived positive environmental impact of measures allowing a rational use of the car.

	estimates	t-stat
Transport Policies Support		
EC on Expanding & Upgrading Roads	-0.392	-5.405
EC on Road Tolls & Gas Taxes	0.581	6.600
ACF on Road Tolls & Gas Taxes	-0.091	-1.389
EC on Bike Lanes & Speed Controls	0.532	8.507
EC on Reducing Car Emissions	0.478	7.944
ACF on Reducing Car Emissions	0.295	7.856
EC on High Occupancy Vehicles & Transit Priorities	0.606	8.143
EC on Improving Transit Service	0.491	8.352
EC on Promoting Compact Communities	0.206	2.994
EC on Encouraging Short Work Weeks	0.396	7.159
Transport Problems Evaluation		
EC on Traffic Congestion	0.735	9.154
EC on Traffic Noise	0.901	9.495
EC on Poor Local Air Quality	1.000	-
ACF on Poor Local Air Quality	-0.061	-1.416
EC on Accidents Caused by Bad Drivers	0.837	14.472
EC on Emissions & Global Warming	1.113	16.200
EC on Speeding Drivers in Neighborhoods	1.107	15.790
Car Attributes Importance		
ACF on Purchase Price	-0.004	-0.087
ACF on Fuel Economy Importance	0.259	5.008
ACF on Horsepower Importance	0.433	7.220
ACF on Safety Importance	1.000	-
ACF on Seating Capacity Importance	0.684	11.563
ACF on Reliability Importance	0.537	16.241
ACF on Styling	0.371	6.660
Income Class		
REV on rev	1.00	-

TABLE 5 Structural Model Covariates

Note also that the effect on the same indicator of the *appreciation of new car features* ACF, the other transport related latent variable considered, is negative although not significant. This sign can be explained because of the perceived negative impact of this kind of car use restrictions, especially if the user is considering to buy a new car. A similar analysis can be done for the other indicators. For example, the positive sign of the effect of both EC and ACF on the support for reducing vehicle emissions with regular testing and manufacturer emissions standards: it is perceived with a positive environmental impact but also it is positively perceived by consumers as a good attribute of a potential new car. It is worth noting that such results permit us to establish a consumer profile in a way not possible with standard discrete choice models.

In sum, using real data about virtual personal vehicle choices we showed that HCM is genuinely capable of adapting to practical situations. Although we obtained similar effects for the alternative attributes in the car choice model for both the Hybrid and MNL models, and the gain in fit of HCM is only marginally better than MNL, nevertheless, HCM offers an enormous improvement in modeling behavior. For Hybrid models, the choice model is only a part of the whole behavioral process. HCM combines the direct effect of underlying latent variables on the choice probabilities, with the socio-demographic characteristics that now enter the choice probabilities through the latent variables. HCM also takes into account opinions and attitudes through the consumer's response to attitudinal rating exercises. These responses are taken as indicators of the latent variables. Using simple questions we can enrich the model and obtain better knowledge about the user's characteristics and his behavioral attitudes and perceptions.

5 Bayesian HCM estimation: analysis of a simple case

5.1 Introducing HCM Bayesian estimation

Including numerous attitudes and perceptions in HCM with large sets of potentially interrelated choices directly entails the simulation of high dimensional integrals. We can address this problem using classical methods, which use an efficient choice probability simulator through maximum simulated likelihood (MSL) estimation, the technical details of which are described in section 3. Although feasible in practice (as shown in section 4), the MSL approach necessary for classical HCM estimation is very demanding in situations with a huge choice set of interdependent alternatives with a large number of latent variables.

For these reasons, we propose to go beyond classical methods by introducing Hierarchical Bayes techniques. Building on the rapid development of Markov Chain Monte Carlo (MCMC) techniques, and on the idea that Bayesian tools (with appropriate priors) could be used to produce estimators that are asymptotically equivalent to those obtained using classical methods, we define the goal of both theoretically and empirically implementing

a Bayesian approach to Hybrid Choice modeling. In a first attempt to achieve this goal we develop and implement the HCM sequence of full conditional posterior distributions for a simple case: a trinomial Probit for the choice model with 1 latent variable, 1 continuous indicator and known variances. We will justify the assumption of a Probit kernel by showing that for this kernel the derivation of the HCM Bayesian estimator is straightforward.

5.2 Choice Model

Let us consider a trinomial choice for $n = 1..., N$ individuals, where three alternatives $i = 1, 2, 3\ \forall n$ are described by two different attributes X_1 and X_2 – plus two alternative specific constants, and one latent variable z^* which we will discuss later. The error terms υ are assumed to be independently Normal distributed.

Structural equation

For each individual n, we thus have

$$\begin{aligned} U_{1n} &= \beta_1 X_{11n} + \beta_2 X_{12n} + \upsilon_{1n} && (28)\\ U_{2n} &= ASC_2 + \beta_1 X_{21n} + \beta_2 X_{22n} + \Gamma_2 z_n^* + \upsilon_{2n} && (29)\\ U_{3n} &= ASC_3 + \beta_1 X_{31n} + \beta_2 X_{32n} + \Gamma_3 z_n^* + \upsilon_{3n}. && (30) \end{aligned}$$

In discrete choice models, decisions are based on utility differences so that we can consider a deviated model that, in this particular case, leads us to write the structural equation in two equations:

$$\begin{aligned} U_{2n} - U_{1n} &= ASC_2 + \beta_1(X_{21n} - X_{11n}) + \beta_2(X_{22n} - X_{12n}) + \Gamma_2 z_n^* + \upsilon_{2n} - \upsilon_{1n} && (31)\\ U_{3n} - U_{1n} &= ASC_3 + \beta_1(X_{31n} - X_{11n}) + \beta_2(X_{32n} - X_{12n}) + \Gamma_3 z_n^* + \upsilon_{3n} - \upsilon_{1n}, && (32) \end{aligned}$$

or, using the difference operator $\Delta_{21} U_n = U_{2n} - U_{1n}$:

$$\begin{aligned} \Delta_{21} U_n &= ASC_2 + \beta_1 \Delta_{21} X_{1n} + \beta_2 \Delta_{21} X_{2n} + \Gamma_2 z_n^* + \Delta_{21} \upsilon_n = \Delta_{21} V_n + \Delta_{21} \upsilon_n && (33)\\ \Delta_{31} U_n &= ASC_3 + \beta_1 \Delta_{31} X_{1n} + \beta_2 \Delta_{31} X_{2n} + \Gamma_3 z_n^* + \Delta_{31} \upsilon_n = \Delta_{31} V_n + \Delta_{31} \upsilon_n, && (34) \end{aligned}$$

which is equivalent to

$$\begin{pmatrix} \Delta_{21} U_n \\ \Delta_{31} U_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta_{21} X_{1n} & \Delta_{21} X_{2n} & z_n^* & 0 \\ 0 & 1 & \Delta_{31} X_{1n} & \Delta_{31} X_{2n} & 0 & z_n^* \end{pmatrix} \begin{pmatrix} ASC_2 \\ ASC_3 \\ \beta_1 \\ \beta_2 \\ \Gamma_2 \\ \Gamma_3 \end{pmatrix} + \begin{pmatrix} \Delta_{21} \upsilon_n \\ \Delta_{31} \upsilon_n \end{pmatrix}. \quad (35)$$

In this HCM independent Probit kernel, we assume that $\upsilon_n \sim i.i.d.\ N(0, \frac{\sigma^2}{2}) \Rightarrow \Delta\upsilon_n \sim N(0, \sigma^2)$. We will normalize the model setting the scale factor equal to 1 by assuming $\sigma^2 = 1$.

<u>*Measurement equation*</u>

$$y_n = \begin{cases} 1 & \text{if } U_{1n} = max(U_{1n}, U_{2n}, U_{3n}) \Leftrightarrow (\Delta_{21}U_n < 0) \wedge (\Delta_{31}U_n < 0) \\ 2 & \text{if } U_{2n} = max(U_{1n}, U_{2n}, U_{3n}) \Leftrightarrow (\Delta_{21}U_n \geq 0) \wedge (\Delta_{21}U_n > \Delta_{31}U_n) \\ 3 & \text{if } U_{3n} = max(U_{1n}, U_{2n}, U_{3n}) \Leftrightarrow (\Delta_{31}U_n \geq 0) \wedge (\Delta_{31}U_n > \Delta_{21}U_n) \end{cases} . \quad (36)$$

5.3 Latent Variable Model

Consider now one latent variable, with one observable continuous indicator.

<u>*Structural equation*</u>

$$z_n^* = bw_n + \zeta_n, \quad (37)$$

where we will consider $\zeta_n \sim N(0, 1), \forall n$,

<u>*Measurement equation*</u>

$$I_n = \lambda z_n^* + \varepsilon_n, \quad (38)$$

with $\varepsilon_n \sim N(0, 1), \forall n$.

5.4 Gibbs sampler implementation

The parameters to estimate in the case we are analyzing are $\theta' = [ASC_2\ ASC_3\ \beta_1\ \beta_2\ \Gamma_2\ \Gamma_3]$, b and λ. Bayes estimation implementation for these parameters requires making draws from the joint posterior distribution:

$$P(\theta, b, \lambda | y, I), \quad (39)$$

or, using data augmentation, from:

$$P(\Delta U, z^*, \theta, b, \lambda | y, I), \quad (40)$$

where $\Delta U = [\Delta_{21}U, \Delta_{31}U]$, and $z^* = (z_1^*, \ldots, z_N^*)'$, $y = (y_1, \ldots, y_N)'$ and $I = (I_1, \ldots, I_N)'$ capture the information for the full group of individuals.

Using Gibbs sampling the estimators are obtained from draws inside an iterative process involving the set of *full conditional distributions*. Namely, at the g-*th* iteration:

$$\Delta U_n^{(g)} \sim \pi(\Delta U_n | z_n^{*(g-1)}, \theta^{(g-1)}, b^{(g-1)}, \lambda^{(g-1)}, {y_n}^{(g-1)}, {I_n}^{(g-1)}), \forall n \quad (41)$$

$$z_n^{*(g)} \sim \pi(z_n^* | \Delta U^{(g-1)}, \theta^{(g-1)}, b^{(g-1)}, \lambda^{(g-1)}, {y_n}^{(g-1)}, {I_n}^{(g-1)}), \forall n \quad (42)$$

$$\theta^{(g)} \sim \pi(\theta | \Delta U^{(g-1)}, z^{*(g-1)}, b^{(g-1)}, \lambda^{(g-1)}, y^{(g-1)}, I^{(g-1)}) \quad (43)$$

$$b^{(g)} \sim \pi(b | \Delta U^{(g-1)}, z^{*(g-1)}, \theta^{(g-1)}, b^{(g-1)}, y^{(g-1)}, I^{(g-1)}) \quad (44)$$

$$\lambda^{(g)} \sim \pi(\lambda | \Delta U^{(g-1)}, z^{*(g-1)}, \theta^{(g-1)}, b^{(g-1)}, y^{(g-1)}, I^{(g-1)}). \quad (45)$$

First, note that $\Delta_{21}U_n \sim N(\Delta_{21}V_n, 1)$, and that $\Delta_{31}U_n \sim N(\Delta_{31}V_n, 1)$. But we also know that $(\Delta_{21}U_n < 0) \wedge (\Delta_{31}U_n < 0)$ when y_n equals 1, that $(\Delta_{21}U_n \geq 0) \wedge (\Delta_{21}U_n > \Delta_{31}U_n)$ when y_n equals 2, and that $(\Delta_{31}U_n \geq 0) \wedge (\Delta_{31}U_n > \Delta_{21}U_n)$ when y_n equals 3. This implies that conditional on y_n, ΔU_n follows a truncated multivariate Normal (TMVN) distribution, where the truncation region $\Re$ is defined by y_n:

$$\pi(\Delta U_n | z_n^*, \theta, b, \lambda, y_n, I_n) \sim \text{TMVN}_{\Re|y_n}\left(\begin{pmatrix} \Delta_{21}V_n \\ \Delta_{31}V_n \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \forall n. \tag{46}$$

In addition, and since the latent variable z_n^* is not observable, we need to incorporate the information provided by the indicator I_n on z_n^*. This information is explicitly given by the conditional probability $\pi(z_n^*|I_n)$ whose expression depends on the assumptions we make. If we assume an independent bivariate normal distribution for the error terms of the latent variable model, then we have

$$\begin{pmatrix} z_n^* \\ I_n \end{pmatrix} \sim N\left(\begin{pmatrix} bw_n \\ \lambda bw_n \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & 1+\lambda^2 \end{pmatrix}\right), \forall n, \tag{47}$$

which implies

$$\pi(z_n^*|\Delta U, \theta, b, \lambda, y_n, I_n) \sim N(\mu_{z_n|I_n}, \sigma^2_{z_n|I_n}), \forall n, \tag{48}$$

where

$$\mu_{z_n|I_n} = bw_n + \frac{\lambda}{1+\lambda^2}(I_n - \lambda bw_n) \tag{49}$$

$$\sigma^2_{z_n|I_n} = \frac{1}{1+\lambda^2}. \tag{50}$$

Note that the latter expression is independent of the individual n, so we can write $\sigma^2_{z|I}$.

When using data augmentation, the latent variables become 'observable', as draws of both ΔU and z^* are made. This fact implies that the rest of the conditional distributions simply correspond to ordinary Bayesian regressions:

$$\pi(\theta|\Delta U, z^*, b, \lambda, y, I) \sim N(\bar{\theta}, \bar{V}_\theta) \tag{51}$$

$$\pi(b|\Delta U, z^*, \theta, b, y, I) \sim N(\bar{b}, \bar{V}_b) \tag{52}$$

$$\pi(\lambda|\Delta U, z_n^*, \theta, b, y, I) \sim N(\bar{\lambda}, \bar{V}_\lambda). \tag{53}$$

If prior beliefs for θ, b and λ are described by $p(\theta) \sim N(\check{\theta}, \check{V}_\theta)$, $p(b) \sim N(\check{b}, \check{V}_b)$, and $p(\lambda) \sim N(\check{\lambda}, \check{V}_\lambda)$ respectively, then we can show that

$$\bar{V}_b = (\check{V}_b^{-1} + w'w)^{-1}, \;\; \bar{b} = \bar{V}_b(\check{V}_b^{-1} + w'z^*) \tag{54}$$

$$\bar{V}_\lambda = (\check{V}_\lambda^{-1} + z^{*\prime}z^*)^{-1}, \;\; \bar{\lambda} = \bar{V}_\lambda(\check{V}_\lambda^{-1} + z^{*\prime}I^*) \tag{55}$$

$$\bar{V}_\theta = (\check{V}_\theta^{-1} + \widetilde{X}'\widetilde{X})^{-1}, \;\; \bar{\theta} = \bar{V}_\theta(\check{V}_\theta^{-1} + \widetilde{X}'\Delta U), \tag{56}$$

where $\widetilde{X}$ is a matrix built stacking the matrices

$$\widetilde{X}_n = \begin{pmatrix} 1 & 0 & \Delta_{21}X_{1n} & \Delta_{21}X_{2n} & z_n^* & 0 \\ 0 & 1 & \Delta_{31}X_{1n} & \Delta_{31}X_{2n} & 0 & z_n^* \end{pmatrix}, \quad \forall n.$$

Note that in order to have an unbiased estimator of θ, the unconditional z_n^* must enter $\widetilde{X}_n$.

5.5 Simulation Experiment

<u>*Choice Model*</u>

We considered a situation with three alternatives $i = \{1, 2, 3\}$ described by two attributes X_1 and X_2, for N=50,000 individuals. The attributes were built taking random draws from a Uniform distribution according to the following table:

$\sim U[min, max]$	min	max
X_1	2.0	5.0
X_2	5.0	15.0

TABLE 6 Attributes experimental design

Then, the systematic utility functions were built assuming a linear specification on the attributes and one alternative specific constant for alternatives 2 and 3. In the following table we show the taste parameters that were assumed. These values were fixed in order to achieve a relatively balanced choice. In other words, the taste values assure that each alternative has enough choices to correctly estimate the model.

Taste Parameters	
ASC_2	0.20
ASC_3	0.40
β_1	-0.05
β_2	-0.10
Γ_2	-0.50
Γ_3	-0.60

TABLE 7 Taste parameters for the choice model

The error terms v were assumed i.i.d Normal and such that $\Delta v_n \sim N(0, 1)$. The scale factor considered was then equal to 1.0 and – in this particular case – it also assures that the choice process is not completely deterministic or completely random. In fact, we checked that with the values shown above around 25% of the simulated individuals changed their choices because of the random term. Previous experiments results tell this is a good trade-off to replicate a standard discrete choice model.

Latent Variable Model

To simulate the latent component of the model we built 1 binary individual characteristic w_n, following a Bernoulli distribution with $p = 0.5$. The associated parameter, as we can see in equation 37, corresponds to b, whose assumed value is reported in Table 8. Using these values it was possible to build one - in this case observed - latent variable which was added into the utility function of alternatives 2 and 3 using Γ_2 and Γ_3 as weight.

Taste Parameters	
b	0.50
λ	0.70

TABLE 8 Latent choice model parameters

Note also that we assumed $\zeta_n \sim N(0,1)$ and $\varepsilon_n \sim N(0,1), \forall n$.

Estimated Parameters

We implemented the iterative Gibbs sampling routine presented earlier using the R language. As starting values we consider

δ	value
ASC_2	0.00
ASC_3	0.00
β_1	0.00
β_2	0.00
Γ_2	0.00
Γ_3	0.00
b	1.00
λ	1.00
$\Delta_{21}U$	0.00
$\Delta_{31}U$	0.00
z^*	0.00

TABLE 9 Starting values

In order to construct the reported results shown in Table 10 we consider a randomly selected sub-sample of 1,000 individuals from the original database of N = 50,000 individuals. The estimated parameters correspond to the means of the results for 15 repetitions of the sub-sampling process.

We observe that the estimates are not only significant but also that they significantly replicate the target values, which correspond to those used to build our simulated database. Specifically, the t-target value is calculated to test the null hypothesis of each estimate of each parameter of interest being equal to its target value.

δ	target	$\hat{\delta}$	s.e.	t test	t target
ASC_2	0.20	0.227	0.064	3.54	0.42
ASC_3	0.40	0.396	0.063	6.30	-0.07
β_1	-0.05	-0.052	0.031	-1.68	-0.07
β_2	-0.10	-0.099	0.009	-10.80	0.12
Γ_2	-0.50	-0.532	0.080	-6.62	-0.39
Γ_3	-0.60	-0.655	0.080	-8.19	-0.69
b	0.50	0.497	0.083	5.98	-0.03
λ	0.80	0.794	0.052	15.32	-0.12

TABLE 10 Estimation results - 1000 rep - 200 burn in

5.6 Probit kernel vs. MMNL kernel

When we want to incorporate flexibility into a standard choice model we have two options: Probit vs. Logit Mixture. In the simple case we studied in this section, we developed the HCM Gibbs sampler assuming a Probit kernel. When using classical simulated maximum likelihood (SML) techniques, the intricate classical Multinomial Probit estimation process reduces the practicability of the standard Probit model. In fact, the Logit Mixture (MMNL) SML estimation outperforms the Probit SML estimation because of the good statistical properties that can be derived for the MMNL estimator (Munizaga and Alvarez-Daziano, 2005). This explains the choice of a MMNL kernel when we presented the simulated maximum likelihood solution for Hybrid Choice Models in Section 3. However, working with Bayesian methods breaks down the complexity of Probit classical estimation (Bolduc, Fortin and Gordon, 1997). As we show in our simulation experiment, HCM Bayesian estimation with a Probit kernel is also straightforward. The properties of the Normal distribution allow us to exploit data augmentation techniques basically because ΔU follows a normal distribution (Albert and Chib, 1993 and McCulloch and Rossi, 2000). In fact, it is trivial to make an extension of the Gibbs sampler we developed to a Normal error component model – such as randomly Normal distributed taste variations – with a Probit kernel for the mode choice.

On the other hand, when modeling a Normal error component model with a Multinomial Logit kernel – which results in a MMNL model – we no longer have the advantageous properties that make implementing the Probit-kernel Gibbs sampling easier. The Bayesian procedure for a standard MMNL – without an associated structure of latent variables – is described by Train (2003). Since the MMNL distribution of ΔU is hard to describe, MMNL Bayesian estimation uses Metropolis-Hastings methods.

Plugging this Metropolis-Hastings procedure into our iterative process of conditional draws, we are able to estimate the whole set of HCM parameters (δ). The choice model parameters, i.e. θ, are obtained using the Metropolis-Hastings step. Note that data augmentation is no longer necessary, which explains the loss in modeling simplicity we found

for the Probit kernel case – and that we implemented in our simulation experiment. As for b and λ, note that equations (54) and (55) are still valid.

6 Conclusions

In the last decade, discrete choice modeling has evolved towards an explicit incorporation of psychological factors affecting decision making. Traditionally, discrete choice models represented the decision process as an obscure black box for which attributes of alternatives and characteristics of individuals were inputs and where the observed choice made by the individual corresponded to the output of the system. The new trend in discrete choice modeling is to enhance the behavioral representation of the choice process. As a direct result, the model specification is improved and the model gains in predictive power. Econometrically, the improved representation – called the Hybrid Choice model – involves dealing with a choice model formulation that contains latent psychometric variables among the set of explanatory variables. Since perceptions and attitudes are now incorporated, this leads to more realistic models.

In this paper we have described the Hybrid Choice model, composed of a group of structural equations describing the (potentially interrelated) latent variables in terms of observable exogenous variables, and a group of measurement relationships linking latent variables to certain observable (continuous or discrete) indicators. We have shown that although HCM estimation requires the evaluation of complex multi-dimensional integrals, both simulated maximum likelihood and Bayesian methods can be successfully implemented and offers an unbiased, consistent and smooth estimator of the true probabilities. Additionally, analyzing a simple case we showed that working with Bayesian methods can break down the complexity of classical estimation.

Using real data about virtual personal vehicle choices made by Canadian consumers when faced with technological innovations, we showed that HCM is genuinely capable of adapting to practical situations. Indeed, the results provide a better description of the profile of consumers and their adoption of new private transportation technologies. We identified two travel related dimensions with a significant impact: environmental concern and appreciation of new car features. We also successfully included in the choice model a latent income variable to account for the measurement errors associated with the reported income classes.

Further research is needed to generalize the HCM Gibbs sampler we developed in this paper. By testing the general HCM Gibbs sampler, we expect to determine when Bayesian MCMC outperforms MSL according to empirical results based on correct identification restrictions and accurate predictions.

References

[1] Albert, J.H. and S. Chib. Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, Vol. 88, No. 422, 1993, pp. 669-679.

[2] Ben-Akiva, M. Structure of passenger travel demand models. Ph.D. dissertation, Department of Civil Engineering, MIT, Cambridge, Mass. 1973.

[3] Ben-Akiva, M. and B. Boccara. Integrated Framework for Travel Behavior Analysis. IATBR Conference, Aix-en-provence, France, 1987.

[4] Ben-Akiva, M., D. Bolduc and J.Q. Park. Discrete Choice Analysis of Shipper's Preferences, in Recent Developments in Transport Modeling: Lessons for the Freight Sector, Eddy Van de Voorde editor, Elsevier, 2008.

[5] Ben-Akiva, M., D. Bolduc, and J. Walker. Specification, identification and estimation of the Logit Kernel (or Continuous Mixed Logit) model". Working Paper, MIT, 2001.

[6] Ben-Akiva M. and S.R. Lerman. *Discrete choice analysis: Theory and application to travel demand*. The MIT Press, Cambridge, Massachusetts, 1985.

[7] Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Boersch-Supan, D. Brownstone, D. Bunch, A. Daly, A. de Palma, D. Gopinath, A. Karlstrom and M.A. Munizaga. Hybrid choice models: progress and challenges. Marketing Letters 13 (3), 2002, 163-175.

[8] Bhat, C. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research B*, 35, 2001, pp. 677-693.

[9] Bolduc, D. Generalized autoregressive errors in the multinomial Probit model. Transportation Research B, 26, 1992, pp. 155-170.

[10] Bolduc, D. and M. Ben-Akiva. A multinomial probit formulation for large choice sets. *Proceedings of the 6th international conference on Travel Behaviour*, Volume 2, 1991, pp. 243-258.

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

[12] Bolduc, D., N. Boucher and R. Alvarez-Daziano. Hybrid Choice Modeling of New Technologies for Car Choice in Canada. *Transportation Research Record*: Journal of the Transportation Research Board, No.2082, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 63-71.

[13] Bolduc, D., B. Fortin, and S. Gordon. Multinomial probit estimation of spatially interdependent choices: An empirical comparison of two new techniques. *International Regional Science Review* 20, 1997, pp. 77-101.

[14] Bolduc, D. and A. Giroux. The Integrated Choice and Latent Variable (ICLV) Model: Handout to accompany the estimation software, Département d'économique, Université Laval, 2005.

[15] Brownstone, D. and K.E. Train. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89, 1999, pp. 109-129.

[16] Bunch, D. A. Estimability in the multinomial Probit model. *Transportation Research B*, 25, 1991. pp. 1-12.

[17] Dansie, B. R. Parameter estimability in the multinomial Probit model. *Transportation Research*, 19, 1985, pp. 526-528.

[18] Horne, M. Incorporating Preferences for Personal Urban Transportation Technologies into a Hybrid Energy-Economy Model. Master's Degree Thesis. School of Resource and Environmental Management. Simon Fraser University, 2003.

[19] Horne, M., M. Jaccard, and K. Tiedman. Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. *Energy Economics*, 27(1), 2005, pp. 59-77.

[20] Jöreskog, K. and D. Sörbom. LISREL VI: Analysis of linear structural relations by maximum likelihood, instrumental variables and least squares methods, User's guide, Department of Statistics, University of Uppsala, Uppsala, Sweden, 1984.

[21] McCulloch, R.R., N.G. Polson and P.E. Rossi. Bayesian Analysis of the Multinomial Probit Model wit Fully Identified Parameters. *Journal of Econometrics*, 99, 2000, pp. 173-193.

[22] McFadden, D. Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontier in Econometrics*. Academic Press, New York, 1974.

[23] McFadden, D. The choice theory approach to market research. *Marketing Science*, Vol. 5, No. 4, 1986, pp. 275-297.

[24] McFadden, D. and K.E. Train. Mixed MNL models of discrete response. *Journal of Applied Econometrics,* 15, 2000, pp. 447-470.

[25] Morikawa, T., M. Ben-Akiva and D. McFadden. Discrete choice models incorporating revealed preferences and psychometric data. *Econometric Models in Marketing* 16, 2002, pp. 27-53.

[26] Munizaga, M.A. and R Alvarez-Daziano. Testing Mixed Logit and Probit models by simulation. *Transportation Research Record* 1921, 2005, pp. 53-62.

[27] Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2003.

[28] Walker, J. The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification. *Transportation Research Record* 1805, 2002, pp. 86-98.

[29] Walker, J. and M. Ben-Akiva. Generalized random utility model. *Mathematical Social Sciences*, Vol. 43, 2002, No. 3, pp. 303-343.