RICARDO ALVAREZ DAZIANO

A Bayesian approach to Hybrid Choice models

Thèse présentée
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de doctorat en économique
pour l'obtention du grade de Philosophiæ Doctor (PhD)

Sciences Sociales
UNIVERSITÉ LAVAL
QUÉBEC

2010

Abstract

Microeconometric discrete choice models aim to explain the process of individual choice by consumers among a mutually exclusive, exhaustive and finite group of alternatives. Hybrid choice models are a generalization of standard discrete choice models where independent expanded models are considered simultaneously. In my dissertation I analyze, implement, and apply simultaneous estimation techniques for a hybrid choice model that, in the form of a complex generalized structural equation model, simultaneously integrates discrete choice and latent explanatory variables, such as attitudes and qualitative attributes. The motivation behind hybrid choice models is that the key to understanding choice comes through incorporating attitudinal and perceptual data to conventional economic models of decision making, taking elements from cognitive science and social psychology.

The Bayesian Gibbs sampler I derive for simultaneous estimation of hybrid choice models offers a consistent and efficient estimator that outperforms frequentist full information simulated maximum likelihood. Whereas the frequentist estimator becomes fairly complex in situations with a large choice set of interdependent alternatives with a large number of latent variables, the inclusion of latent variables in the Bayesian approach translates into adding independent ordinary regressions. I also find that when using the Bayesian estimates it is easier to consider behavioral uncertainty; in fact, I show that forecasting and deriving confidence intervals for willingness to pay measures is straightforward.

Finally, I confirm the capacity of hybrid choice modeling to adapt to practical situations. In particular, I analyze consumer response to innovation. For instance, I incorporate pro-environmental preferences toward low-emission vehicles into an economic model of purchase behavior where environmentally-conscious consumers are willing to pay more for sustainable solutions despite potential drawbacks. In addition, using a probit kernel and dichotomous effect indicators I show that knowledge as well as a positive attitude toward the adoption of new technologies favor the adoption of IP telephony.

Résumé

Les modèles microéconométriques de choix discrets ont pour but d'expliquer le processus du choix individuel des consommateurs parmi un ensemble limité et exhaustive d'options mutuellement exclusives. Les modèles dits de choix hybrides sont une généralisation des modèles de choix discrets standard, où des modèles indépendants plus sophistiqués sont considérés simultanément. Dans cette thèse des techniques d'estimation simultanée sont analysées et appliquées pour un modèle de choix hybride qui, sous la forme d'un système complexe d'équations structurelles généralisées, intègre à la fois des choix discrets et des variables latentes en tant que facteurs explicatifs des processus décisionnels. Ce qui motive l'étude de ce genre de modèles est que pour comprendre le processus du choix il faut incorporer des attitudes, des perceptions et des attributs qualitatifs à l'intérieur de modèles décisionnels économiques conventionnels, tout en prenant ce qui dit la recherche en sciences cognitives ainsi qu'en psychologie sociale.

Quoique l'estimation du système d'équations d'un modèle de choix hybride requière l'évaluation d'intégrales multidimensionnelles complexes, on résoudre empiriquement ce problème en appliquant la méthode du maximum de vraisemblance simulée. Ensuite on dérive une procédure d'échantillonnage de Gibbs pour l'estimation simultanée bayésienne du modèle qui offre des estimateurs convergents et efficaces. Ceci devient une méthode plus avantageuse comparativement aux méthodes classiques dans un cadre analytique avec un grand nombre de variables latentes. En effet, en vertu de l'approche bayésienne il suffit de considérer des régressions ordinaires pour les variables latentes. Par ailleurs, dériver les intervalles de confiance bayésiennes pour les parts de marché ainsi que pour des dispositions à payer devient trivial.

De par sa grande géneralité, le modèle de choix hybride est capable de s'adapter à des situations pratiques. En particulier, la réponse des consommateurs suite à l'innovation technologique est analysée. Par exemple, on étudie les préférences pro-environnementales dans un modèle économique des décisions d'achat de véhicules verts selon lequel les consommateurs soucieux de l'environnement sont prêts à payer davantage pour des véhicules à faibles émissions, en dépit des inconvénients potentiels. En outre, en utilisant un noyau probit et des indicateurs dichotomiques on montre que des connaissances préalables ainsi que des attitudes positives envers l'adoption de nouvelles technologies favorisent l'adoption de la téléphonie IP.

Acknowledgements

NEL MEZZO DEL CAMMIN DI NOSTRA VITA, MI RITROVAI PER UNA SELVA OSCURA, CHÉ LA DIRITTA VIA ERA SMARRITA. Dante, Inferno

Not that going through a PhD is an inferno, in fact it is far beyond that, but in more than one occasion I did wonder if a guy who was making a profitable career as a consultant made a mistake in moving from sunny and Mediterranean central Chile to (really) cold but charming Québec City to begin a PhD with the idea of pursuing an academic career. Many thanks to those who proved me wrong in my doubts: this selva selvaggia was not really that terrible after all.

I would like to start thanking all the members of my dissertation committee, professors Philippe Barla, Michel Bierlaire, and Stephen Gordon, for their participation in evaluating my work. I am very thankful to Denis Bolduc, my advisor and now my friend, who gave me all the freedom to pursue my research without any constraint.

Because of Denis' confidence in my work, I had an incredible opportunity to establish my own directions of research with plenty of flexibility. This flexibility not only permitted me to test my autonomy and capabilities for academic research but also afforded me an impressive freedom of movement. I presented papers at various conferences and I value all the comments and questions that helped me to improve several sections of this dissertation. But my movements also included quite a few getaways. However, I never forgot to take my laptop with me and I can testify that inspiration can appear suddenly anywhere. Parts of my dissertation were written in such disparate places as Chilean Patagonia and the Amalfi coast. An important part of my attention deficit disorder was dissipated at the Penn Bookstore Café in Philadelphia, my second city in North America, where amidst my simulations I enjoyed various pauses santé, facebooking with a cappuccino at hand and taking a look at a free copy of the Corriere della Sera (and some weekly gossip mags, I have to admit.)

I also thank professors David Brownstone and Stefen Proost, who made very interesting remarks on an early version of my JMP during that magnificent experience that was my participation at the Venice International University Summer School in Resources and

Environmental Economics. The small and slightly claustrophobic island of San Servolo will always remain in my memory, especially because of escapades in the different sestieri of Venice with the other students.

I am also grateful of the Government of Canada, who directly financed my first two years at Laval through the Government of Canada Award program, and then indirectly through funds coming from Natural Resources Canada.

I also have to mention the Chilean team: Francisca Yáñez, Sebastián Raveau, and Juan de Dios Ortúzar, whom I really enjoyed collaborating with and meeting in Harrogate, Padova, DC, and Lisbon. Special thanks to members of the transportation research group at Universidad de Chile.

Je dois remercier aussi mes chers collègues étudiants du département d'économique avec lesquels j'ai partagé toutes ces années. J'en remercie notamment Myra, par nos discussions sur l'humain et le divin, notre soutien mutuel comme futurs économètres appliqués, et... ben, aussi par nos pauses thé.

À Rölfi, qu'on le couvre d'or! O SERENISSIME ROLFE, AMICE MAGNE, SINE TE NON VALEO. Glōriam tuam & perpetuum amōrem tuum semper laudō.

Last but not least (do people always have to say this?), I thank my extended family. Especialmente a mis padres, que siempre me hicieron llegar su apoyo pese a la lejanía. Thanks to my parents for all the effort and work that made everything possible for me.

The Parker house, Oregon, October 2010

Ἅδιον οὐδὲν ἔρωτος· ἃ δ᾽ ὄλβια δεύτερα πάντα
ἐστίν· ἀπὸ στόματος δ᾽ ἔπτυσα καὶ τὸ μέλι.
τοῦτο λέγει Νοσσίς· τίνα δ᾽ ἁ Κύπρις οὐκ ἐφίλασεν
οὐκ οἶδεν κήνα γ᾽ ἄνθεα ποῖα ῥόδα.

NOTHING is ſweeter than Love; & every other joy
is ſecond to it: even the honey I ſpit out of my mouth.
Thus Noſſis ſays: and he who hath never loved Cypris,
knows not at all what roſes her flowers are.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Dissertation statement and methodology

The study of uncertainty has become a paramount topic for several fields in economics, statistics, and psychology. Uncertainty, broadly defined, accounts for a world that is probabilistic in nature. As economists we aim at modeling the decision-making process of consumers, but we need to take into account behavioral uncertainty. In my dissertation I address and integrate different dimensions of behavioral uncertainty into one microeconometric model of choice.

Individuals act according to the expected consequences, or payoffs, of their actions (behavior). From the point of view of economics, the cognitive evaluation of behavior[1] and the consequent behavioral response[2] are reflected by *preferences* through an unobservable construct that is known as *utility*. However, psychology theories propose that to evaluate the potential outcomes of behavior we construct internal unobservable measures denominated *attitudes*. Standard economic models of choice neglect the role of attitudes. In addition, standard choice models often neglect the impact of other variables that do not have a natural order or an overt measurement scale, such as quality.

In the general context of *random utility maximization* and of discrete choice models in particular, in this dissertation I analyze how the decision-making process can be modeled

[1]The cognitive evaluation is associated with the expected consequences.

[2]Choice.

to account for the impact of diverse unobservable entities that are related to choice. I build on hybrid choice models (HCMs), which are a generalization of standard discrete choice models where independent expanded models are considered simultaneously. One of these expansions is a generalized structural equation model that simultaneously accommodates a discrete choice model with latent explanatory variables[3]. Even though the idea of introducing latent variables into the choice process is not new, I identify several econometric challenges associated with practical implementation of HCMs.

Hence, the main objective of this dissertation is to study both theoretically and empirically the application of simultaneous econometric estimation of HCMs. First, I discuss that omitting relevant latent variables or measuring the underlying latent concept with error result in problems of endogeneity. Second, the system of equations that describes an HCM needs to be solved simultaneously to derive estimators that are both consistent and efficient.

To achieve the main objective I pursue the following interrelated themes: *advanced econometric estimation techniques for HCMs* and *empirical performance of HCM estimation methods*. The first project involves derivation of the estimators as well as computer implementation of the approaches analyzed in this research. In the second project, I apply the estimators to practical situations.

For the project *advanced econometric estimation techniques for HCMs*, I address simultaneous estimation of HCMs, first using an efficient choice probability simulator through maximum simulated likelihood (MSL) estimation. Although feasible, the MSL approach necessary for classical estimation is very demanding in situations with a large choice set of interdependent alternatives with a large number of latent variables. In fact, the latent variables affect the behavior of the simulated likelihood function in such a way that a standard optimization algorithm may require a huge number of iterations to converge.

For these reasons, I then propose to go beyond classical methods by introducing Bayesian econometrics and exploring whether Bayesian techniques represent an attractive alternative to HCM frequentist estimation. Building on the rapid development of Markov chain Monte Carlo (MCMC) techniques and on the idea that Bayesian tools could be used to produce estimators that are asymptotically equivalent to those obtained using classical methods, I take as my goal to implement a Bayesian approach to hybrid choice modeling. To ensure calculation speed, the software needed is written in the R programming language, which is widely used among Bayesian practitioners.

[3]Attitudes, but also quality.

Interestingly, the introduction of Bayesian tools adds another dimension to handling uncertainty. When embracing the Bayesian approach, we recognize that we are uncertain about the true state of the world, which is expressed by the true parameters of the econometric model. As opposed to the frequentist approach, which handles uncertainty about the true parameters by considering them as fixed but unknown constants, the Bayesian approach considers the true parameters to be random variables. This difference is fundamental, because in Bayesian econometrics to make inferences about the parameters we can introduce prior knowledge or beliefs and apply the rules of probability directly. Hence, the Bayesian approach appears more akin to a choice model that accounts for perceptions and beliefs.

Finally, the contribution of my dissertation consists not only in solving econometric challenges related to simultaneous estimation of HCMs. I also discuss the theoretical foundations that support the integration of economics and a cognitive model of agent behavior. In addition, the relevance of my work also comes from empirical applications which are not only used to test the performance of the estimator, but also provide interesting results that cannot be derived when using standard choice models. In fact, results of the project *empirical performance of HCM estimation methods* are key to understand consumers' response to innovation.

1.2 Principles of Bayesian econometrics

Consider the general dominated parametric model[4]

$$(\mathcal{Y}, \mathcal{P} = P_\theta = \ell(y;\theta) \cdot \mu, \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1) \tag{1.1}$$

where $\mathcal{Y}$ is the sample space, $\mathcal{P}$ is a parameterized family of probability density functions P_θ on $\mathcal{Y}$, $\ell(y;\theta)$ is the likelihood function, μ is the dominating measure, θ is a vector of p parameters, and Θ is the parameter space. In the case of a sampling model, elements in $\mathcal{Y} \subseteq \mathbb{R}^N$ are composed by random samples with i.i.d components $\{y_i\}_{i=1}^N$.

In parametric statistics, the (point) estimation problem reduces to propose a value $\hat{\theta}$ to the true but unknown parameter θ in the model $(\mathcal{Y}, \mathcal{P} = P_\theta)$. One of the most popular point estimation methods in frequentist statistics is maximum likelihood estimation.

Suppose that the joint distribution of $Y = (Y_1, \ldots, Y_N)'$ of a parametric model $(\mathcal{Y}, \mathcal{P} =$

[4] The section overviews concepts that are treated in detail in **?**, **?**, **?**, and **?**.

P_θ) admits a density $\ell(y;\theta) = f(y_1, \ldots, y_N; \theta)$. Once y is observed, the maximum likelihood (ML) method keeps the value $\hat{\theta}(y)$ that maximizes $\ell(y;\theta)$ as the ML estimate of θ.

Interestingly, the general estimation problem can be regarded as statistical decision making, which is parallel to the decision-making problem in economics. In statistics, we have to choose $\hat{\theta}$ among various possible values. For this, we need decision rules that can be found in decision-making theory.

1.2.1 Elements of decision theory

Given a model $(\mathcal{Y}, \mathcal{P})$, a nonrandomized decision rule δ serves to provide a response $\delta(y) \in \mathcal{D}$ to every possible $y \in \mathcal{Y}$, where $\delta(y)$ is called a decision, and $\mathcal{D}$ is the space of all possible decisions. For instance, in the case of point estimation[5] δ is an estimator, whereas $\delta(y)$ is an estimate. Note that the point estimation problem comes from the fact that the true parameter θ is not known. If θ were known, the decision rule would deterministically choose the correct decision $\delta(P_\theta) = \theta$. Since θ is unknown, correct decisions are not feasible and the decisions that we take come with an associated cost. A loss function $L(\delta(y), \theta)$ is a nonnegative function indicating the cost or loss incurred by taking the decision $\delta(y)$ when the true parameter is θ. Under a correct decision, the loss function is $L(\theta, \theta) = 0$. In fact, $L(\delta(y), \theta) = 0 \iff \delta(y) = \theta, \forall y \in \mathcal{Y}$.

Suppose that we want point estimation of a scalar function $g(\theta) \in G = g(\Theta) \subseteq \mathbb{R}$. The typical loss function for this problem is the scalar quadratic loss function $L(\delta(y), \theta) = (\delta(y) - g(\theta))^2$. Note that the loss function depends ex post on y, and that prior to the observations the loss function $L(\delta(Y), \theta)$ is a random variable. Thus, we can define an average loss, which is operationalized through the risk function $R(\delta, \theta) = \mathbb{E}_\theta \left(\delta(y) - g(\theta)\right)^2$. The risk function provides a partial preorder allowing us to rank decision rules. In general, the risk function of a nonrandomized decision rule δ is given by

$$R(\delta, \theta) = \mathbb{E}_y L(\delta(Y), \theta) = \int_{\mathcal{Y}} L(\delta(y), \theta)\ell(y;\theta) d\mu(y).$$

Note that frequentist maximum likelihood estimation, which seeks the parameter values that are most likely to have produced the distribution of the observations $y \in \mathcal{Y}$, can be

[5] Other problems in statistics that can be viewed as a decision problem are the problems of interval estimation, testing, model selection, and prediction.

interpreted as minimizing the empirical risk with an appropriately chosen loss function (the negative log-likelihood).

1.2.2 Overview of Bayesian decision making

Bayesian decisions represent a decision-making process under uncertainty.

Definition: Bayesian decision problem. *To choose an **action** $\mathbf{a} \in A \subseteq \mathbb{R}^p$, the decision maker minimizes the Bayes risk function*

$$R(\mathbf{a}) = \int_{\Omega}\int_{\mathcal{Y}} L(\mathbf{a},\omega)\ell(y;\omega)p(\omega)d\mu(y)d\mu(\omega) = \mathbb{E}_{\omega}\mathbb{E}_{y}L(A,\omega), \tag{1.2}$$

*where the loss function $L(\mathbf{a},\omega)$ sets the decision criterion when action $\mathbf{a}$ is taken and the **true state of the world** is $\omega \in \Omega$. The probability $p(\omega)$ expresses the **uncertainty** about the true state of the world before new evidence y is provided.*

The idea behind Bayesian decision making is that when a decision maker has to make a choice under uncertainty, the optimal decision is based on the minimum expected cost of being wrong given the beliefs about the state of the world: $\hat{\mathbf{a}} = \text{argmin} R(\mathbf{a})$.

Even though I will apply the Bayesian approach to statistical inference, it is instructive to present first the analogous general approach to decision making, basically because this motivates the analysis of the role of the decision maker's beliefs about the unknown state of the world. In fact, Bayesian decision theory is analogous to the concept of **consumer behavior under uncertainty** in economics. In economics a Bayes decision is taken according to a maximum expected utility principle. Note that the loss function described above can be interpreted as a **disutility** to be minimized.

1.2.3 Bayesian inference

In a Bayesian setting of statistical decision problems, parameters of a model $(\mathcal{Y}, \mathcal{P} = P_\theta)$ are assumed to have a prior statistical distribution $p(\theta)$ that describes the probability distribution of θ before the observation of y. The consideration of θ being a random variable is what distinguishes the Bayesian approach from classical statistics. This notion is fundamental for Bayesian inference and is derived from the concept of subjective probabilities[6] (probability laws under uncertainty). The combination of the prior distribution $p(\theta)$ with the information coming in via the sample data $y \in \mathcal{Y}$ determines the posterior distribution of the parameters $p(\theta|y)$ with respect to a measure μ. The posterior and prior distributions are related following Bayes' theorem according to

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)},$$

where $p(y|\theta)$ represents the distribution of the observations y for every particular value of θ, and $p(y)$ is the marginal distribution of the data, which is sometimes called the predictive density of y. Note that $p(y|\theta) = \ell(y;\theta)$ by definition. Since $p(y)$ is a constant that does not depend on the observations y[7], for inference purposes Bayes' theorem is rewritten as

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

which emphasizes the Bayesian notion of updating knowledge through evidence.

Using the concepts of (Bayesian) decision making, the action **a** corresponds to finding an estimate $\hat{\theta} = \delta(y) \in \mathcal{D}$ of the true parameter $\theta \in \Theta$, which describes the state of the nature $\omega = \omega(\theta)$. The latter distinction puts in evidence the presence of an econometric model, with a structural relation linking $\mathcal{Y}$ and $\mathcal{P} = P_\theta$. The Bayes decision function is built by minimizing the Bayes risk $R(\hat{\theta} = \delta(y))$ as defined in equation **??**. Thus, a Bayes estimator is the decision that minimizes the Bayes risk. Note however that if the decision is taken ex post (after the observation of y), the Bayes decision can be taken using a posterior Bayes risk, according to

$$R(\delta|y) = \int_\Theta L(\hat{\theta}, \theta)p(\theta|y)d\mu(\theta). \tag{1.3}$$

The estimator that minimizes the posterior Bayes risk $\forall y \in \mathcal{Y}$ also minimizes the Bayes risk and therefore is a Bayes estimator.

The most common loss function $L(\hat{\theta}, \theta)$ used for Bayesian estimation is the general quadratic loss

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)'Q(\hat{\theta} - \theta), \tag{1.4}$$

where Q is a positive definite matrix.

[6]Subjective probabilities measure the beliefs about the occurrence of a particular event.

[7]$p(y) = \int_\Theta p(y|\theta)p(\theta)d\mu(\theta)$.

Theorem 1: Bayes decision with a quadratic loss. *With a quadractic loss function* **??**, *the Bayes decision $\hat{\theta}$ is unique and corresponds to the mean of the posterior distribution $p(\theta|y)$*

$$\hat{\theta} = \int_{\Theta} \theta p(\theta|y) d\mu(\theta) = \mathbb{E}(\theta|y). \quad (1.5)$$

The use of alternative risk functions yields different point estimators. For instance, when using a linear loss function, the Bayesian estimate is the median of the posterior distribution. In general, a linear-linear loss function yields to a Bayes decision equal to the posterior q^{th} quantile.

Corollary 1: Precision of the Bayes decision with a quadratic loss. *With a quadractic loss function* **??**, *an unbiased estimator in the Bayesian sense of the precision (risk) of the Bayes decision $\hat{\theta} = \mathbb{E}(\theta|y)$ is the posterior variance*

$$\mathbb{V}(\hat{\theta} - \theta) = \mathbb{V}\mathbb{E}(\hat{\theta} - \theta|y) + \mathbb{E}\mathbb{V}(\hat{\theta} - \theta|y) = \mathbb{E}\mathbb{V}(\theta|y). \quad (1.6)$$

In sum, with a quadratic matrix risk function, the calculation of the first and second moments of the posterior distribution are of fundamental interest. It is important to remark that $\hat{\theta} = \mathbb{E}(\theta|y)$ is the Bayes decision for the point estimation problem, which is equivalent to assessing that $\hat{\theta}$ is the best estimator of θ in the Bayesian sense with a quadratic loss, but still in Bayesian econometrics the parameter θ is random as opposed to the frequentist problem where the true parameter is fixed. Hence, the description of the whole posterior distribution is of interest. This distinction is relevant in the sense that Bayesian econometrics provides tools for finite samples, whereas classical inference is based on unobserved independent repeated samples. In effect, the difference between Bayesian and classical inference is clearly exemplified by the interval estimation problem.

Definition: Credible region or Bayesian confidence region. *A set $\mathcal{C} \subseteq \Theta$ such that*

$$P(\theta \in \mathcal{C}) = \int_{\mathcal{C}} p(\theta|y) d\mu(\theta) = 1 - \alpha, \quad (1.7)$$

where $(1 - \alpha)$ is a credibility level.

Note that a credible region is a fixed area containing θ with a specified coverage probability $(1 - \alpha)$, conditional on the observed data y. The frequentist confidence region is a completely different concept. First, under a classical perspective θ is fixed and as such there is no sense in constructing a region based on its distribution. Second, a non-Bayesian confidence region is constructed using the unobserved sampling distribution of the estimator. This sampling distribution, which reflects the idea that different point estimates are generated over independent repeated replications of the data, cannot be obtained for small samples. For this reason, a classical confidence region is asymptotic in the sense that the region depends on the distribution of unobserved realizations of the data; this distribution can be described using large sample theory.

Even though Bayesian and classical inference are intrinsically different, a Bayesian estimate depends on the data and a different sample will generate a different estimate. If we simultaneously[8] consider different independent repeated replications of the data, one can generate a sequence of Bayesian decisions (point estimates) for which an asymptotic distribution can be derived. This asymptotic distribution can be seen as a classical interpretation of the Bayesian estimator. In fact, the large sample properties of the Bayesian point estimate $\hat{\theta}$ in **??** are closely related to classical properties of the maximum likelihood estimator.

Theorem 2: the Bernstein-von Mises theorem. *Consider the Bayesian point estimator $\hat{\theta} = \mathbb{E}(\theta|y)$. Let $\hat{\theta}_{ML}$ be the maximum likelihood estimator[9]. If we denote the true value of θ by θ_0[10], it can be shown[11] that*

1. $\sqrt{N}(\theta - \hat{\theta}) \xrightarrow{d} MVN\,(0, \mathcal{I}^{-1}(\theta_0))$,
2. $\sqrt{N}(\hat{\theta} - \hat{\theta}_{ML}) \xrightarrow{d} 0$,
3. $\hat{\theta} \overset{a}{\sim} MVN\,(0, \mathcal{I}^{-1}(\theta_0)/N)$,

where $\mathcal{I}(\theta_0)$ is the asymptotic Fisher information matrix

$$I(\theta_0) = \underset{N}{plim} - \frac{1}{N}\frac{\partial^2 \ln p(y|\theta_0)}{\partial\theta\partial\theta'}. \tag{1.8}$$

[8] In principle under the Bayesian approach, the posterior of a previous realization could be taken as prior for the next realization. For analyzing asymptotic performance we rule out this possibility.

[9] The maximum likelihood estimator $\hat{\theta}_{ML}$ converges to θ_0 at the rate $1/\sqrt{N}$ and satisfies the first order conditions $\partial \ln p(y|\hat{\theta}_{ML})/\partial\theta = 0$.

[10] $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ represents the true but unknown parameter that generates the observations y.

[11] Note that all convergence results are defined in the classical sense. The prior distribution is assumed continuous with the Lebesgue measure on $\mathbb{R}^p$ as dominating measure.

The Bernstein-von Mises theorem not only establishes that $\hat{\theta}$ is asymptotically unbiased (consistent), normal, and efficient[12], but it also proves the relationship between the maximum likelihood and Bayes estimators. In fact, when N is large, the Bayes and ML estimators are approximately equal. If the prior distribution $p(\theta)$ is continuous with a strictly positive density on a neighborhood of the true θ_0, the asymptotic properties of consistency and efficiency of $\hat{\theta}$ do not depend on the choice of $p(\theta)$: if N is large, the evidence provided by the observations y is such that a priori information can be neglected.

I now return to the prior distribution $p(\theta)$. A prior reflects knowledge and beliefs, but mathematically prior distributions are chosen inside a family of parametric probability distributions that allow us to incorporate beliefs in a convenient way. Note that the posterior distribution may become the prior distribution for a subsequent problem. As we just saw, the importance of the prior distribution disappears as the sample size increases. However, the use of Bayesian inference is particularly interesting for small samples where the role of the prior distribution is potentially relevant. In general, even for small samples the relative importance of the prior distribution is proportional to its precision: the effect of the prior gradually disappears as the prior variance increases. The last result translates into the notion of diffuse or noninformative priors. A diffuse or noninformative prior is a distribution that is widely dispersed, at least over Θ where the likelihood function is concentrated. A flat prior distribution with an infinite integral is called an improper prior.

1.2.4 Markov chain Monte Carlo methods

In the previous subsection we saw that Bayesian inference examines the posterior distribution $p(y|\theta)$, and in particular the posterior first and second moments. Although the density of the posterior distribution can be obtained using Bayes' theorem, the main difficulty concerns the characterization of this distribution, because $p(y|\theta)$ is not always known explicitly. For this, Bayesian inference makes use of Markov chain Monte Carlo methods.

Markov chain Monte Carlo (MCMC) methods are a class of stochastic sampling algo-

[12] The posterior variance is approximately equal to the estimated Fisher information matrix, i.e. $N\mathbb{V}(\theta|y) \approx \mathcal{I}^{-1}(\theta_0)$.

rithms based on constructing a Markov chain[13] that has the desired distribution[14] as its equilibrium distribution. There are different MCMC methods, but Gibbs sampling and Metropolis Hastings are the most typically applied.

1.2.5 Gibbs sampling

Let $\mathfrak{P}$ be a partition of $\theta \in \Theta$ such that $\theta' = (\theta'_{(1)}, \dots, \theta'_{(\mathfrak{P})})$. For every subvector $\theta_{(\mathfrak{p})}$, we define $\theta'_{<(\mathfrak{p})} = (\theta'_{(1)}, \dots, \theta'_{(\mathfrak{p}-1)})$, $\theta'_{>(\mathfrak{p})} = (\theta'_{(\mathfrak{p}+1)}, \dots, \theta'_{(\mathfrak{P})})$, $\theta'_{<(1)} = \{\emptyset\}$, $\theta'_{>(\mathfrak{P})} = \{\emptyset\}$, and $\theta'_{-(\mathfrak{p})} = (\theta'_{<(\mathfrak{p})}, \theta'_{>(\mathfrak{p})})$. The partition is chosen such that the full conditional distributions $\pi(\theta_{(\mathfrak{p})}|\theta_{-(\mathfrak{p})})$ are easy to describe in the sense that it is possible to draw directly from each $\pi(\theta_{(\mathfrak{p})}|\theta_{-(\mathfrak{p})}), \forall \mathfrak{p}$. The Gibbs sampler is an algorithm based on an MCMC where the transition process from $\theta^{(g-1)}$ to $\theta^{(g)}$ for $p(\theta|y)$ at the g^{th} iteration is defined through

$$\theta^{(g)}_{(\mathfrak{p})} \sim \pi(\theta_{(\mathfrak{p})}|\theta^{(g)}_{<(\mathfrak{p})}, \theta^{(g-1)}_{>(\mathfrak{p})}, y), \forall \mathfrak{p} \in \mathfrak{P}.$$

It can be shown that this reversible Markov chain generates an instance from the posterior distribution at each iteration, i.e. $\theta^{(g)}_{(\mathfrak{p})} \sim p(y|\theta), \forall g$.

1.2.6 Metropolis Hastings

Sometimes direct sampling for one or more of the conditional distributions inside the Gibbs sampler is difficult. For Metropolis-Hastings implementation, a candidate $\theta^{cand} \in \Theta$ is drawn from the transition probability $q(\theta^{cand}|\theta^{curr})$ of generating candidate θ^{cand} given $\theta^{curr} \in \Theta$, such that $\theta^{curr} \sim p(\theta, y)$. The candidate realization θ^{cand} is then compared to the current $\theta^{curr} \in \Theta$ through the acceptance ratio:

$$\alpha = \min \left\{ 1, \frac{p(y|\theta^{cand})p(\theta^{cand})}{p(y|\theta^{curr})p(\theta^{curr})} \cdot \frac{q(\theta^{cand}|\theta^{curr})}{q(\theta^{curr}|\theta^{cand})} \right\}.$$

Starting with an arbitrary value $\theta^{(0)}$, in the Metropolis Hastings algorithm at the g^{th} iteration the candidate is accepted as the new $\theta^{(g)} = \theta^{cand}$ with probability α, while the old one is preserved $\theta^{(g)} = \theta^{curr}$ with probability $1 - \alpha$. In the case $q(\theta^{cand}|\theta^{curr}) = q(\theta^{cand} - \theta^{curr})$ the generating process of the candidate θ^{cand} is a random-walk Metropolis

[13] A Markov chain is a stochastic process that possesses the Markov property, namely that given the present, it is possible to forecast the future independently of the past.

[14] In Bayesian inference, the desired distribution is the posterior $p(y|\theta)$.

chain. If the proposal density is such that $q(\theta^{cand}|\theta^{curr}) = q(\theta^{cand})$, the process is a Metropolis independence chain. It can be shown that $p(\theta^g = \theta^{curr}) = p(\theta^{g-1} = \theta^{curr})$. Since $\theta^{curr} \sim p(\theta, y)$, it follows that $\theta^g \sim p(\theta|y), \forall g$. Thus, realizations in the Metropolis Hastings algorithm generates instances from the posterior distribution. Note that the Gibbs sampler is a special case of the Metropolis Hastings algorithm, where the proposal density is given by the conditional distributions and the acceptance ratio equals 1.

1.3 Dissertation Outline

The rest of this dissertation is organized as follows.

In Chapter 2 I motivate the introduction of hybrid choice models through an overview of the role of attitudes in explaining behavior. Then, in Chapter 3 I describe the classical estimation techniques for a full information 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.

In Chapter 4, aiming at understanding the effects of climate change and energy security concerns on travel behavior I analyze pro-environmental preferences toward low-emission vehicles. For this, using real data about purchase intentions of low-emission vehicles by Canadian consumers, I analyze the practical feasibility of a Bayesian estimator for hybrid choice models.

Chapter 5 generalizes the MCMC method for Bayesian estimation of HCMs. After describing the system of equations and deriving the estimators following Gibbs sampling, a Monte Carlo study is performed. Using a virtual case of travel mode choice, I compare the empirical performance of HCM Bayesian and classical point estimates in terms of accuracy, statistical significance, and efficiency.

In Chapter 6 I apply the hybrid choice modeling framework (and the general estimator derived in Chapter 5) to another empirical application with real data. In a telecom context, this application seeks to understand consumer behavior toward IP telephony access in Japan.

Finally, in Chapter 7 I conclude and identify lines of future research.

References

J. Geweke. *Contemporary Bayesian Econometrics and Statistics.* Wiley-Interscience, Hoboken, NJ, 2005.

C. Gourieroux and A. Monfort. *Statistics and econometric models.* Cambridge University Press, New York, NY, 1995.

E. Greenberg. *Introduction to Bayesian Econometrics.* Cambridge University Press, New York, NY, 2004.

T. Lancaster. *An Introduction to Modern Bayesian Econometrics.* Wiley-Blackwell, Malden, MA, 2004.

Chapter 2

Attitudinal response in discrete choice: the hybrid choice model

Recent research in social psychology has been centered on the relationship between attitudes and behavioral intentions that may eventuate in actual behavior. In this chapter, through an overview of the role of attitudes in explaining behavior I motivate the introduction of latent endogenous constructs as explanatory variables inside a utility function representing economic preferences. I review how attitudinal constructs, which represent the individual predisposition to evaluate entities in favor or disfavor, can be measured in practice. I also review the main models of the interaction between attitudes and behavior, as well as economic models of consumer behavior. Then, I recast hybrid choice models in a general microeconometric framework that is capable of including both qualitative attributes and attitudes as latent variables in a standard discrete choice model. Since psychometric studies provide theories and evidence on how attitudes affect behavior, if we omit attitudes in the context of economic choice we expect to face endogeneity problems. Thus, the estimators of the taste parameters will be inconsistent. This is also a problem for other estimators based on the estimates of the marginal utilities, such as elasticities, willingness to pay measures, and derived demands. As a result, the hybrid choice model emerges not only as a tool to improve the behavioral representation of the choice process, but also as a way to deal with endogeneity in discrete choice models.

2.1 Introduction

The literature on attitudes and the relation of attitude-to-behavior is extremely vast, and it reflects one of the most relevant topics of research in social psychology. In addition, consumer behavior is the original and fundamental focus of microeconomics and thus the literature on economic preferences is huge. My review is therefore selective in the sense that I present the fundamentals of the theoretical concepts that are relevant for understanding and developing an integrated attitudinal model of economic preferences.

2.2 The role of attitudes in explaining behavior

In both social and behavioral sciences, an important subject of study is the impact of **attitudes** on behavior. Attitudes measure the psychological continuous evaluation of *favor*[1] or *disfavor* assigned by the individual to a particular entity (the *attitude object*), including behavior itself (**???**). Attitudes are generally described as being rather stable, and yet subject to some change. However, when there is a change in attitudes, the old attitude is not completely replaced. Thus, when attitudes change, the new attitude dominates but may not replace the old attitude (**?**)[2].

Attitudes have a perceptual nature: diverse stimuli generate a perception that is evaluated by the individual according to imperfect information leading to an attitudinal response toward the attitude object (**?**). This evaluation, which is not directly observable, can be affective, cognitive or behavioral, reflecting the multidimensionality of attitudes (**??**); this model, sketched in Figure **??**, is also known as the tripartite model of attitude structure (**??**). The affective component reflects the feelings that the evaluated entity evokes in terms of emotions such as desire, happiness, fear and empathy. The cognitive component exhibits individual beliefs about the likelihood of a relationship between the evaluated entity and a particular outcome. Cognitive attitudes may reflect knowledge (past experience) as well as prejudices (stereotypes). Finally, the behavioral component is related to intended actions. In Figure **??** I present an example of a tripartite attitudinal evaluation regarding the adoption of low-emission cars.

In sum, attitudes are unobservable but are reflected by individual feelings, beliefs and

[1] Positive, pro, support, like, agree.

[2] Whereas the new attitude is explicit, the old attitude remains as an implicit construct.

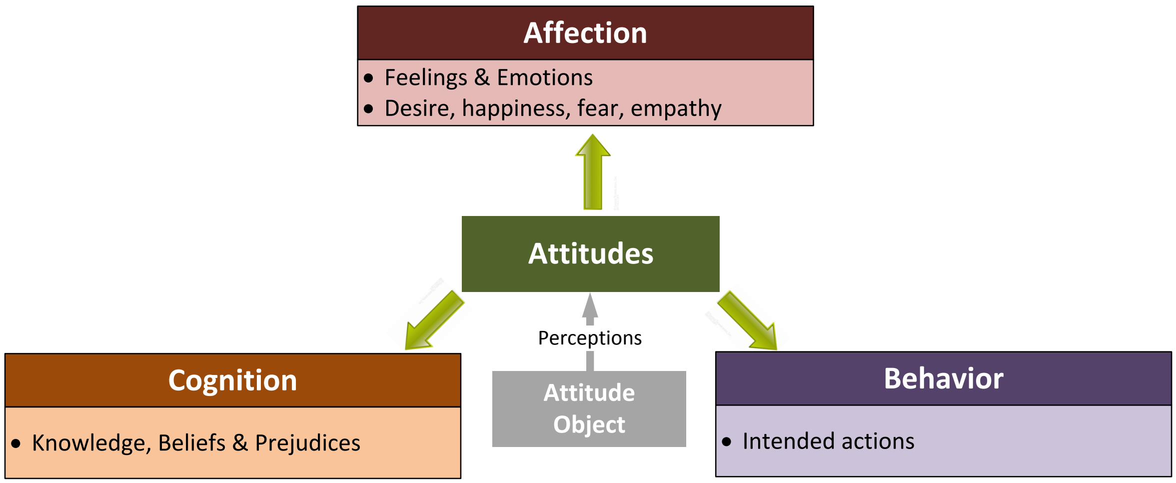

Figure 2.1: The tripartite model of attitude structure

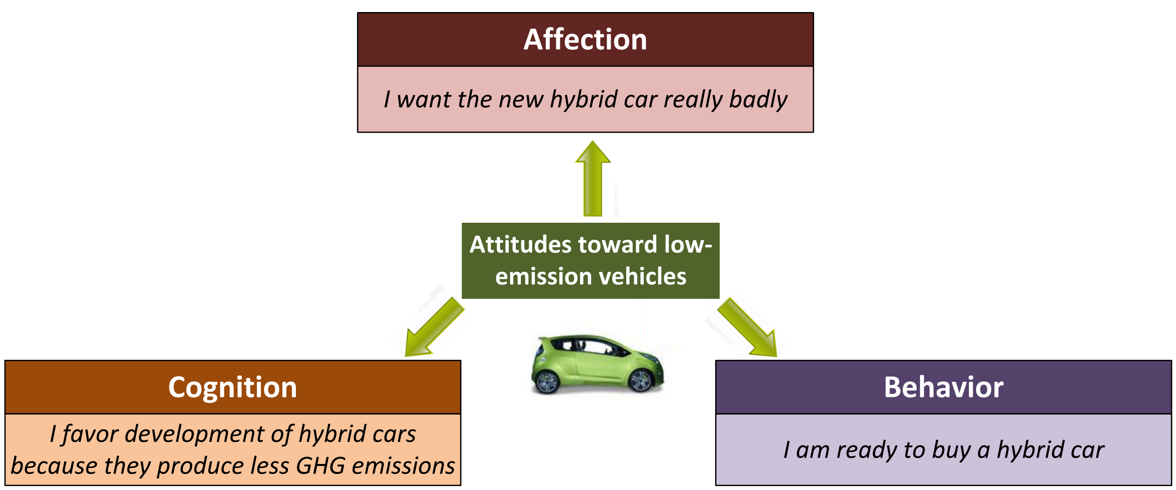

Figure 2.2: The tripartite model of attitude structure: vehicle choice

intended behavior.

Whereas it seems rational to think that an individual acts according to his personal beliefs, in a large number of occasions observed actions and stated intentions do not coincide. Since the beginning of the scientific analysis of attitudes, social psychology theorists have noted a discrepancy between intentions and actions. In fact, until the first half of the 1970s, researchers were rather pessimistic about the existence of a relationship between stated attitudes and actual behavior (**??**). In other words, we could not, apparently, predict behavior based on what people say they will do. However, **??** argued that there is a predictive value in the attitude-behavior relationship, but only within a context of correspondence between the concepts of target, action, context, and time.

To explain the attitude-behavior link, in the mid-1970s Fishbein and Ajzen developed the theory of reasoned action (TRA), revising and expanding Fishbein's expectancy-value theory (EVT). According to EVT (**?**), the evaluative dimension of an attitude depends on a combination of two factors: the beliefs that individuals have regarding the results implied by behavior (expectancy) multiplied by the importance that the individuals assign to these possible results (value). Expectancies are related to the perceived probability of a particular behavior producing a certain outcome (beliefs about the consequences of performing the behavior). Values correspond to the individual's valuation of these consequences. Note that the combination defining the attitude A toward behavior can be viewed as a factorial index of beliefs (expected success, b_i) and values v_i:

$$A = \sum_i b_i v_i \tag{2.1}$$

If more than one behavior is possible, EVT predicts that the chosen action will be the one with the largest attitudinal index. Since the evaluative nature of attitudes in EVT is rather utilitarian, one of the criticisms of this theory has been the lack of affect as a predictor of attitudes.

The Theory of Reasoned Action (TRA; **?**) holds that the best predictor of behavior is intention. According to TRA, behavioral intention is the immediate antecedent of behavior and corresponds to the cognitive representation of the strength of an individual's willingness to perform a given action. In TRA (see Figure **??**), this intention is not only determined by the attitudes toward the specific behavior (which is constructed following an EVT model based on behavioral beliefs) but also by subjective norms (a result of normative beliefs). Subjective norm is the combination of beliefs (perceived expectations) from the individual's social network (i.e. beliefs about how people that are relevant to the individual will view the behavior in question) along with intentions to fulfill these expectations.

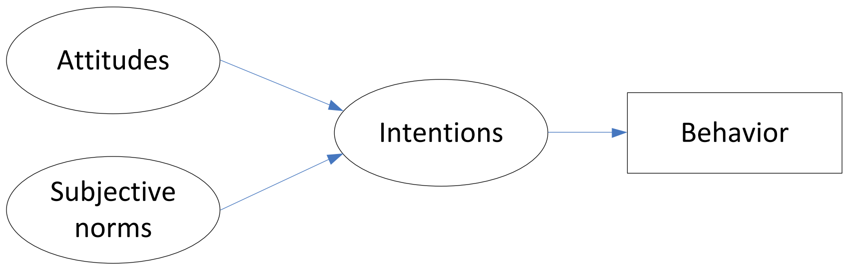

Figure 2.3: Theory of Reasoned Action

Behavioral intentions explain actions according to TRA; yet consider how many people have problems accomplishing New Year's resolutions. The problem may be that many resolutions may not in fact be under one's full volitional control, and this may explain the gap between intentions and behavior. An extension of TRA, the theory of Planned Behavior (TPB; **?**) states that attitudes and subjective norms may be insufficient to explain volitional behavior if the individual's control over the behavior is incomplete. TPB adds to TRA the concept of perceived behavioral control (see Figure **??**), i.e. individual perceptions regarding the ability to perform a given action. Perceived behavioral control has a direct impact on actual behavior, as well as an indirect impact through behavioral intentions (**?**). Note that perceived behavioral control may reflect actual behavioral control, i.e. internal control factors such as skills and abilities, as well as external control factors, such as environmental constraints in resources such as time and money.

2.3 Structural equation modeling of latent variables: a mathematical representation of attitudes

Latent variables provide a key concept in the statistical modeling of attitudes. While recognizing the existence of several different formal definitions of latent variables, **?** provides the following conceptualization: "a latent random (or nonrandom) variable is a random (or nonrandom) variable for which there is no sample realization for at least some observations in a given sample." Since there is no sample realization, a latent variable can be described as a factor that cannot (properly) be quantified in practice. Note that

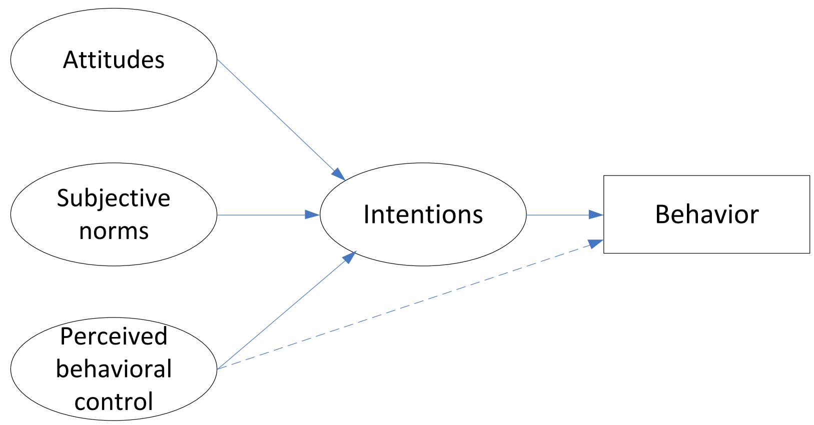

Figure 2.4: Theory of Planned Behavior

a latent variable can be unobservable either by nature (hypothetical constructs) or by practical reasons (a true variable measured with error). For instance, attitudes are by definition not directly observable; likewise, total demand for a theatrical performance cannot be measured from the total number of people that attended the show if they filled the capacity of the theater.

? also makes a distinction in the degree of abstractness of a latent variable. Abstractness is related to the existence of a measurement scale. Attitudes are abstract constructs, just as social class and mental state are. Other less abstract latent variables can be subject to measurement, but, for example, the measurement of income can be indirect because of misclassification errors. Beyond their use as a tool for overcoming the measurement problem, latent variables are also used to model unobserved heterogeneity, missing values, and latent responses (discrete choice), as well as to generate flexible distributions, and to combine information from different sources (for a discussion, see **?**). Finally, a latent variable can be endogenous (dependent) or exogenous (independent).

When a latent variable, whether a dependent or independent variable, is used, standard regression techniques cannot be applied. Although latent variables are not observable, they can be manifested through overt variables that serve to identify the concepts underlying them. These variables are called *effect indicators* or *manifest* variables. Latent variable models describe the relationship between the latent variable and (observable) manifest or indicator variables. Such models are often used as a means of dimension reduction, where the latent variable represents an underlying concept explaining the group of manifest variables. Dimension reduction is technically desirable when the number of manifest variables is large, or when there arise problems such as multicollinearity (provoked by direct incorporation of highly correlated manifest variables as explanatory variables in a regression).

Structural equation models view the relationship of attitude-to-indicators as a system of simultaneous equations[3]. Two main sub-models can be distinguished in structural equation modeling (SEM): first is the *structural model* describing potential causal relations between endogenous and exogenous variables; second is the measurement model specifying the relations of latent variables explaining their manifest variables. Note that the SEM terminology is somewhat misleading, because the full structural equation model of latent variables is a simultaneous system of structural equations, meaning that the measurement model is also a structural equation.

The JKW model (**???**), also known as LISREL[4], provides a general linear representation of the system of structural equations for both the structural and measurement models. Following the LISREL parametrization, the structural equation is of the form:

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{2.2}$$

where η is a (latent) vector of endogenous variables, ξ is a (latent) vector of exogenous variables, B is a coefficient matrix describing simultaneity in the endogenous variables, Γ is a coefficient matrix measuring the causal effects of ξ on η, and ζ is an error vector.

If both $\eta(y)$ and $\xi(x)$ are observable with no measurement error, then we obtain the structural model that has been the foundation of econometrics as a separate field. In econometrics, the structural model can represent a single-equation economic problem or a simultaneous system of structural equations, such as the equilibrium model of supply and demand that clears the market. General estimation techniques include (generalized) least squares and maximum likelihood. Note that the notation in SEM is somewhat different from ordinary regressions because in SEM the parameters are grouped in a matrix, whereas the exogenous variables are written as vectors. In SEM the resulting covariance structure is important, and the matrix notation of parameters facilitates the derivation of the covariance of the reduced form.

As discussed above, when a latent variable is used, we also need manifest variables that identify what is unobservable. The relationship between the observed variables and the

[3] See **?** for a complete review of SEM applied to travel behavior modeling.

[4] Estimation of SEM was eased due to the LISREL (Linear Structural RELationships) program developed by **?**, **?**, and **?**.

latent variables is operationalized through the measurement equations

$$y = \Lambda_y \eta + \varepsilon$$
$$x = \Lambda_x \xi + \delta$$

where y and x are manifest variables of the latent variables η and ξ, respectively. Λ_y and Λ_x are coefficient matrices; and ε and δ represent measurement errors. The measurement model alone is a key tool for psychometrics, a field concerned with measurement of abstract psychological concepts (such as attitudes, intelligence, and personality traits). Factor analysis (a technique that dates back to **?**) is the estimation method commonly used, since this method seeks to explain variability in the manifest variables in terms of fewer underlying dimensions (the latent variables).

The simultaneous SEM system considers the general model where the structural and measurement equations are combined. This is necessary when either the endogenous or the exogenous (or both) variables in the structural model are unobservable. Maximum likelihood and (generalized) least squares are the statistical methods used most often for estimation of the SEM parameters. Note that if the indicators perfectly measure the latent variables, then we have $\eta = y$ and $\xi = x$, and we obtain the classical econometric problem. Figure **??** depicts the general SEM system using a path diagram. Path analy-

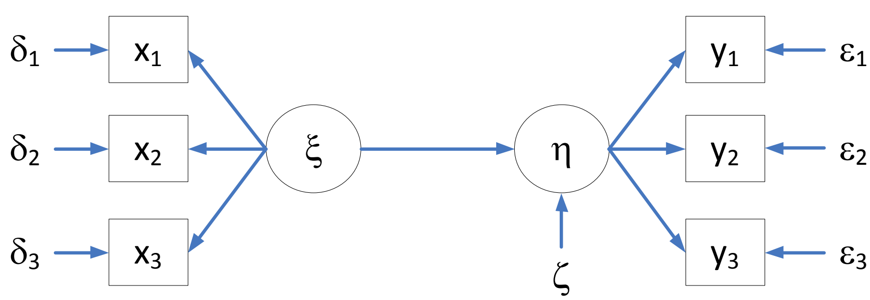

Figure 2.5: LISREL model

sis was created by **??** as a useful way to represent variable dependencies in a system of structural equations. In a path diagram, boxes represent manifest variables, an ellipse stands for a latent variable, disturbance terms are represented unenclosed, and a straight arrow between variables indicates causality. In current applications path diagrams are mostly viewed as a convenient way of representing potentially complex relationships, especially when the underlying concepts are highly abstract. However, Wright's pioneering path analysis was fundamental for the development of structural equation modeling in the 1970s[5].

Figure **??** presents the path diagram for a special case of the SEM system known as the Multiple Indicator Multiple Cause (MIMIC) model (**?**). In a MIMIC model the endogenous latent variable is explained by an observed exogenous variable. Using the LISREL notation, the MIMIC model is a sub-model of the SEM system, where the presence of a causal indicator[6] leads to

$$\begin{aligned} \eta &= B\eta + \Gamma\xi + \zeta \\ y &= \Lambda_y \eta + \varepsilon \\ x &= \xi \end{aligned} \tag{2.3}$$

where the particularity of the model is the vector x of causal indicators, as opposed to the vector y of standard manifest variables (effect indicator). Causal indicators permit us to have a nonrandom exogenous variable in the structural equation. Causal indicators can be guaranteed to exist when it is possible to establish a causal relationship where x explains η. Even though the MIMIC model was originally formulated as a single latent variable problem, it was quickly generalized to multiple latent variables. **?** is one of the earliest examples of a MIMIC model with multiple latent variables. In this work, the latent variables are propensities of individuals to participate in four different modes of politics (voting, campaigning, participating in community activities, and contacting local and national officials personally), the effect-indicators of these propensities are individual actions (such as voting in previous presidential elections, and number of

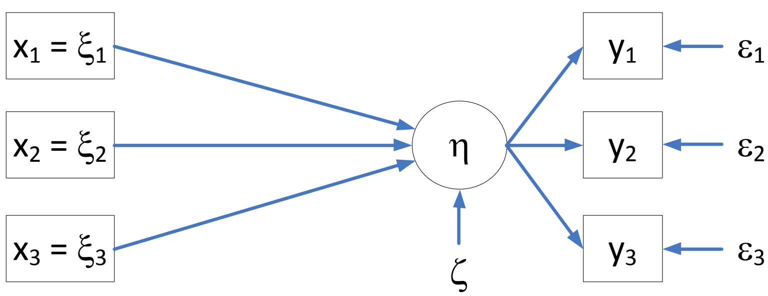

Figure 2.6: MIMIC model

[5]As noted by **?**, Wright was able to derive a complete analysis of covariance using path analysis rather than through matrix algebra. A clear example is Wright's analysis of the system of supply and demand equations, which translated into clear conditions for identification and estimation, far before econometricians derived identification conditions.

[6]A causal indicator measures the latent variable without error, **?**.

active memberships in community problem-solving organizations), and the causes are background characteristics of the individuals (such as socioeconomic status, age, and gender). Working with the same data, **?** presented a slightly different version of the model, analyzing the identification of the parameters and applying a simultaneous estimation approach. Note that the direct integration of the causal indicators implies a direct assessment of potentially heterogeneous latent responses (attitudes) to a particular concept (attitudinal object). In this example, different propensities to participate can be ascertained for different groups according to socioeconomic status, age, or gender (and combinations of these variables).

2.3.1 Causal indicators of attitudes

If the latent variables are attitudes, natural causal indicators that account for heterogeneity are the characteristics of the individuals. For instance, just as in the example of political participation, attitudes toward recycling may present differences across groups that can be tested using a MIMIC model.

However, socio-demographic variables are not always the only possible causal indicators, which is especially the case of latent constructs representing variables with measurement errors. To give an early example, in **?** the authors model the latent market value of a property using different home-value measures as manifest variables (effect indicators such as an appraised value given by a private firm, and the estimate provided by the owner). As causal indicators, the authors use property characteristics such as effective age, number of storeys, number of rooms, and finished area. However, note that the latent variable in Robins and West's study is more than an attitude, but rather expresses a qualitative attribute. I will return to this point later.

In any case, recall that a causal indicator in the MIMIC model represents an exogenous variable that helps to explain the latent variables. For example, current recycling habits are a result of someone's attitudes toward recycling, but the habits do not themselves cause the attitude. (In other words, recycling habits are manifest variables, and they are not causal indicators.) Conversely, some studies based on gender socialization theories establish that women have stronger environmental concerns than men (**??**), and this result favors the inclusion of gender as a causal indicator of recycling attitudes. To avoid violating causality, one must exercise caution when defining the explanatory variables of the structural model. Ideally, the structural model should be based on a specific theory.

2.3.2 Measuring attitudes: obtaining manifest variables

Attitudes reflect an evaluative process that is not directly observable. Thus, when we loosely talk about methods for measuring attitudes, what we mean is the measurement of an overt expression of an attitude, which translates directly into manifest or indicator variables. On the one hand if this overt expression is a verbal self-reported response, then we are making use of direct measurement procedures. Indirect measurement procedures on the other hand are built using non-verbal overt expressions, typically observed behavior such as physical reactions. In both cases, it is clear that attempting to measure attitudes translates into assessing possible manifest variables, which by definition are the overt expression of the attitudes.

In a single-item measure the individual is asked to directly report the attitude toward an attitudinal object, typically on a structured scale. However, the answer provided is a verbalization of the attitude and not the underlying attitude in the mind of the respondent. This verbalization is subject to conscious or unconscious bias; for instance an individual who is neutral toward recycling may declare a favorable attitude because of the perception that being green is a better response to give (social desirability bias). In my view, it is very important to understand the distinction between the true attitude and the verbalization of it, because the latter is merely a manifest variable of the former. Considering stated attitudes as effect indicators is consistent with the theoretical foundations of attitudes, because what one can measure are manifest variables. Note that a self-reported statement about the attitude in question is possible depending on the abstractness of the latent construct. Clearly, one can ask people about their attitude toward recycling, but the task becomes more complex if one asks about a general environmental concern concept. In addition, it is important to note that perceptual indicators may be wrongly interpreted as attitudes causing perception, whereas perceptions explain the formation of attitudes. A perceptual indicator, such as a self-declared level of satisfaction, is merely a statement that elicits the individual's beliefs. Attitudes toward an entity affect this elicited belief.

Another problem with single-item measures is that they presume the hypothesis that the attitude-indicator relationship is unidimensional, i.e. only one indicator provides enough information to identify the underlying attitude. To resolve this problem, we can combine different single-item measures for multiple manifest variables that reflect the complexity of the attitude under study. For instance, attitudes toward recycling can be manifested not only through direct verbalization but also through both the degree of support for other environmental policies and personal statements about current recycling and other eco-friendly habits. This example makes it clear that it is not only verbalization (of both the attitude under study and other related attitudes) which provides identification of the true attitude, but also stated (self-reported) and overt actions (indirectly observed) which may serve as manifest variables. The combination of single-item measures involves collecting information about different aspects of the underlying concepts that will allow us to identify how this concept is built; put in a simple example, if we need to identify a person in a group, the more information about the person's characteristics the better. Continuing the same example, note that even though we can construct a person's profile with enough information to detect the individual we are looking for, this profile is hardly complete: it is quite likely that the set of manifest variables will not provide a full measurement of the latent variable.

Typical single-item measurement scales of attitudes include semantic scales, such as Likert items which measure either positive or negative responses to a particular statement, typically according to a 5 or 7-level scale (see example in Figure **??**). These measures must be constructed with care to avoid potential bias (See **?**). Direct multiple-item measures include the Thurstone equal-appearing interval method, semantic differentials, and Likert summated ratings, i.e. the Likert scale, which is the sum of several Likert items.

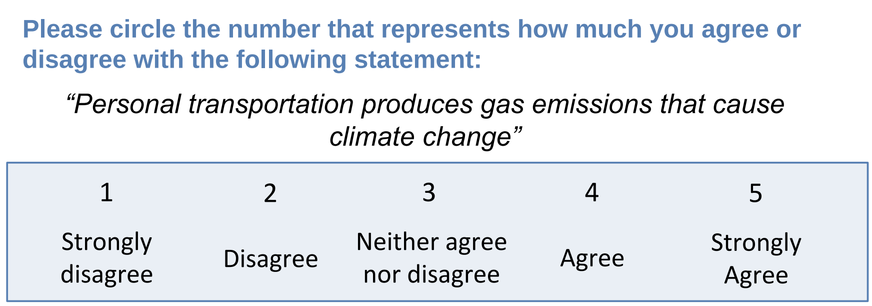

Figure 2.7: Likert item example

2.4 Economic Preference Models

According to microeconomic theory, consumer preferences aid in understanding purchase behavior and can be described as an artificial summary relation that serves to characterize consumer tastes, i.e. likes and dislikes (**?**). This definition is a foundational axiom of microeconomics and is formally expressed through mathematical assumptions of completeness and transitivity (preorder).

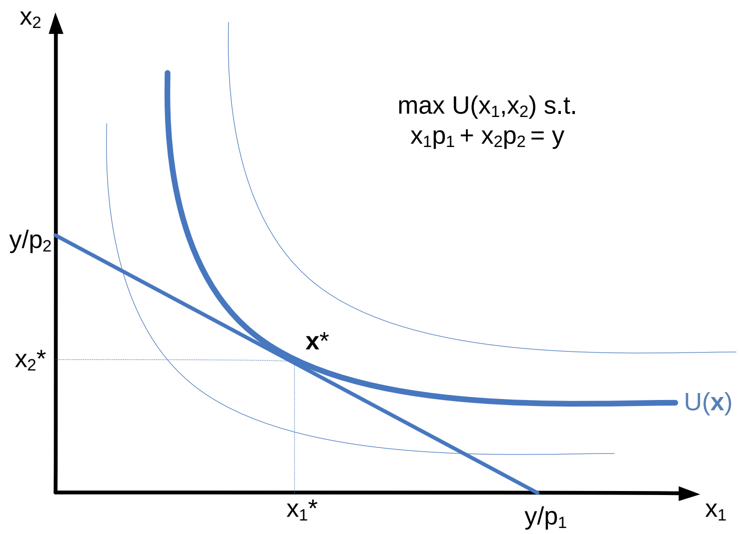

Figure 2.8: Preference maximization and optimal choice

Preferences serve to explain choice, because the preorder assumptions over preference relations allow consumers to rank different consumption bundles (subsets of the consumption set). Ordinarily, however, economists model economic decisions by means of a *utility function*. Even though the concept of utility has a long tradition in economics (for instance, utility is the central concept in utilitarianism), in modern economics utility is conceptualized as a function that summarizes the information conveyed by the consumer's preference relation. Mathematically, existence of a continuous real-valued utility function representing the preference relation is guaranteed if the latter is complete, transitive, continuous, and strictly monotonic. Whereas continuity provides regularity conditions, the assumption of strict monotonicity implies local nonsatiation and is directly related with the economic idea of 'more is better'. If the assumption of strict convexity of preferences is added, then the utility function is continuous, strictly increasing, and strictly quasiconcave. Convex consumer tastes imply that balanced combinations of goods are preferred to extreme combinations, i.e. strict convexity translates into the economic principle of diminishing marginal rates of substitution. The continuous, increasing and quasiconcave utility function implies the usual indifference mapping that is depicted in economic problems. In fact, with these properties, it is possible to represent the consumer's problem as a formal utility-maximization problem, where behavior depends on the set that maximizes utility subject to a budget constraint (Figure **??**): in the context of purchase behavior, actions have an economic price.

As we can see – even though the economic formulation is mathematically stricter –, the entire concept of economic preferences is closely related to the concept of attitudes reviewed in section 2.2. In effect, both preferences and attitudes appear as hypothetical constructs explaining (choice) behavior. The conceptualization of the utility function as an index representing preferences through a combination of quantities or attributes and tastes appears akin to the TRA attitudinal index of beliefs and values. In fact, both indexes are maximized to predict behavior. In addition, note that the budget constraint in the consumer's problem works as a behavioral control function.

Thus, economic preferences and attitudes are related. In fact, economic preferences appear as a special case of attitudes explaining purchase behavior and economic choice in general. However, as noted by Becker (2000), economists ordinarily take preferences as given, whereas other social sciences focus their attention on the origin, structure, and change of preferences/attitudes. In addition, economic preferences can be described as a special sort of attitude dominated by the assumption of rationality in neoclassical economics[7], and the behavioral assumption behind consumer choice is utility maximization.

In the next section I describe the bases of a particular type of consumer preference model, known as discrete choice, which is especially suitable for economically representing choice behavior in a context of alternative selection.

2.4.1 Standard discrete choice models

Discrete choice models aim to explain the process of individual choice among a mutually exclusive, exhaustive and finite group of alternatives (**?**). The choice of a specific brand

[7]Rationality comes from the preference relation being a total preorder. In addition, the attitude object is a choice set of goods and services, and the behavior we try to explain is the choice of a consumption bundle or plan.

of detergent at the supermarket, personal choice of entering or not entering the labor market, a particular household's heating-fuel choice, and individual travel mode choice[8]: these all are examples of discrete choice. According to consumer theory the decision process reflects preferences set by utility-maximization behavior. In the case of standard consumption theory, the utility function representing the preference relation depends on the continuous quantities in the consumption set. However, when the nature of a specific good is discrete, we adopt a hedonic approach and the preference relation is assumed to depend on a group of *attributes* (**?**)[9] combined according to individual tastes. A decision is then made based on the alternative that has the highest level of satisfaction. For example, how much one is satisfied with using his or her car for a specific trip depends on the personal valuation of how much money one spends to use it, how much time one travels in it, how hard it is to park at the destination, among other attributes. In a modal choice context, one might use his or her car but this is choice is made among a group of countable available alternatives (*car*, *bus*, *subway*, *tram*, and *walk*). The decision process is characterized by the utility (expected satisfaction) trade-off according to the attribute values of each alternative, namely cost, travel time, waiting time and walking time. One chooses the alternative that is the most convenient according to one's individual tastes.

Suppose that the whole individual consumer problem is described by a separable consumption set where it is possible to identify n continuous goods $x \in \mathbb{R}^n_+$ and one discrete choice[10]. The discrete choice is characterized by a choice set C with J polytomous different alternatives that the individual considers or perceives, and a group Q_i containing the attribute levels that describe each alternative $i \in C$. Because of the assumption of a mutually exclusive process for the discrete choice, the utility level is conditional on the discrete choice $i \in C$ (represented by its attribute values). Essentially, we can formulate the consumer problem in this case as

$$\max_{x \in \mathbb{R}^n_+, i \in C} U(x_1, x_2, \ldots, x_n, Q_i) \text{ s.t. } \mathbf{p} \cdot \mathbf{x} + c_i \leq y, \tag{2.4}$$

where $\mathbf{p}$ represents the price vector of the continuous goods, c_i is the price of the discrete good i, and y represents income.

This problem can be solved in two stages: one problem associated with the continuous portion of the problem, and another problem characterized by the discrete choice. First

[8]Even though discrete choice models are used in multiple disciplines, research has been led by the analysis of disaggregate behavioral travel demand. This tradition emerged through the work of Domencich and McFadden in the 1970s (See **?**).

[9]Based on ideas introduced by **?**.

[10]Separability of the continuous and discrete goods implies that the satisfaction derived from the consumption of the continuous goods does not depend on the qualitative attributes of the discrete good.

we solve the continuous-goods portion and as a result we obtain the indirect utility conditional on the discrete choice.

$$\max_{x \in \mathbb{R}^n_+} U(x_1, x_2, \ldots, x_n, Q_i) \text{ s.t. } \mathbf{p} \cdot \mathbf{x} \leq y - c_i$$
$$\Rightarrow \mathbf{x}^* = \mathbf{x}(\mathbf{p}, y - c_i, Q_i) \Rightarrow (\mathbf{p}, y - c_i) = U(\mathbf{x}^*, Q_i).$$

Since the continuous problem is solved conditional on the discrete choice, the resulting indirect utility can be evaluated for each discrete alternative. Because of the discrete nature of this second stage, we solve the discrete portion of the problem simply by comparing each conditional indirect utility and choosing the one that achieves a maximum value.

$$\max_{i \in C} V(\mathbf{p}, y - c_i, Q_i) \equiv i^* \{V(\mathbf{p}, y - c_1, Q_1), \ldots, V(\mathbf{p}, y - c_J, Q_J)\}.$$

When comparing the conditional indirect utility functions, only the terms that are alternative-specific survive. The price vector of the continuous goods and the total income are common to all of the conditional utility functions and are not relevant in determining the maximum. The utility of the discrete alternatives is built as a truncated conditional indirect utility function, where we keep only the attributes whose levels vary between alternatives as well as cost. In discrete choice modeling, the truncated conditional indirect utility is simply called the utility of a certain alternative. To incorporate heterogeneity among individuals, empirically it is usual to consider socio-demographic characteristics of the decision-maker. If we assume a linear-in-parameters specification, then V_{in} – which is the utility of alternative i and individual n – can be written as a function of a vector of taste parameters β and the attributes X_{in} – enclosing the alternative-specific attributes Q_i, the cost of choosing the alternative c_i, characteristics of the individual S_n and an alternative-specific constant (ASC):

$$V_{in} = V(c_i, Q_i, S_n) = X_{in}\beta.$$

In discrete choice modeling, the most common approach is based on **random utility theory** (**?**), which introduces the concept of individual choice behavior being intrinsically probabilistic[11]. Whereas the Random Utility Model (RUM) framework recognizes the existence of a systematic component of individual behavior, RUM also takes into account the incapacity of the analyst to observe all the variables that influence the decision (incomplete information that entails uncertainty). **?** identifies four sources of

[11]Random utility theory has its roots in psychology. The idea was introduced by **?**, who built on a model of imperfect discrimination; later **?** linked the concept of random utility with the theory of individual choice behavior developed by **?**.

uncertainty: unobserved alternative attributes, unobserved individual attributes or random *taste variations*, measurement errors (including incorrect perception of attributes), and proxy or instrumental variables. Therefore, utility is modeled as a random variable, consisting of an observable systematic and deterministic component V_{in}, and an unobservable random component ε_{in}:

$$U_{in} = V_{in} + \varepsilon_{in}.$$

Different discrete choice RUMs can be derived based on various assumptions on the distribution of the random term ε_{in}. The probabilistic nature of the choice behavior implied by the RUM framework leads to the individual probabilities of each consumer selecting each available alternative:

$$\mathbb{P}_n(i) = \mathbb{P}(i|C_n) = \mathbb{P}(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) = \mathbb{P}(V_{in} - V_{jn} \geq \varepsilon jn - \varepsilon in, \forall j \in C_n, j \neq i).$$

In 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 mixed logit model (**???**), which offers a flexible covariance structure together with the possibility of approximating any random utility model (**?**). Research in discrete choice modeling in the last two decades has devoted an enormous effort toward understanding the flexibility of the distribution of the error term and developing estimation methods that account for this flexibility.

The link between attitudes and economic preferences expressed through a utility function becomes clearer when we analyze willingness to pay (WTP) measures, which are a usual output of preference models and, in particular, of discrete choice models. Economic WTP reflects how much money a consumer is ready to pay for a good - or for an increase of a desirable attribute in the case of discrete choice models. For example, consider a vehicle purchase situation. Usually low-emission vehicles are pricier, but some consumers are willing to pay this higher price because they want a vehicle that produces less GHG emissions. The WTP of those consumers reflects not only an economic preference but also an environmental preference that can be understood as an (economic) attitude related to the consumer's environmental concerns.

2.4.2 Discrete choice econometrics as SEM

In econometrics, the problem of latent endogenous variables has led to specific models of qualitative dependent variables, including discrete (or qualitative) choice. Note that under the RUM framework, standard discrete choice is a special case of SEM. The (truncated indirect) utility function is a latent construct that measures the individual level of satisfaction conditional on each alternative according to a structural equation derived from microeconomic principles. The structural relationship involves (observable) attributes of the alternatives and socioeconomic characteristics of the individuals, both as exogenous variables. Although the utility function is unobservable, revealed or stated choices serve as indicators of the underlying choice process (Figure **??**):

$$
\begin{aligned}
U &= X\beta + \nu \\
y &= \textit{choice}
\end{aligned}
$$

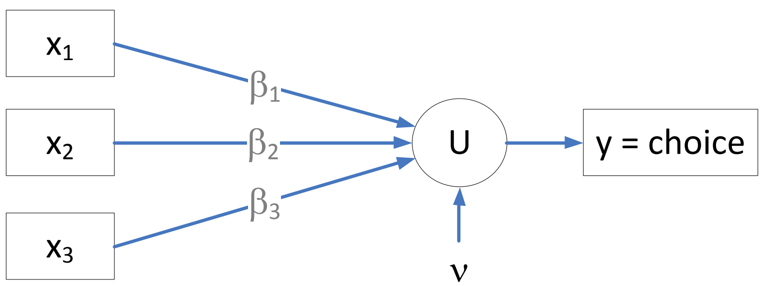

Figure 2.9: Standard discrete choice

Not only does the structural equation representing economic preferences through a utility function make discrete choice a special case of SEM, but also the specific nature of choice as an effect indicator of the preferences. In effect, choice (representing the label of the alternative chosen after utility maximization) is a nominal scale for the discrete response resulting from maximizing the underlying continuous utility function. Thus, the alternatives represent a qualitative concept that is represented by the hedonic attributes. Note further that the discrete response represents a measure of utility differences: we cannot observe the utility of an individual, but after an overt choice we know that the utility is higher for the chosen alternative, i.e. the utility difference with the non-chosen alternatives is positive[12]. Observing choice data for a group of individuals (or repeated data for a single individual in a dynamic choice situation) we can map the utility function, yielding the estimation of the unknown parameters of the structural equation.

[12]Considering the absolute value of each utility.

In a revealed preference (RP) study real choices are observed, but in a stated preference (SP) experiment (also termed conjoint analysis), respondents are faced with hypothetical choice situations. Note that SP choices can be understood as an elicitation of behavioral intentions (or an attitude toward purchase), whereas RP data correspond to manifest variables of current or past behavior.

2.4.3 Extending the discrete choice framework

According to Daniel Kahneman, there still remains a significant difference between economists who develop practical models of decision-making and behavioral scientists who focus on a cognitive understanding of agent behavior beyond utility maximization. Both groups have fundamental interests in behavior but each works with different assumptions and tools. **??** 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 economic models of decision making.

Effectively, under the standard random utility approach **?** discrete choice models represent the decision process as an obscure black box, where **attitudes**, **perceptions** and **knowledge** are neglected (**??**). This is a serious problem if we consider that, as discussed above, psychometric studies have proven that attitudes affect behavior. From an econometric point of view, since we expect attitudes to have a causal impact on economic choice (based on TPB, for instance), omitting attitudinal factors leads to omitting a relevant variable, leading to concomitant endogeneity problems.

The black box described by **?** is sketched in Figure **??**. The idea of the black box is that, even though the study of economic preferences has a solid mathematical basis, the axioms explaining economic consumer behavior do not explain the cognitive process involved in terms of how tastes are formed and how perceptions and attitudes affect preferences.

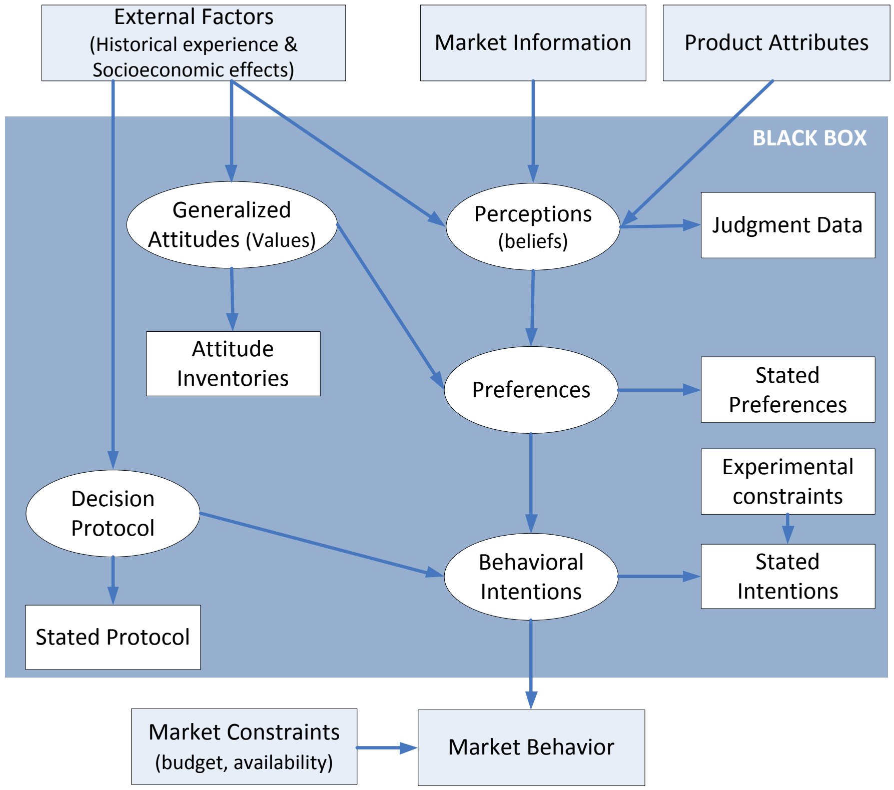

Figure 2.10: The black box as described by **?**

Another representation of the choice process, sketched by McFadden and Ben-Akiva in the late 1990s (see **?**), is presented in Figure **??**. Here not only perceptions, beliefs and attitudes are fundamental in the process, but the role of information, experience, and affect are relevant too. In this model, which is not precisely a path diagram, the thicker arrows correspond to the standard economic preference model of choice, where information is processed as perceived attributes and then a utility maximization cognitive process results in choice. The lighter arrows represent the relevant relationships according to behavioral science. Note that in this model it is clear that the interaction between perceptions, attitudes, and preferences depends on whether we use an economic preference approach or a theory based on psychology. In fact, as McFadden points out, the standard economic model appears as a simplified form of the more general model that takes into account the whole system.

Because attitudes affect the cognition of the choice process (**?**), to generate a more comprehensive economic representation of the decision-making task, it is essential to consider behavioral science research on the links among attitudes, perceptions, behavioral

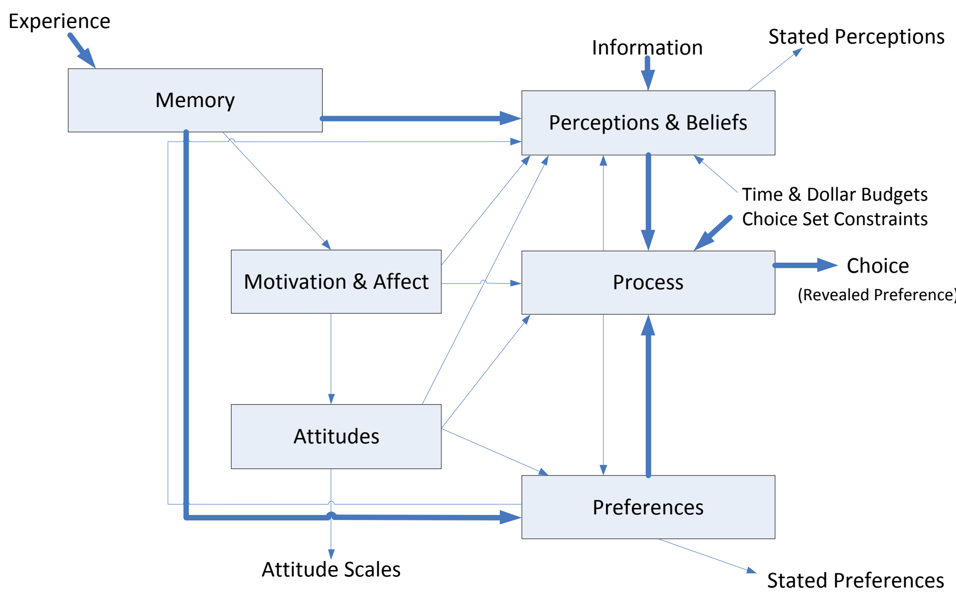

Figure 2.11: Representation of the choice process (?)

intentions, and actual behavior[13]. In economic preference models, choice behavior is explained through the cognitive process of utility maximization. We will see that the effect of attitudes on behavior can be indirectly incorporated through the attitude effects on the utility function.

2.5 Hybrid Choice Modeling

Hybrid choice models are a generalization of standard discrete choice models where different expanded models are considered simultaneously (see ?). A hybrid choice model (HCM) expands on discrete choice modeling by combining the following important modeling extensions (?): heterogeneity through flexible error structures (such as the use of mixed logit); the combination of revealed (RP) and stated preference (SP) data; the presence of latent classes explaining underlying market segments (through a latent class model); and the integration of latent (unobserved) constructs according to an Integrated Choice and Latent Variable (ICLV) model (Figure ??). It is the ICLV model inside the HCM conceptual framework which permits the inclusion of latent attitudinal constructs

[13]For a review of empirical research on the inclusion of attitudes in travel behavior modeling, see ?.

in such a way that understanding of consumer behavior is improved while the model gains in predictive power.

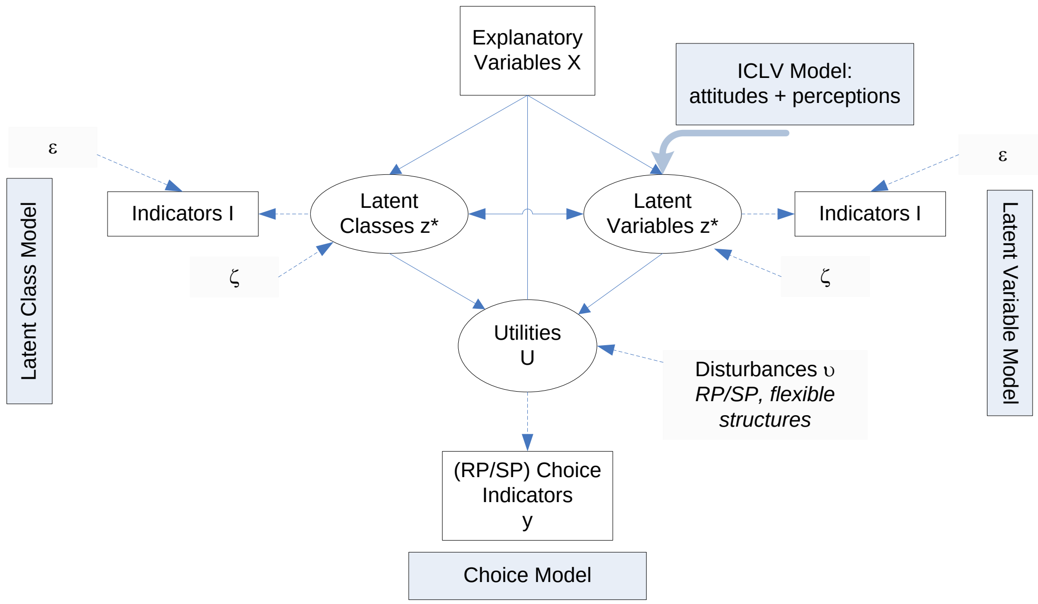

Figure 2.12: Hybrid Choice Model as sketched by **?**

Research on the inclusion of attitudes into discrete choice models started in the late 1970s with the work of **?**, who considered manifest variables directly in the utility function. The next step was to conduct factor analysis to get fitted attitudinal factor scores that are introduced as explanatory variables of the discrete choice (**??**). However, this sequential estimation approach results in estimates that are not only inefficient but also inconsistent. Although **?** and **?** propose different methods for achieving consistency and efficiency in the two-step estimator (**?**), the required procedure is complex and often neglected in empirical work. **?** and **?** set the theoretical fundamentals for later development of a comprehensive framework of hybrid choice models and ICLV simultaneous estimation, where the whole model is viewed as a system of equations involving a standard SEM system and discrete choice. These fundamentals were revisited by Ben-Akiva, Walker and Bolduc in a seminal work (**??**) that has motivated the reemergence of ICLV models as an important research subject in discrete choice modeling.

Methodologically, the modeling challenge in hybrid choice modeling arises in simultaneous estimation of the ICLV model and the consideration of flexible disturbances. In fact, in recent literature the terms hybrid choice and hybrid discrete choice have both been used to describe a joint ICLV model.

Combining the SEM and DCM notations, the hybrid choice model can be written as

$$\begin{aligned} U &= X\beta + \Gamma z + \nu \\ z &= \Pi z + Bw + \zeta \\ y &= choice, \\ I &= \Lambda_z z + \varepsilon \\ x &= \xi_x \\ w &= \xi_w \end{aligned} \tag{2.5}$$

where U and z are latent endogenous variables (although z also enters as an explanatory variable in the structural equation of U). X is a matrix of attribute characteristics and socioeconomic variables; the last equation, which is usually omitted, indicates that this matrix is observable. w is a (latent) vector of exogenous variables. y is a choice indicator and I is a vector of effect-indicators. β, Γ, Π, B, and Λ_z are unknown parameters; ν, ζ, and ε are disturbance terms. A schematic path diagram is presented in Figure **??**.

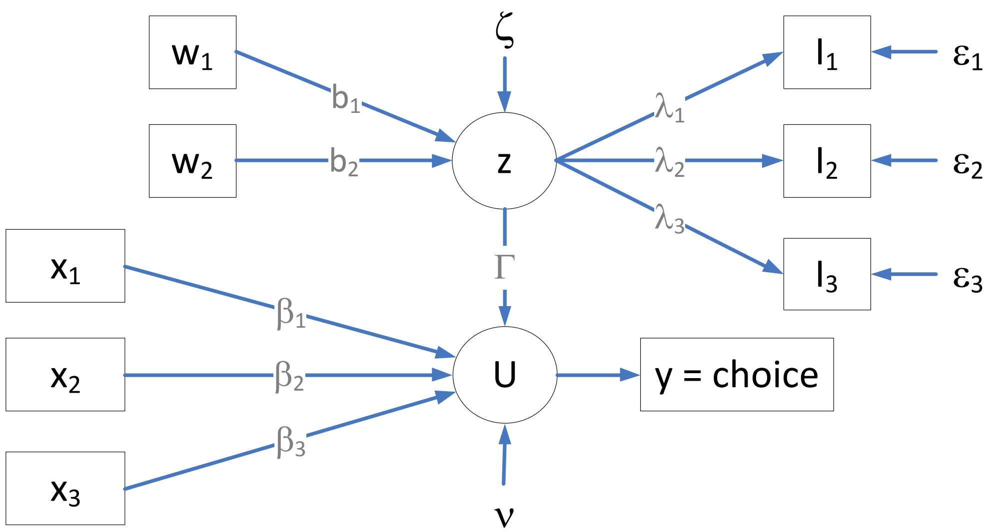

Figure 2.13: Hybrid Choice Model

When we look at the equations describing the hybrid choice model, we can clearly distinguish a particular extension of the SEM system. In a more general form, the hybrid model is nothing other than an SEM where the exogenous[14] latent variables has been

[14]Exogenous to the utility function.

endogenized, and hence there is an additional structural equation that is specific to what I will call a causal latent variable. However, the hybrid model is a particular instance of this generalized SEM because the dependent latent variable (the 'original' endogenous latent variable) represents a utilitarian function of economic preferences and the indicator is relative to choice behavior. In this sense, the label ICLV is somewhat redundant: as noted above, the standard discrete choice model is already an integration of latent variables and economic choice. In fact, the standard discrete choice model can be interpreted somewhat as the MIMIC version of the hybrid choice model. Whereas the name "hybrid choice model" gives the idea of a model that consolidates different portions or sub-models that can be plugged in to gain flexibility, the name itself is not particularly self-explanatory. As we have seen, **the hybrid choice model is a discrete choice model with endogenous latent causal variables**.

What do these endogenous latent causal variables represent? Consider first qualitative attributes. Qualitative attributes, as opposed to quantitative attributes, do not have a natural order or an overt measurement scale. Because quality is hard to conceptualize, it is also hard to measure (owing, for example, to the multidimensional nature of quality), and therefore quality needs to be operationalized as a latent construct.

The most naive approach toward qualitative attributes is to simply ignore them. However, this is equivalent to assuming incorrectly that consumers do not consider quality when evaluating alternatives before purchase. In fact, assuming that quality is a relevant attribute in explaining consuming behavior, the omission of quality will cause potentially severe econometric problems. In the best case, if the omitted relevant variable is independent of the other attributes in the model, then the estimated parameters will be unbiased (although the standard errors will be invalid). It is not difficult, however, to argue that quality and price are closely related. The omission of quality means that part of this variable's effect goes to the error term, but because quality and price are correlated, then price will be correlated with the error term. This leads to the econometric problem of endogeneity which provokes biased parameters.

A less naive approach uses a proxy variable for the qualitative construct. Effectively, qualitative attributes are often introduced as categorical variables on a nominal scale. The nominal scale may be adequate if the qualitative attribute is discrete in nature. But if there is some continuity in the evaluation of quality, the nominal scale becomes a proxy variable that measures the true qualitative attribute with error.

A typical example of the proxy variable approach can be seen in stated preferences (or conjoint analysis) studies in which respondents are faced with alternatives described as having 'poor', 'standard', or 'high' quality. Represented through these discrete categories, quality is inserted into the utility function using dummy variables or effects coding (representing the nominal scale), or even by means of an arbitrary continuous scale. If we replace quality for a proxy variable, without taking into account the measurement error involved, then we again have endogeneity. In the SEM terms, the proxy variables correspond to the effect-indicators or manifest variables of the latent construct. Taking SEM as modeling framework, we know that we do not insert the indicators directly into the structural equation, but we have a measurement model that allows us to identify and hence correctly integrate the latent variable inside the structural equation. For instance, income is very likely to be misspecified. Using hybrid choice modeling, the solution is simple: we include in the choice model a latent income variable to account for the measurement errors associated with reported income classes (which can be taken as effect-indicators of income).

Attitudes present a second kind of endogenous causal latent variable. The theory of planned behavior asserts that attitudes cause action through their role in explaining behavioral intentions[15]. In hybrid choice modeling, the structural relationship between choice behavior and attitudes is given by the presence of attitudes as latent causal variables in the utility function. In an HCM attitudes play a role in the choice process through a corresponding marginal utility. For instance, a consumer with positive attitudes toward sustainability is ready to pay more for organic produce because food that is farmed following socially responsible practices provides more satisfaction to the consumer.

Note that the attitudes (or qualitative attributes) that we incorporate in an HCM are endogenous to the economic choice model (the structural equation of the utility function), but we manage to treat this endogeneity by means of the structural and measurement equations of the latent variable. Since the simultaneous system of equation yields efficient and consistent estimators, HCMs appear as a tool for dealing with endogeneity in discrete choice models.

Other methods for treating endogeneity have also been discussed (**?**). In particular, the control function method (**?**) has recently been attracting the attention of DCM researchers. **?** compare the control function method with a specific HCM where fitted errors of an IV regression for price are used as effect-indicators. The authors conclude

[15] In addition, TPB provides a theoretical framework that allows us to consider attitudes as relevant variables inside a structural equation.

that the HCM performs better in terms of consistency. My discussion here, however, extends beyond instrumental variables, because the latent variable approach allows us to explicitly incorporate the variable that, when omitted, provokes endogeneity problems.

Several examples of empirical applications of the HCM framework are found in the literature[16], most of them using a two-step estimator. In the last couple of years, a vivid interest has arisen in developing a generalized efficient and consistent simultaneous estimator. Effectively, although the idea behind the ICLV model is not new, recent research on simultaneous estimation of the HCM has renewed interest in the integration of psychometric models in the field of discrete choice microeconometrics. **?** present the first example of such an analysis of a general situation characterized by a large number of latent variables and a large number of choices[17]. A recent application of the Hybrid Choice setting applied to the freight sector appears in **?**. For a personal-vehicle-technology choice, **?**[18] analyze the practical use of a large number of indicators.

2.5.1 HCM as an attitudinal model of choice

The HCM system of equations introduces attitudes in different dimensions of the choice process. Interpreting the model from an attitudinal point of view, we model the consumer's problem as a special case of planned behavior. For instance, in the context of purchase behavior, we can distinguish purchase intentions from actual purchase behavior. Purchase intentions represent general behavioral-utilitarian attitudes toward the purchase of a specific good. These utilitarian or economic attitudes can be understood as the economic preference valuation summarized by a utility function. Purchase intentions reflect a consumer's desires as well as the evaluative process prior to actual purchase. Of course, purchase intentions are also affected by a behavioral control function, specifically by the budget constraint. Since economic preferences are unobservable, this utility function is treated as a latent variable. The economic preferences that underlie purchase intentions are manifested through self-reported stated choices.

In practice, purchase intentions are measured in *stated preference* (SP) experiments using conjoint analysis. The economic preferences (summarized in the utility function), the

[16]However, in the vast literature on discrete choice, the inclusion of attitudinal factors is rather sporadic.

[17]This paper is also the first attempt to develop a Bayesian estimator. However, the authors provide only the outline of a Gibbs sampler which is not ready for full implementation.

[18]See chapters 3 and 4.

budget constraint, and the stated choices conjointly determine purchase intentions. Once the taste parameters are determined, it is possible to forecast the consumer's response in terms of purchase intentions beyond the experimental values of the attributes considered in the SP choice situations. Even though purchase intentions do not necessarily reflect actual purchase behavior, measuring purchase intentions is especially important when introducing new products. Moreover, according to the theory of planned behavior, intentions should serve to explain actual behavior. In a discrete choice framework, actual behavior is also determined by the individual utility function. Actual purchase behavior is measured by actual choice (current or previous purchase behavior.) In practice, this measurement is done using revealed preference (RP) studies which take choice as a behavioral manifest variable.

In its current state of development, the HCM handles the effect of behavioral intentions on actual behavior through joint estimation of the utility function mixing RP and SP data. In practice, joint RP-SP yields more accurate predictions of purchase behavior, basically because the experimental trade-offs of the SP choice situations can be used to identify the taste parameters that explain revealed choice. However, research on the causal effect of intentions on actual choice needs to be pursued further.

Additionally, the HCM allows us to include latent variables as attributes of the alternatives. These latent variables may represent qualitative attributes, but may also represent other attitudinal dimensions related to consumer behavior, beyond the behavioral dimension provided by the utility function. Effectively, we can introduce into the choice process affective and cognitive attitudes. In the HCM these attitudinal dimensions indirectly affect behavior through their impact on the behavioral dimension determined by the utility function.

For example, in the case of a *purchase of low-emission vehicles*, the cognitive attitude related to a *positive evaluation of low-emission vehicles as a means to alleviate oil dependency* may have an effect on *readiness to buy an enviro-friendly car*, and this disposition is the behavioral attitude determined by the utility function. We can also include behavioral attitudes associated with other behavioral contexts. For example, the *willingness to buy an iPad* (as determined by the utility function in this particular purchase context) can be affected by the behavioral attitude determined by *promptness to adopt new technologies* in general. For these explanatory latent variables (Figure **??**), the HCM considers both causal and effect indicators that must be introduced into the choice process simultaneously with the preference model for consumer behavior. In sum, the HCM is an effective tool for incorporating the economic grounds of decision making into well-

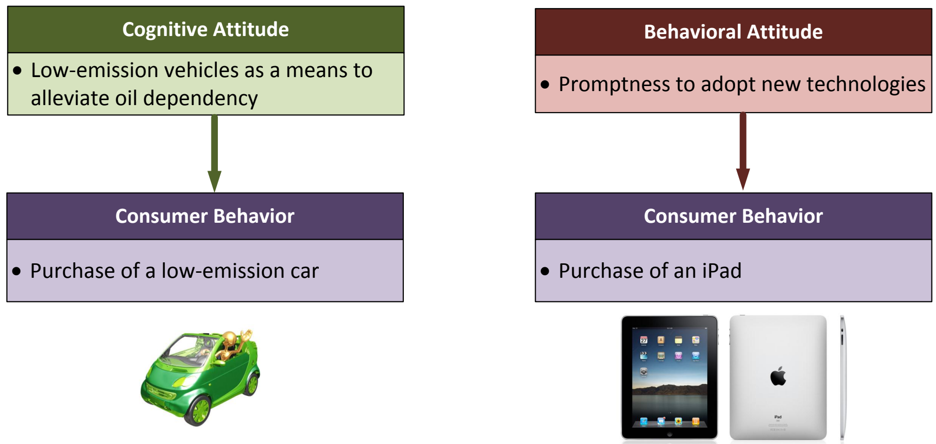

Figure 2.14: Integrating attitudes toward new technologies

established psychometric models of attitude-behavior. However, additional research is needed to explore other HCM expansions such as: the inclusion of subjective norms (impact on demand by reference groups); the integration of cognitive theories explaining the formation of, and stability and change in, attitudes; the impact of information, knowledge and habits on both stated intentions and actual actions; and potential semiotic effects (**?**) on the dynamics of choice behavior.

2.5.2 Extended Examples of HCMs

To clarify the concepts covered in this chapter I will discuss two examples of potential latent explanatory variables. Consider first voice-service reliability of mobile telephony. Reliability is a qualitative variable that is intangible. A naive approach will take an indicator variable as a proxy for the original concept of quality. For example, an SP survey could define a 3-level proxy of reliability as 'low number of dropped calls', 'medium number of dropped calls' and 'high number of dropped calls'. Clearly, if reliability is a relevant variable in the choice of mobile telephony provider and we take the proxy as an explanatory variable we will have measurement error problems such as endogeneity. According to SEM, we recognize the latent nature of reliability and assume that it can be indirectly observed (manifested) through other variables such as the 'average number of dropped calls', the stated 'perceived reliability of the service', and the 'strength of the signal measured by the number of bars', among other indicators[19]. At the same time, we take as granted that the reliability of the cell phone service depends on variables such as the 'volume of network traffic', 'weather conditions' and the 'cell site network density in the service area'[20]. Although we do not observe the latent reliability, the simultaneous system of SEM equations allows us to estimate the unknown parameters to explain the latent construct.

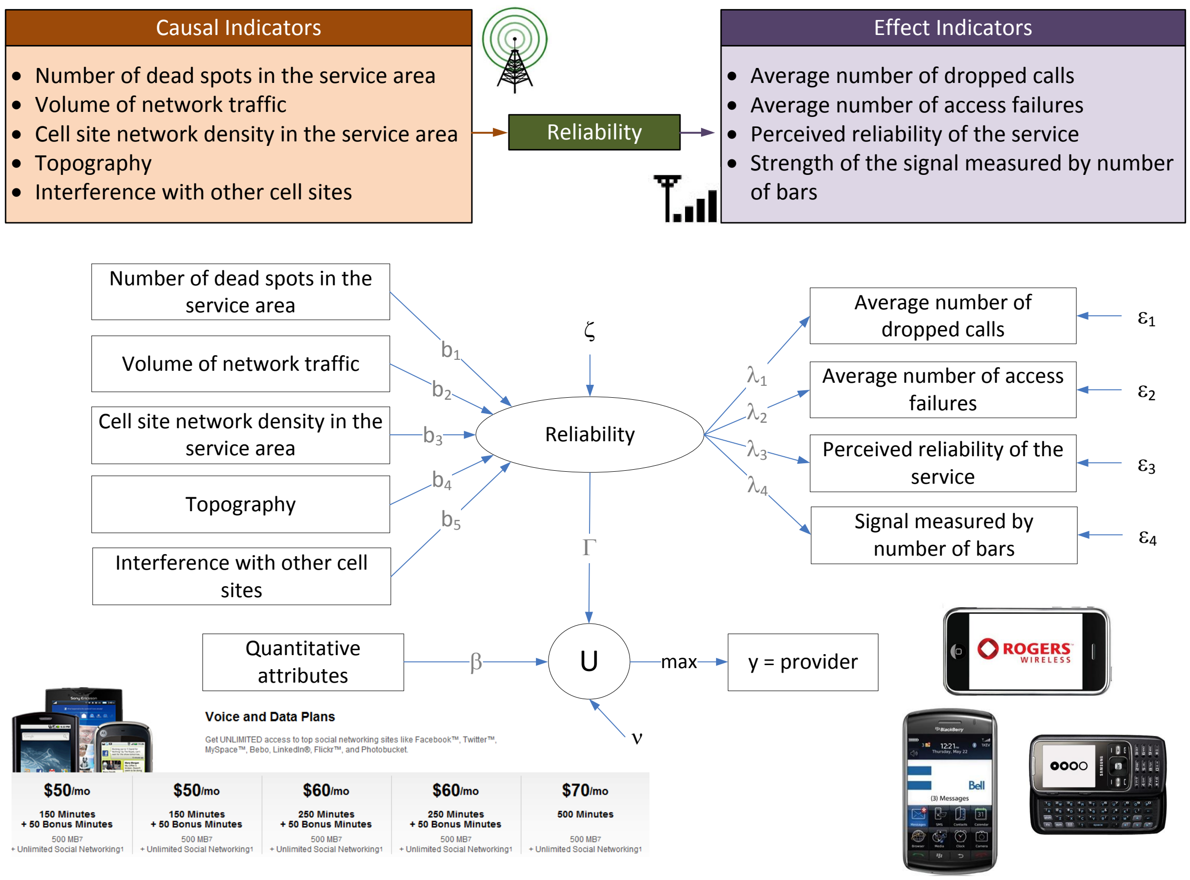

Figure 2.15: Example: Hybrid Choice Model including reliability of mobile telephony

In an HCM for the decision of mobile telephony provider we consider two problems simultaneously (Figure ??). The first is the estimation of the utility function that describes the individual economic choice process. The second is the estimation of the equations

[19]The relationship between reliability and its indicators is represented through measurement equations.

[20]The causal relationship between reliability and its explanatory variables is represented through a structural equation.

that are related to reliability, which we assume as a relevant explanatory variable of the utility function in the first problem.

It is important to understand the role of each equation of the SEM system defining the HCM. Continuing with the mobile telephony example, the utility function is a latent variable that is manifested through choice; the latter in turn indicates that the utility function is greater for the chosen provider than for the rest of the alternative providers[21]. As mentioned above, we have assumed that reliability enters the utility function as a qualitative explanatory variable and as such it is recognized as a latent variable. Thus, reliability is modeled using the SEM combination of the structural and measurement equations outlined previously. But since reliability enters into the utility function, there is also an interaction between the SEM describing the consumer's preference and the SEM describing the qualitative reliability. All the equations need to be considered simultaneously in order to estimate the unknown parameters.

Although it is clear how the system works for estimation, it is also relevant to discuss how the system works for forecasting. If a provider decides to make an effort to improve the reliability of the company, then the consumer faces a different scenario than the one used for estimation of the parameters. Assuming that tastes are stable, the company's efforts toward improved reliability will positively affect the likelihood of this particular provider being chosen. Even though for improved reliability the indicators will change (i.e. the number of dropped calls will fall and the customers should report an improved level of satisfaction), the manifest variables for a future situation are not known a priori. This is not a problem, however, because the current indicators served only for identification of the latent variable: once the parameters have been estimated, we need only the structural model to make predictions. Effectively, if the provider decides to expand their network coverage aiming at improving the reliability of their services, then by using the structural model for reliability of the HCM, we can ascertain the impact of such an expansion on the reliability of the provider, and then obtain the change in the market shares of each provider that this expansion provokes.

The discussion above is valid for latent variables as qualitative attributes. When modeling attitudes the causal and effect indicators may however have a different nature. For example, when choosing a cell phone plan, the consumer's intentions and actual purchase of a smartphone can be explained by his or her behavioral attitude toward adopting new technologies (Figure **??**). Someone who has a more favorable view of new technologies will be more likely to have a smartphone. When incorporating the latent promptness to

[21] This is a direct result of utility maximization behavior.

adopt new technologies in an HCM, we need to identify both causal and effect indicators. In this example, the adoption of new technologies may present differences across socio-demographic groups that can be tested using MIMIC models.

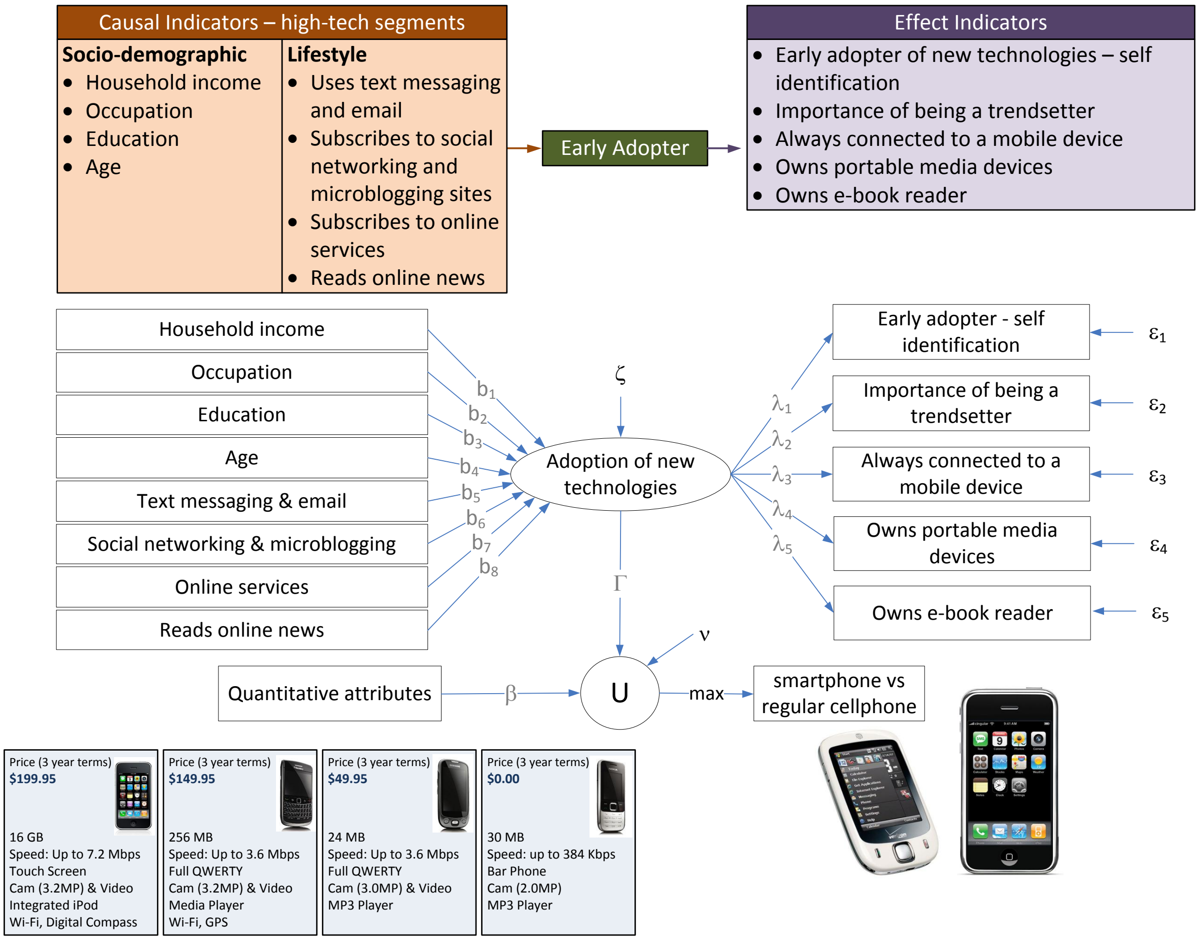

Figure 2.16: Example: Latent disposition to adopt new technologies

In effect, causal indicators such as social class, age, and education translate the structural equation of the latent variable into a model for market segmentation of new technology adoption. For instance, a young professional may be more likely to adopt new technologies than a retired worker. Or, lifestyle characteristics may also serve as causal indicators. For example, someone who texts and emails, and who belongs to a social networking or microblogging website, may be more inclined to demand a smartphone. Effect indicators for the adoption of new technologies can be self-reported responses, such as 'self-perception of the consumer as an early adopter of new technologies', or the 'importance of being a trendsetter', as well as previous or current purchase behavior, such as 'currently using e-mail' and 'Internet browsing in a non-smartphone', 'owning portable media devices', 'using a laptop', or 'owning an e-book reader'.

Thus, using an HCM for this particular choice context not only are we able to understand the consumer's preferences in terms of, for example, willingness to pay to have access to a smartphone, but we can also assess how much more likely someone is to choose a smartphone on a high-tech segment, where the different segments are simultaneously obtained.

As can be inferred from this example, usual manifest variables for attitudes are perceptual-attitudinal (opinions) or behavioral (overt actions) indicators. Additionally, natural causal indicators that account for heterogeneity are characteristics of individuals. Effectively, empirical HCM work that employs a structural model of attitudes makes use of explanatory variables are basically sociodemographic (for example **?**). Conversely, in HCMs with qualitative attributes (**??**) the structural equation makes use of both socioeconomic variables and observable attributes.

Two generalizations can be made from both of the telecom examples. When introducing latent variables, on the one hand we have qualitative attributes, for which causal indicators can be alternative-specific. For example, in a travel mode HCM **?** use age, terminal time (rail-auto), number of transfers by rail, and availability of free parking for auto as exogenous variables causing a latent convenience attribute. On the other hand, we have attitudes, for which causal indicators are often individual-specific. For example, in a travel mode HCM **?** use gender, age, and the presence of children in the household as causal variables to explain general attitudes toward the environment.) Hence, it is of fundamental importance to determine the causality relationship necessary for building the structural and measurement equations.

These distinctions are important in forecasting, because by incorporating observable attributes in the structural model of the latent attribute we are able to predict the effects on choice after an expected change in the latent attribute; the discussion on forecasting with HCMs is also outlined by **?**. For example, in a travel mode choice example a change in the transportation system that reduces the number of transfers by rail entails greater convenience for rail and hence rail becomes more attractive and more likely to be chosen. Likewise, frequency can explain comfort inasmuch as low frequency of high-demand buses entails a loss in comfort because the buses are crowded. So in an HCM context, a policy aiming to improve frequencies can affect choice not only through a direct effect but also through the effect of frequency on the structural equation of comfort. The role of certain other variables, such as passenger density, is less obvious, especially when there is simultaneity between the qualitative attribute and the variable being analyzed: passenger density might be a factor explaining comfort, or it might instead be a manifestation of this qualitative variable.

2.6 Summary and concluding remarks

Throughout this chapter I have discussed the importance of developing a more comprehensive economic model that takes into account the potentially complex relationships among perceptions, attitudes, economic choice behavioral intentions, and overt behavior. Hybrid choice models (HCMs) are a generalized structural equation model (SEM) system that simultaneously accommodates a discrete choice model with latent explanatory variables, where these variables in turn enter the system as a standard SEM.

As discussed earlier, there are several reasons why it is desirable to include latent variables as explanatory variables in standard economic preference models. Explanatory latent variables can represent variables that are difficult to measure or that do not possess an associated measurement scale. A variable that can be measured (but with difficulty) may entail working with measurement errors. Variables that are unobservable by nature need a special treatment. Two relevant types of latent explanatory variables of consumers' preferences appear: qualitative attributes and attitudes.

Note that whereas the study of the attitude-behavior relationship is fundamental in social psychology and cognitive science research, in economics the impact of attitudes as explanatory variables of consumer behavior has mostly been neglected, yielding econometric problems arising from the omission of a relevant variable.

In econometrics, latent variables are modeled using structural equation models (SEMs) that represent the unobserved construct as a system of equations. First, the structural equation links the endogenous latent variables[22] to exogenous (latent or observed) variables[23] according to a causal relation. Second, the measurement equation provides identification of the latent variable through manifest variables or effect-indicators[24] (See Figure **??**).

[22] I.e. qualitative attributes or attitudes, or both.

[23] I.e. explanatory variables of the qualitative attributes or attitudes according to some general model.

[24] Variables that are dictated by the level of the qualitative attribute or attitude.

Attitudes		Quality	
Causal Indicators	**Effect Indicators**	**Causal Indicators**	**Effect Indicators**
• Sociodemographic • Lifestyle	• Perceptual • Behavioral	• Sociodemographic • Alternative-specific attributes	• Perceptual • Alternative-specific attributes

Figure 2.17: Causal and effect indicators for attitudinal and qualitative attributes

Finally, because attitudes together with qualitative attributes motivate the development of an integrated model of economic preferences and latent variables, the HCM is ultimately a more realistic econometric model. Because the HCM integrates the effect of attitudes on consumer behavior, the economic choice model of an HCM is only a part of the whole behavioral process incorporates individual attitudes, opinions and perceptions. This chapter has made clear the need to incorporate attitudinal data into discrete choice models. Even though some effort has been made in this direction, several econometric challenges remain to be addressed. In the following chapters I analyze econometric estimation of HCMs, using both classical and Bayesian techniques.

References

Integrated Framework for Travel Behavior Analysis, number 17 in IATBR Conference, Aix-en-Provence, France, 1987.

I. Ajzen. From intentions to actions: A theory of planned behavior. In J. Kuhl and J. Beckmann, editors, *Action-control: From cognition to behavior?* Springer, Heidelberg, 1985.

I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211, 1991.

I. Ajzen and M. Fishbein. *Understanding attitudes and predicting social behavior.* Prentice-Hall, Englewood Cliffs, NJ, 1980.

I. Ajzen and M. Fishbein. Attitudes and the attitude-behavior relation: reasoned and automatic processes. In W. Stroebe and M. Hewstone, editors, *European Review of Social Psychology.* Wiley, Chichester, England, 2000.

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

M. Ben-Akiva, D. Bolduc, and J. Walker. Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. Working Paper, MIT, 2001.

M. Ben-Akiva, 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): 163–175, 2002.

M. Ben-Akiva, D. Bolduc, and J.Q. Park. Discrete choice analysis of shipper's preferences. In E. Van de Voorde, editor, *Recent Developments in Transport Modeling: Lessons for the Freight Sector*. Emerald Publisher, 2008.

A.T. Bernardino. *Telecommuting: Modeling the Employer's and the Employee's Decision-Making Process*. Garland Publishing, New York, NY, 1996.

R.D. Blackwell, P.W. Miniard, and J.F. Engel. *Consumer behavior*. Harcourt College, Forth Worth, TX, 2001.

W. Bohte, K. Maat, and B. van Wee. Measuring attitudes in research on residential self- selection and travel behaviour: A review of theories and empirical research. *Transport Reviews*, 29(3):325–357, 2009.

D. Bolduc and R. Alvarez-Daziano. On estimation of hybrid choice models. In S. Hess and A. Daly, editors, *Choice Modelling: the state-of-the-art and the state-of-practice, Proceedings from the Inaugural International Choice Modelling Conference*. Emerald, England, 2010.

D. Bolduc and M. Ben-Akiva. A multinominal probit formulation for large choice sets. Technical report, Proceedings of the 6th International Conference on Travel Behaviour 2, 1991.

D. Bolduc, M. Ben-Akiva, J. Walker, and A. Michaud. Hybrid choice models with logit kernel: applicability to large scale models. In M. Lee-Gosselin and S. Doherty, editors, *Integrated Land-Use and Transportation Models: Behavioral Foundations*, pages 275–302. Elsevier, New York, 2005.

K.A. Bollen. *Structural equations with latent variables*. John Wiley and Sons, Chichester, 1989.

K.A. Bollen. Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.*, 53: 605–634, 2002.

D. Brownstone and K.E. Train. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89:109–129, 1999.

G. Debreu. Review of R.D. Luce individual choice behavior. *American Economic Review*, 50: 186–188, 1960.

A.H. Eagly and S. Chaiken. *The psychology of attitudes*. Harcourt Brace Jovanovich, Fort Worth, TX, 1993.

L.R. Fabrigar, J.A. Krosnick, and B.L. MacDougall. Attitude measurement. In T.C. Brock and M.C. Green, editors, *Persuasion: Psychological Insights and Perspectives*. Sage, Thousand Oaks, CA, 2005.

M. Fishbein. An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations*, 16:233–240, 1963.

M. Fishbein and I. Ajzen. Attitudes toward objects as predictos of single and multiple behavioural criteria. *Psychological Review*, 81(1):59–74, 1974.

M. Fishbein and I. Ajzen. Misconceptions about the Fishbein model: reflection on a study by songer-nocks. *Journal of Experimental Social Psychology*, 12(6):579–584, 1977.

T.F. Golob. Structural equation modeling for travel behavior research. Technical report, UC Irvine: Center for Activity Systems Analysis, 2001.

Z. Griliches. Hedonic price indices for automobiles: An econometric analysis of quality change. Price statistics of the federal government, National Bureau of Economic Research, 1961.

C.A. Guevara and M. Ben-Akiva. Addressing endogeneity in discrete choice models: Assessing control-function and latent- variable methods. Working Paper, Massachusetts Institute of Technology, 2009.

R.R. Heffner, K.S. Kurani, and T.S. Turrentine. Symbolism in California's early market for hybrid electric vehicles. *Transportation Research D: Transport and Environment*, 12(6):396–413, 2007.

L.M. Hunter, A. Hatch, and A. Johnson. Cross-national gender variation in environmental behaviors. *Social Science Quarterly*, 85:677–694, 2004.

M. Vredin Johansson, T. Heldt, and P. Johansson. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40:507–525, 2006.

K.G. Jöreskog. A general method for analysis of covariance structures. *Biometrika*, 57:239–251, 1970.

K.G. Jöreskog. A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan, editors, *Structural Models in the Social Sciences*. Academic Press, New York, NY, 1973.

K.G. Jöreskog and A.S. Goldberger. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351): 631–639, 1975.

K.G. Jöreskog and D. Sörbom. *Advances in Factor Analysis and Structural Equation Models*. Abt Books, Cambridge, MA, 1979.

K.G. Jöreskog, G. T. Gruvaeus, and M. van Thillo. ACOVS: A general computer program for analysis of covariance structures. Technical report, Educational Testing Services, Princeton, NJ, 1972.

J.W. Keesling. Maximum likelihood approaches to causal analysis. Ph.D Thesis, University of Chicago, 1972.

T.C. Kinnear and J.R. Taylor. *Marketing reasearch: an applied approach*. McGraw-Hill, New York, NY, 1996.

F. Koppelman and J. Hauser. Destination choice behavior for non-grocery-shopping trips. *Transportation Research Record*, 673:157–165, 1978.

F. Koppelman and E. Pas. Travel choice behavior: Models of perceptions, feelings, preference and choice. *Transportation Research Record*, 765:26–33, 1980.

K. Lancaster. A new approach to consumer theory. *Journal of Political Economy*, 74:132–157, 1966.

R.T. LaPiere. Attitude vs. actions. *Social Forces*, 13:230–237, 1934.

J. Louviere, K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T. Cameron, C. Carson, J. Deshazo, D. Fiebig, W. Greene, D. Hensher, and D. Waldman. Recent progress on endogeneity in choice modeling. *Marketing Letters*, 16(3):225–265, 2005.

R.D. Luce. *Individual choice behavior*. Wiley, New York, NY, 1959.

C. Manski. The structure of random utility models. *Theory and Decision*, 8:229–254, 1977.

J. Marschak. Binary choice constraints on random utility indicators. In K. Arrow, editor, *Stanford Symposium on Mathematical Methods in the Social Sciences*. Stanford Univ. Press, Stanford, 1960.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontier in Econometrics*. Academic Press, New York, 1974.

D. McFadden. The choice theory approach to market research. *Marketing Science*, 5(4):275–297, 1986. Special issue on consumer choice models.

D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57:995–1026, 1989.

D. McFadden. Economic choices. Nobel Prize Lecture, 2000.

D. McFadden and K.E. Train. Mixed MNL models of discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

T. Morikawa. Incorporating stated preference data in travel demand analysis. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1989.

T. Morikawa, M. Ben-Akiva, and D. McFadden. Incorporating psychometric data in econometric choice models. Working Paper, Massachusetts Institute of Technology, 1996.

A. Petrin and K. Train. Omitted product attributes in discrete choice models. Working Paper, National Bureau of Economic Research, 2005.

A. Polydoropoulou. Modeling user response to advanced traveler information systems (ATIS). Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1997.

J.A. Prashker. Mode choice models with perceived reliability measures. *Transportation Engineering Journal*, 105(TE3):251–262, 1979.

P.K. Robins and R.W. West. Measurement errors in the estimation of home value. *Journal of the American Statistical Association*, 72:290–294, 1977.

M.J. Rosenberg and C.I. Hovland. Cognitive, affective, and behavioral components of attitudes. In C.I. Hovland and M.J. Rosenberg, editors, *Attitude organization and change: an analysis of consistency among attitude components*. Yale University Press, New Haven, CT, 1960.

L.G. Schiffman and L.L. Kanuk. *Consumer behaviour*. Pearson Education, Upper Saddle River, NY, 2001.

A. Skrondal and S. Rabe-Hesketh. *Generalised latent variable modelling: multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC, Boca Raton, Florida, 2004.

M.B. Smith. The personal setting of public opinions: a study of attitudes toward Russia. *Public Opinion Quarterly*, 11:507–523, 1947.

C. Spearman. General intelligence, objectively determired and measured. *American Journal of Psychology*, 15:201–293, 1904.

D. C. Stapleton. Analyzing political participation data with a MIMIC model. *Sociological Methodology*, 9:52–74, 1978.

L. Thurstone. A law of comparative judgment. *Psychological Review*, 34:273–286, 1927.

L.L. Thurstone. The measurement of social attitudes. *Journal of Abnorm. Soc. Psychol.*, 26: 249–269, 1931.

K. Train, D. McFadden, and A. Goett. The incorporation of attitudes in econometric models of consumer choice. Working paper, Cambridge Systematics, 1986.

S. Verba and N. Nie. *Participation in America: Political Democracy and Social Equality*. Harper & Row, New York, NY, 1979.

J. Walker. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. PhD dissertation, MIT, Massachusetts, 2001.

J. Walker and M. Ben-Akiva. Generalized random utility model. *Mathematical Social Sciences*, 43(3):303–343, 2002.

A.W. Wicker. Attitude versus actions: the relationship of verbal and over behavioural responses to attitude objects. *Journal of Social Issues*, 24:41–77, 1969.

D.E. Wiley. The identification problem for structural equation models with unmeasured variables. In A.S. Goldberger and O.D. Duncan, editors, *Structural Models in the Social Sciences*. Academic Press, New York, NY, 1973.

T.D. Wilson, S. Lindsey, and T.Y. Scholler. A model of dual attitudes. *Psychology Review*, 107:101–126, 2000.

S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–585, 1921.

S. Wright. The method of path coefficients. *Annals of Mathematical Studies*, 5:161–215, 1934.

M.F. Yáñez, P. Mansilla, and J. de D. Ortúzar. The Santiago panel: measuring the effects of implementing Transantiago. *Transportation*, 2009. In press.

L. Zelezny. Elaborating on gender differences in environmentalism. *Journal of Social Issues*, 56 (3):443–458, 2000.

Chapter 3

On classical estimation of hybrid choice models

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 mixed logit model. This search for flexibility has continuously faced the problem of a more involved and extremely demanding estimation process: the curse of dimensionality. In the last few years the flexibility search has been extended to the next level, integrating attitudinal data into standard choice models through hybrid choice models (HCMs). The estimation of HCMs is not absent of econometric challenges to be addressed and solved.

In this chapter I describe the classical estimation techniques for a full information 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. I also discuss the formulation and problems of a limited information maximum likelihood solution, which is the method mostly used in practice. A case study using real data where the joint estimation is applied completes the chapter.

3.1 HCM estimation: an econometric challenge

As discussed in Chapter 2, hybrid choice models (HCMs) integrate standard discrete choice modeling (DCM) and structural equation modeling (SEM), taking into account the impact of attitudes on the decision process. The econometric representation of a general HCM setting involves solving a simultaneous equation system defined by structural[1] and measurement[2] equations for both the DCM and SEM sub-models.[3]

Assuming a linear specification, the HCM system of structural and measurement equations may be written as follows:

Structural equations

$$z_n^* = \Pi z_n^* + Bw_n + \zeta_n = (\mathrm{I}_L - \Pi)^{-1} Bw_n + (\mathrm{I}_L - \Pi)^{-1}\zeta_n, \quad \zeta_n \sim N(0, \Psi) \tag{3.1}$$

$$U_n = X_n\beta + \Gamma z_n^* + \upsilon_n \tag{3.2}$$

Measurement equations

$$I_n = \alpha + \Lambda z_n^* + \varepsilon_n, \quad \varepsilon_n \sim N(0, \Theta) \tag{3.3}$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

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 $(\mathrm{I}_L - \Pi)$ 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 $(\mathrm{I}_L - \Pi)^{-1} Bw_n$ 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 (**??**) 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

[1] For unobservable dependent variables.
[2] For manifest variables.
[3] For the structural equation latent variable sub-model we use a MIMIC specification.

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 (**??**). Since latent variables are not actually observed, the choice probability is obtained by integrating the latter expression over the whole space of z_n^*:

$$P_n(i|X_n, w_n, \theta, B, \Pi, \Psi) = \int\limits_{z_n^*} P_n(i\,|z_n^*, X_n, \theta\,) g(z_n^*|w_n, B, \Pi, \Psi) dz_n^*, \tag{3.5}$$

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 (**??**).

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 (**??**) and (**??**). 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:

$$P_n(i, I|X_n, w_n, \delta) = \int\limits_{z_n^*} P_n(i\,|z_n^*, X_n, \theta\,) f(I_n|z_n^*, \Lambda, \Theta) g(z_n^*\,|w_n, B, \Pi, \Psi\,) dz_n^*, \tag{3.6}$$

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

3.2 Full information maximum likelihood

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

3.2.1 Evaluating the joint choice probability

For efficiency reasons, we only focus on a full information solution. Following **?**, HCM classical full information estimation requires the evaluation of the joint probability $P_n(i, I|X_n, w_n, \delta)$ defined in equation (**??**). 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 (**??**).

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:

$$P_n(i, I|X_n, w_n, \delta) = \int\limits_{z^*} \frac{\exp(X_{in}\beta + C_i z_n^*)}{\sum\limits_{j \in C_n} \exp(X_{jn}\beta + C_j z_n^*)} f(I_n|z_n^*, \Lambda, \Theta) g(z_n^*|w_n, B, \Pi, \Psi) dz_n^*. \quad (3.7)$$

Assuming an MNL 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 mixed logit (MMNL) kernel is the most convenient assumption to model flexible error structures. We will decompose υ_n assuming a Normal distributed factor analytic structure:

$$\upsilon_n = PT\xi_n + \nu_n, \tag{3.8}$$

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 mixed logit kernel adds an additional F-dimensional integral to the joint probability $P_n(i, I|X_n, w_n, \delta)$, which now implies solving:

$$P_n(i, I|X_n, w_n, \delta) = \int\limits_{\xi_n}\int\limits_{z_n^*} P_n(i\,|z_n^*, X_n, \theta, \xi_n\,) f(I_n|z_n^*, \Lambda, \Theta) g(z_n^*\,|w_n, B, \Pi, \Psi\,) N_\xi(0, I_F) dz_n^* d\xi_n. \tag{3.9}$$

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:

$$P_n(i\,|z_n^*, X_n, \theta, \xi_n\,) = \frac{\exp(X_{in}\beta + C_i z_n^* + P_i T\xi_n)}{\sum\limits_{j\in C_n} \exp(X_{jn}\beta + C_j z_n^* + P_j T\xi_n)}, \tag{3.10}$$

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

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:

$$I_{rn}^* = \alpha_r + \Lambda_r z_n^* + \varepsilon_{rn}, \quad \varepsilon_{rn} \sim N(0, \theta_r^2). \tag{3.11}$$

In the binary case, we rather get instead:

$$I_{rn} = \begin{cases} 1 & \text{if } I_{rn}^* \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.12}$$

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

$$I_{rn} = \begin{cases} 1 & \text{if } \gamma_0 < I_{rn}^* \leq \gamma_1 \\ 2 & \text{if } \gamma_1 < I_{rn}^* \leq \gamma_2 \\ \vdots & \\ Q & \text{if } \gamma_{Q-1} < I_{rn}^* \leq \gamma_Q, \end{cases} \tag{3.13}$$

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:

$$f(I_n) = \prod_{r=1}^{R} f(I_{rn}). \tag{3.14}$$

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

$$f(I_{rn}) = \frac{1}{\theta_r}\phi\left(\frac{I_{rn} - \alpha_r - \Lambda_r z_n^*}{\theta_r}\right), \tag{3.15}$$

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

$$f(I_{rn}) = \Phi\left(\frac{\alpha_r + \Lambda_r z_n^*}{\theta_r}\right)^{I_{rn}} \left(1 - \Phi\left(\frac{\alpha_r + \Lambda_r z_n^*}{\theta_r}\right)\right)^{(1-I_{rn})}, \tag{3.16}$$

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

$$f(I_{rn} = q) = \Phi\left(\frac{\gamma_q - \Lambda_r z_n^*}{\theta_r}\right) - \Phi\left(\frac{\gamma_{q-1} - \Lambda_r z_n^*}{\theta_r}\right) \quad . \tag{3.17}$$

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

3.2.2 Simulated maximum likelihood solution

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

$$\ell(\delta) = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i, I|X_n, w_n, \delta)^{y_{in}}, \tag{3.18}$$

which leads to the following maximum log-likelihood problem:

$$\max_{\delta} \mathcal{L}(\delta) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i, I|X_n, w_n, \delta). \tag{3.19}$$

The evaluation of the joint probability $P_n(i, I|X_n, w_n, \delta)$ is required to find the solution of the problem (**??**) $\hat{\delta} = argmax\{\mathcal{L}(\delta)\}$. The number of latent variables has an impact on the computation of this probability, since each additional latent variable adds an additional dimension to the integral. In fact, for the case of a MMNL kernel, note that equation (**??**) 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, the evaluation of the joint probability rapidly becomes intractable and simulation would be required (see **?**). 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 (**??**), we can take an empirical mean that provides a valid estimator of the true probability:

$$\tilde{P}_n(i, I\,|X_n, \delta\,) = \frac{1}{S} \sum_{s=1}^{S} P_n(i\,|z_n^*, X_n, \theta, \xi_n^s\,) f(I_n\,|z_n^{*s}, \Lambda, \Theta\,), \tag{3.20}$$

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

Calling $h(i_n, I_n|X_n, \delta, \xi_n, z_n^*) \equiv P_n(i\,|z_n^*, X_n, \theta, \xi_n\,) f(I_n\,|z_n^*, \Lambda, \Theta\,)$, the first order conditions (FOC) of the SAA problem imply calculating the derivative:

$$\begin{aligned}\frac{\partial \tilde{\mathcal{L}}(\delta)}{\partial \delta} &= \sum_{n=1}^{N}\sum_{i\in C_n}\frac{1}{\tilde{P}_n(i, I\,|X_n, w_n, \delta\,)}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial h(i_n, I_n|X_n, \delta, \xi_n^s, z_n^{*s})}{\partial \delta} \\ &= \sum_{n=1}^{N}\sum_{i\in C_n}\frac{1}{\tilde{P}_n(i, I\,|X_n, w_n, \delta\,)}\frac{1}{S}\sum_{s=1}^{S}\partial h(i_n, I_n|X_n, \delta, \xi_n^s, z_n^{*s}) \times \\ &\quad \left[\frac{\partial \ln P_n(i\,|z_n^{*s}, X_n, \theta, \xi_n^s\,)}{\partial \delta} + \frac{\partial \ln f(I_n\,|z_n^{*s}, \Lambda, \Theta)}{\partial \delta}\right].\end{aligned} \tag{3.21}$$

Numerical maximization of the simulated likelihood function requires finding analytical expressions for both derivatives:

$$\frac{\partial \ln P_n(i\,|z_n^{*}, X_n, \theta, \xi_n\,)}{\partial \delta}$$

and

$$\frac{\partial \ln f(I_n\,|z_n^{*}, \Lambda, \Theta\,)}{\partial \delta}.$$

Also, note that in the case of an MMNL kernel we have

$$\tilde{P}_n(i, I\,|X_n, w_n, \delta\,) = \frac{1}{S}\sum_{s=1}^{S}\frac{\exp(X_{in}\beta + C_i z_n^{*s} + P_i T\xi_n^s)}{\sum\limits_{j\in C_n}\exp(X_{jn}\beta + C_j z_n^{*s} + P_j T\xi_n^s)} f(I_n\,|z_n^{*s}, \Lambda, \Theta\,). \tag{3.22}$$

In the past few years, a lot of progress has been made regarding MSL estimation. **?** gives an in-depth analysis of the properties of MSL estimators. Recent results, based mainly on the analysis of mixed logit models and mostly attributable to **?**, 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 (**??**). Currently, the HCM estimation software 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 (**??**) 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 likelihood function: $\sum_{n=1}^{N}\sum_{i\in C_n} y_{in}\ln P_n(i, I\,|X_n, w_n, \delta\,)$.

3.2.3 Identification discussion

Whereas identification issues are now well understood in the context of traditional discrete choice models (**??**), 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 (**?**).

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. **?**, **?** and **?** study the identification conditions for the case of multinomial probit models, which is equivalent to the case of the mixed logit model, discussed specifically in **?** and **?**.

On the other hand, conditions for identification of some specific latent variable models can be applied. First, the measurement scale of the latent variables is unknown and hence normalization is required. The normalization can be achieved either by setting a unit variance for each latent variable[4] or by setting to 1 one nonzero coefficient in each column of the matrix Λ (see **?**). When the measurement equations are not correlated (i.e. Θ is assumed diagonal), then the matrix Ψ is identified. To complete identification of the parameters of the measurement equation of the latent variable model (equation **??**), 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]Normalizing the variance of the SEM latent variables is equivalent to the normalization of the scale of the utility function.

3.3 Simultaneous vs Sequential Estimation

3.3.1 Sequential Estimation

Because of the complexity associated with the simultaneous SML approach, most of the current applications of HCMs use sequential estimation that makes use of a two-step estimator (**???**). First is the solution of the MIMIC model; second is the estimation of the DCM parameter vector through maximum likelihood using the predicted conditional mean of the latent variables.

First, we describe a limited information maximum likelihood (LIML) solution. Note that by inserting equation **??** into equation **??**, we obtain the following regression equation[5]:

$$I_n = \alpha + \Lambda \tilde{B} w_n + \tilde{\varepsilon}_n, \quad \tilde{\varepsilon}_n \sim N(0, \Lambda \tilde{\Psi} \Lambda' + \Theta). \tag{3.23}$$

Now that we have rewritten the MIMIC model as one equation, we can estimate the unknown parameters δ_{SEM}[6] by standard maximum likelihood techniques:

$$\max_{\delta_{SEM}} \mathcal{L}(\delta_{SEM}) = -\frac{N}{2} \ln |\Lambda \tilde{\Psi} \Lambda' + \Theta| + \sum_{n=1}^{N} (I_n - \alpha - \Lambda \tilde{B} w_n)' (\Lambda \tilde{\Psi} \Lambda' + \Theta)^{-1} (I_n - \alpha - \Lambda \tilde{B} w_n). \tag{3.24}$$

Taking $\hat{\delta}_{SEM} = \arg\max\{\mathcal{L}(\delta_{SEM})\}$, we can evaluate the conditional mean $\hat{\mathbb{E}}(z_n^*|w_n, I_n) = \mathbb{E}(z_n^*|\hat{\delta}_{SEM}, w_n, I_n)$. Let δ_{DCM} be the vector of unknown parameters in the discrete choice model. Consider the conditional mean of the utility function:

$$\hat{\mathbb{E}}(U_n|I_n) = X_n \beta + \Gamma \hat{\mathbb{E}}(z_n^*|w_n, I_n). \tag{3.25}$$

This conditional mean together with the conditional variance $\mathbb{V}(z_n^*|\hat{\delta}_{SEM}, w_n, I_n)$, can be used to obtain the choice probabilities $P_n(i|I_n, X_n, w_n, \delta_{DCM}, \hat{\delta}_{SEM})$.

To find $\hat{\delta}_{DCM}$, we need to solve the maximum log-likelihood problem:

$$\max_{\delta_{DCM}} \mathcal{L}(\delta_{DCM}; \hat{\delta}_{SEM}) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i|I_n, X_n, w_n, \delta_{DCM}, \hat{\delta}_{SEM}). \tag{3.26}$$

This maximization problem is not the same as equation **??**. Even more, since we are conditioning on I_n and because $\hat{\delta}_{SEM}$ allows us to use a predictor of the latent variable, the problem reduces to a standard discrete choice model.

[5]To simplify notation, we define $\tilde{B} = (I_L - \Pi)^{-1} B$, and $\tilde{\Psi} = [(I_L - \Pi)^{-1}] \Psi [(I_L - \Pi)^{-1}]'$

[6]δ_{SEM} represents the vector of unknown parameters for the MIMIC model.

Note that the sequential solution $\delta_{seq} = (\delta_{DCM}, \delta_{SEM})$ is not the maximum likelihood estimator of the hybrid model but an approximation. In fact, the LIML solution results in consistent but inefficient estimates. Even though the asymptotic variance of the second step can be corrected, additional computation is needed (?).

In practice an even simpler approach is used, where after estimating δ_{SEM} fitted values for the latent variable $\hat{z}_n^*$ are calculated. These fitted values are taken as non-stochastic, exogenous, and observable attributes that enter the discrete choice model. Then, keeping the hypotheses of the choice model, δ_{SEM} is calculated using maximum likelihood. Despite the simplicity of this sequential approach, the resulting estimates are not only inefficient but also inconsistent. A solution to this problem is to take into account the distribution of the fitted latent variables (as we did in the LIML case) either considering numerical integration of the expectation of the choice probabilities conditional on the fitted latent variables (?), or an alternative approach that makes use of a method of simulated moments (as proposed by ?).

In sum, proper use of a sequential method for HCM estimation implies a sufficient degree of complexity that vanishes the simplicity of method's conceptualization.

3.3.2 Comparing Simultaneous and Sequential methods using Monte Carlo

We carried out a simulation experiment with the purpose of checking the empirical performance of both the simultaneous and the sequential estimation methods for hybrid choice models (?). In particular, with this controlled experiment we aim to study the effect of data variability on both estimation methods.

We consider a situation with two alternatives $i = 1, 2$ described by two attributes X_1 and X_2, for a population of $N = 50,000$ individuals. The systematic utility function was built assuming an incremental specification, linear both on the attributes and the latent variables, considering also one alternative specific constant associated with alternative 2. In addition to the attributes of the alternatives, we make the hypothesis that underlying attitudes and perceptions are incorporated into the choice process as latent variables. Although the simultaneous estimation method does not impose any restriction on the number of latent variables involved, the sequential estimation method, because of identification constraints, requires at least 2 different latent variables. Hence we consider two latent variables z_1 and z_2, both directly affecting the utility function of alternative 2. Thus, for each individual n we have:

$$U_{2n}-U_{1n} = \beta_0+\beta_1(X_{21,n}-X_{11,n})+\beta_2(X_{22,n}-X_{12,n})+\Gamma_1 z^*_{1n}+\Gamma_2 z^*_{2n}+(\upsilon_{2n}-\upsilon_{1n}). \quad (3.27)$$

The two attribute differences $(X_{21,n}-X_{11,n})$ and $(X_{22,n}-X_{12,n})$ were built taking random draws from independent truncated Normal distribution functions with arbitrary lower and upper bounds.[7] To test the impact of the variability of the data on the estimated parameters – an aspect related with empirical identification and parameter recovery (**?**), we vary the coefficient of variation (CV) of each truncated Normal distribution according to the following table:

Case	$(X_{21,n} - X_{11,n}) \sim TN_{\mathfrak{R}}(\mu, \sigma^2)$			$(X_{22,n} - X_{12,n}) \sim TN_{\mathfrak{R}}(\mu, \sigma^2)$		
	μ	σ	CV	μ	σ	CV
1	5.00	2.50	0.50	-5.00	10.00	-2.00
2	5.00	0.50	0.10	-5.00	10.00	-2.00
3	5.00	0.25	0.05	-5.00	10.00	-2.00
4	5.00	0.10	0.02	-5.00	10.00	-2.00

Table 3.1: Parametric definition of the distribution of the attribute differences

Note that the second attribute difference has a (high-variance) constant CV, whereas the first attribute difference has a decreasing low-variance CV. In fact, in our experimental databases the CV of the first attribute difference goes from a moderately low-variance ($Case$ 1, with $CV = 0.50$) to a case with an extremely low-variance ($Case$ 4, with $CV = 0.02$).

The error terms υ_{1n} and υ_{2n} are i.i.d. $Gumbel(0,1)$ leading to an MNL kernel for the discrete choice model.[8] By setting the *Gumbel* scale factor to 1, we are able to directly analyze the estimated parameters without scale concerns. By setting the scale factor to 1, the resulting percentage error in the data in $Cases$ 1 to 4 is about 20% (i.e. about one of every five simulated individuals change their choices because of the random term). This number assures that the choice process is neither completely deterministic nor completely random.

The choice model is completed with the underlying choice process reflected by the measurement equations:

$$y_{1n} = \begin{cases} 1 & \text{if } U_{1n} \geq U_{2n} \\ 0 & \text{otherwise,} \end{cases} ,\forall n \in N \; ; y_{2n} = \begin{cases} 1 & \text{if } U_{2n} \geq U_{1n} \\ 0 & \text{otherwise,} \end{cases} ,\forall n \in N \; . \quad (3.28)$$

[7]We considered a range of $\mathfrak{R} =[\ -100,\ 100\ ]$

[8]Conditional on the unobserved latent variables

We assume that each latent variable is respectively explained by two different variables w_{1n} and w_{2n}, through the structural equations:

$$\begin{aligned} z^*_{1n} &= b_1 w_{1n} + \zeta_{1n} \\ z^*_{2n} &= b_1 w_{1n} + \zeta_{2n}, \end{aligned} \tag{3.29}$$

where $\zeta_{1n} \sim N(0,1)$ and $\zeta_{2n} \sim N(0,1)$, $\forall n \in N$. In general, the variables w_n correspond to individual characteristics that usually take the form of dummy variables. To simulate this effect we consider that w_{1n} and w_{2n} are i.i.d. $Bernoulli(p = 0.5)$. Regarding the number of explanatory variables to be considered in the MIMIC model, the sequential approach requires at least as many explanatory variables as latent variables for parameter identification. In this case, working with two different individual characteristics for each latent variable is necessary for the sequential estimation method as an identification constraint that the simultaneous method does not require.

Finally, since the variables z^*_{1n} and z^*_{2n} are unobserved, we need measurement equations for the latent variable model. A minimum of two indicator variables per latent variable is required by the sequential estimation method. We thus assume that each latent variable is measured by two different continuous indicator variables, leading to 4 distinct measurement equations:

$$\begin{aligned} I_{11n} &= \lambda_{11} z^*_{1n} + \varepsilon_{1n} \\ I_{21n} &= \lambda_{21} z^*_{1n} + \varepsilon_{2n} \\ I_{32n} &= \lambda_{32} z^*_{2n} + \varepsilon_{3n} \\ I_{42n} &= \lambda_{42} z^*_{2n} + \varepsilon_{4n}, \end{aligned} \tag{3.30}$$

where $\varepsilon_{\cdot n} \sim N(0,1), \forall n \in N$.

The path diagram for the set of structural and measurement equations of the experimental HCM is sketched in Figure **??**.

To construct the estimated parameters for each coefficient of variation of the data (*cases* 1 to 4 defined in Table **??** on page **??**), we considered 15 randomly selected sub-samples of 1500 individuals from an original database of N = 50,000 individuals. In effect, the reported results correspond to the means of the estimates for the 15 repetitions of the sub-sampling process. The results of the parameter estimation using both approaches, as well as the respective t-values, are presented in Tables **??**,**??**,**??**, and **??**; the t-values against the (known) target values are also shown. The simultaneous results were obtained using a full information simulated maximum likelihood code in Fortran for hybrid choice

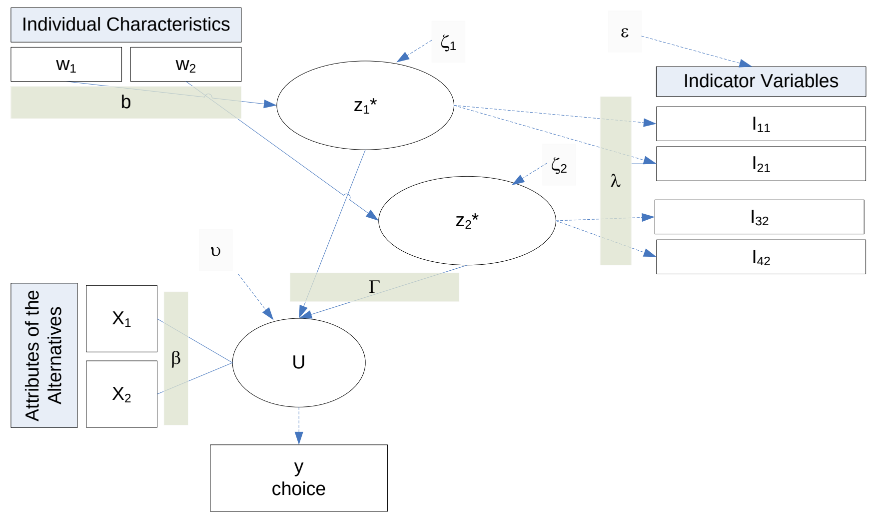

Figure 3.1: Experimental Hybrid Choice Model

models. To approximate the maximum log-likelihood solution we used 250 repetitions based on Halton draws. The results of the sequential method are not corrected for efficiency.

In the first case ($Case$ 1, $CV_{(X_{21,n}-X_{11,n})} = 0.50$), we observe not only that the estimates are 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 that each estimate of each parameter of interest is equal to its target value.

However, note that for the lower-variance cases there is a problem recovering a significant constant (Cases 2 to 4). In fact, in the two most extreme low-variance cases (Cases 3 and 4), there is also a problem recovering a significant parameter for the first attribute (which is the parameter associated with the attribute with a low-variance). Even though in the extremely low-variance cases we are still replicating the target values, the (asymptotic) point estimates for both β_0 and β_1 are far from the target value (compare to Case 1). This situation results from inflated standard errors due to working with a variable with low-variance among the individuals (i.e. the data are not rich enough to estimate the model).

Regarding the latent variables, even though both methods recover the true parameters, there is a persistent problem with the t statistics of the sequential results. When using

δ	Target	Sequential Estimation				Simultaneous Estimation			
		$\hat{\delta}$	s.e.	t-stat	t-target	$\hat{\delta}$	s.e.	t-stat	t-target
β_0	0.50	0.506	0.187	2.70	0.03	0.517	0.172	2.99	0.10
β_1	-0.40	-0.393	0.032	-12.16	0.22	-0.408	0.034	-12.02	-0.22
β_2	-0.20	-0.192	0.011	-18.12	0.78	-0.200	0.011	-17.73	-0.02
Γ_1	-0.30	-0.284	0.177	-1.61	0.09	-0.280	0.089	-3.10	0.22
Γ_2	-0.40	-0.364	0.299	-1.22	0.12	-0.373	0.093	-3.99	0.29
b_1	0.80	0.784	0.077	10.25	-0.21	0.802	0.058	13.75	0.03
b_2	0.50	0.471	0.070	6.75	-0.42	0.484	0.050	9.62	-0.32
λ_{11}	0.40	0.382	0.032	11.88	-0.58	0.381	0.029	13.04	-0.65
λ_{21}	1.00	0.978	0.036	27.20	-0.62	0.980	0.036	27.50	-0.56
λ_{32}	0.60	0.614	0.034	18.18	0.42	0.611	0.032	19.23	0.34
λ_{42}	1.00	0.993	0.036	27.72	-0.21	0.998	0.036	28.08	-0.05

Table 3.2: Estimation results, *Case* 1 $CV_{(X_{21,n}-X_{11,n})} = 0.50$

δ	Target	Sequential Estimation				Simultaneous Estimation			
		$\hat{\delta}$	s.e.	t-stat	t-target	$\hat{\delta}$	s.e.	t-stat	t-target
β_0	0.50	0.299	0.686	0.44	-0.29	0.381	0.711	0.52	-0.17
β_1	-0.40	-0.343	0.138	-2.49	0.41	-0.369	0.142	-2.59	0.22
β_2	-0.20	-0.200	0.027	-7.31	0.01	-0.202	0.012	-17.27	-0.15
Γ_1	-0.30	-0.313	0.168	-1.87	-0.08	-0.370	0.089	-4.16	-0.79
Γ_2	-0.40	-0.430	0.272	-1.58	-0.11	-0.408	0.092	-4.41	-0.09
b_1	0.80	0.800	0.076	10.52	-0.002	0.797	0.057	13.91	-0.05
b_2	0.50	0.491	0.070	7.02	-0.13	0.488	0.050	9.84	-0.24
λ_{11}	0.40	0.385	0.032	12.10	-0.48	0.385	0.029	13.12	-0.51
λ_{21}	1.00	0.995	0.036	27.70	-0.13	0.996	0.036	27.81	-0.11
λ_{32}	0.60	0.598	0.034	17.78	-0.05	0.605	0.033	18.49	0.15
λ_{42}	1.00	0.998	0.036	27.89	-0.05	0.997	0.035	28.39	-0.09

Table 3.3: Estimation results, *Case* 2 $CV_{(X_{21,n}-X_{11,n})} = 0.10$

δ	Target	Sequential Estimation				Simultaneous Estimation			
		$\hat{\delta}$	s.e.	t-stat	t-target	$\hat{\delta}$	s.e.	t-stat	t-target
β_0	0.50	0.497	1.348	0.37	-0.002	0.610	1.406	0.44	0.08
β_1	-0.40	-0.386	0.269	-1.43	0.05	-0.421	0.282	-1.50	-0.07
β_2	-0.20	-0.190	0.010	-18.50	0.99	-0.199	0.011	-17.92	0.09
Γ_1	-0.30	-0.255	0.165	-1.54	0.27	-0.306	0.086	-3.56	-0.07
Γ_2	-0.40	-0.440	0.280	-1.57	-0.14	-0.406	0.090	-4.52	-0.07
b_1	0.80	0.816	0.076	10.73	0.21	0.823	0.058	14.17	0.40
b_2	0.50	0.490	0.070	7.03	-0.15	0.485	0.050	9.78	-0.31
λ_{11}	0.40	0.385	0.032	12.15	-0.49	0.389	0.030	13.16	-0.39
λ_{21}	1.00	1.001	0.036	27.86	0.02	1.003	0.036	28.07	0.08
λ_{32}	0.60	0.603	0.034	17.96	0.10	0.605	0.033	18.41	0.15
λ_{42}	1.00	1.010	0.036	28.17	0.27	1.012	0.036	28.04	0.32

Table 3.4: Estimation results, *Case* 3 $CV_{(X_{21,n}-X_{11,n})} = 0.05$

δ	Target	Sequential Estimation				Simultaneous Estimation			
		$\hat{\delta}$	s.e.	t-stat	t-target	$\hat{\delta}$	s.e.	t-stat	t-target
β_0	0.50	1.721	3.349	0.51	0.37	1.993	3.416	0.57	0.44
β_1	-0.40	-0.628	0.670	-0.94	-0.34	-0.698	0.684	-1.01	-0.44
β_2	-0.20	-0.191	0.010	-18.61	0.88	-0.199	0.011	-17.97	0.10
Γ_1	-0.30	-0.307	0.173	-1.78	-0.04	-0.325	0.086	-3.78	-0.29
Γ_2	-0.40	-0.447	0.288	-1.55	-0.16	-0.340	0.089	-3.78	0.68
b_1	0.80	0.781	0.075	10.37	-0.25	0.790	0.056	14.16	-0.17
b_2	0.50	0.476	0.070	6.84	-0.35	0.489	0.050	9.75	-0.22
λ_{11}	0.40	0.391	0.032	12.27	-0.28	0.396	0.030	13.40	-0.14
λ_{21}	1.00	1.004	0.036	27.94	0.13	1.005	0.036	27.98	0.14
λ_{32}	0.60	0.599	0.034	17.81	-0.03	0.598	0.033	18.19	-0.05
λ_{42}	1.00	1.011	0.036	28.19	0.31	1.014	0.036	28.42	0.39

Table 3.5: Estimation results, *Case* 4 $CV_{(X_{21,n}-X_{11,n})} = 0.02$

the sequential estimation method and for all experiments[9], the latent variables are not significant at a confidence level of 95%. This is a very important issue, as the t-stat is commonly used to define which variables are included in the utility function. Thus, even though the latent variables in this controlled experiment we know that the latent variables have an effect on choice, the associated parameters are statistically significant only by using the simultaneous full information SML solution.

Finally, the full information SML estimates capture efficiency gains that are reflected by smaller standard errors than those obtained with the sequential LIML.

3.4 Case Study: travel mode choice in Santiago

This case study presents an empirical application to modal choice of the code developed for the full-information simulated likelihood solution for HCMs. The data corresponds to revealed preferences from the fourth wave of the *Santiago Panel* (**?**), a five-day pseudo-diary reporting information about commute trips in the morning rush hour in Santiago de Chile.[10] On the fourth wave of the panel, survey participants were asked to state their perceptions on different characteristics of the modes of transportation. Having attitudinal data makes it possible to build an HCM for the Santiago Panel (**?**).

The sample of the fourth wave consists of 258 individuals who live in Santiago and commute to work at one of the five campuses of the Pontifical Catholic University of Chile. After a clean-up of inconsistencies on the perception module of the survey, there remain 1107 usable observations for HCM estimation.[11] Ten different modes of transportation are considered, both pure and combined: car-driver, car-passenger, shared taxi[12], metro, bus, car-driver & metro, car-passenger & metro, shared taxi & metro, bus & metro, and shared taxi & bus. For each mode, the following level of service variables are available: travel time (TT), walking time ($WALKT$), waiting time ($WAITT$), travel cost ($COST$), and number of transfers made ($TRANSF$). Socioeconomic information about the respondents is also available.

[9]Even for the best case with the highest variability.

[10]The whole study considers a panel of four different waves, from December 2006 to October 2008. Each wave consists on a five-day pseudo-diary.

[11]Each individual reports five choices, one per each working day.

[12]Shared taxis, or *colectivos* in Spanish, are a common mode of transportation in Latin America. In Santiago, shared taxis take up to 4 passengers and drive along a fixed route.

Regarding the attitudinal data, respondents were asked to rate their satisfaction on various aspects of the pure modes using a scale from 1 to 7.[13] The rated aspects are: safety regarding accidents, safety regarding theft, ease of access, comfort during the trip, availability of suitable information, possibility of calculating the travel time prior to the trip, and possibility of calculating the waiting time prior to the trip. Using factor analysis, three dimensions were identified: accessibility-comfort ($ACC\&COM$), reliability (REL), and safety SAF. The path diagram for hybrid choice modeling of the travel mode choice for the Santiago Panel is presented in Figure **??**.

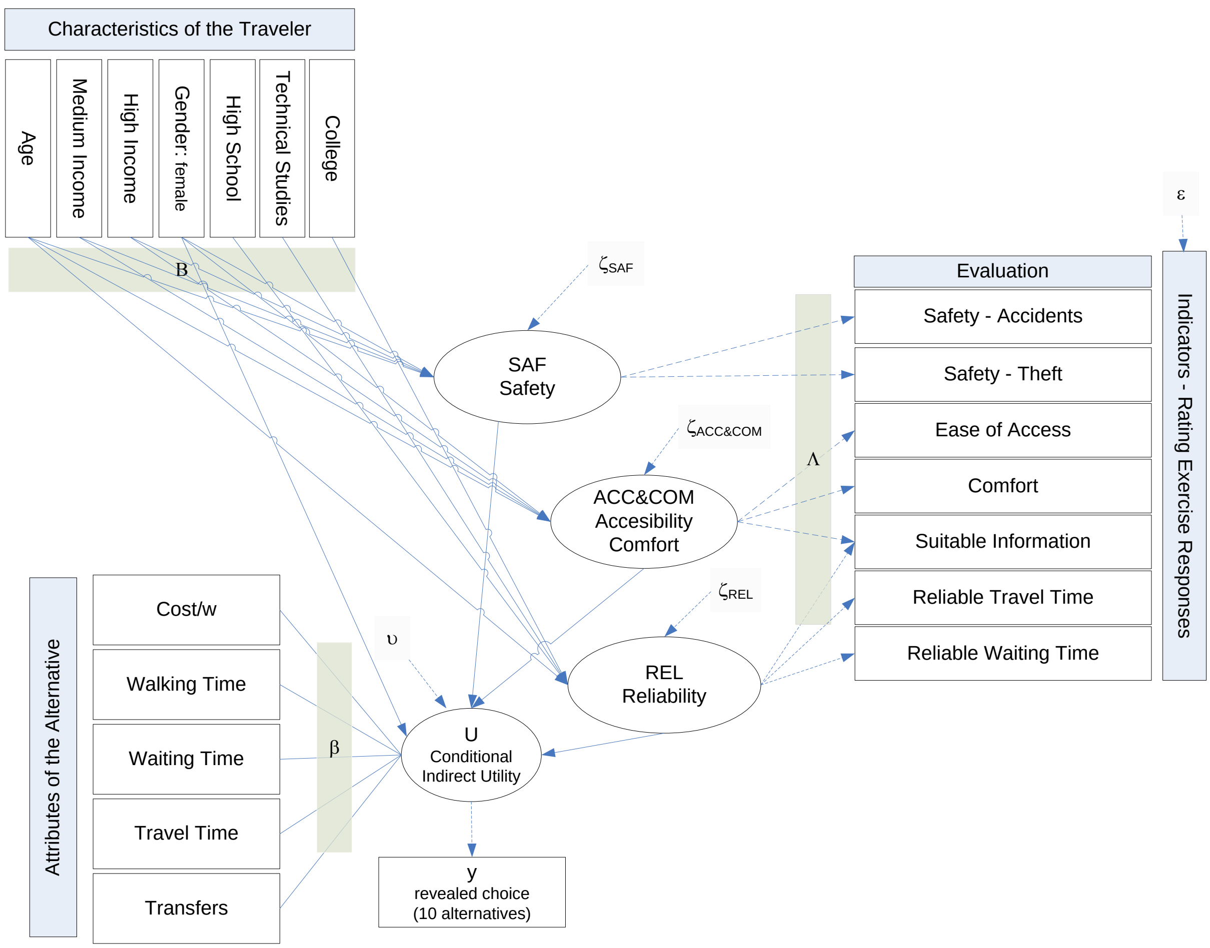

Figure 3.2: Travel Mode Choice HCM

For the indirect utility function, a specification based on the wage rate ω_n is used.

[13] Ratings from 1 up to 7 are based on the Chilean grading system. Grading in Chile considers a linear scale, where 1 stands for *very deficient* and 7 means *outstanding*.

Systematic taste variations in travel time are also incorporated (See **?**).

$$\begin{aligned} V_{in} &= \beta_{cost}\frac{COST_{in}}{\omega_n} + (\beta_{TT} + \beta_{TTw}Fem_n)TT_{in} + \beta_{walk}WALKT_{in} + \beta_{wait}WAITT_{in} \quad (3.31) \\ &\quad + \beta_{transf}TRANSF_{in} + \Gamma_{acc\&com}ACC\&COM_{in} + \Gamma_{rel}REL_{in} + \Gamma_{saf}SAF_{in} + \nu_{in}, \end{aligned}$$

where $\beta_{(\cdot)}$ and $\Gamma_{(\cdot)}$ represent unknown parameters. Note that systematic taste variations for the travel time (TT) variable are introduced; since Fem is a gender indicator for women, the marginal disutility of travel time is β_{TT} for men, and $\beta_{TT} + \beta_{TTw}$ for women.

Since levels of satisfaction were provided only for the pure modes (without combinations), the alternative-specific latent variables $ACC\&COM_{in}$, REL_{in}, and SAF_{in} affect the utility functions of the travel modes car-driver, car-passenger, shared taxi, metro, and bus. Also, because of the alternative-specific nature of the latent variables in this case, the three concepts (accessibility-comfort, reliability, and safety) are operationalized as 15 different latent variables.

The results for the choice model are presented in Table **??**. Estimation of the parameters was performed using 250 Halton-based repetitions for the simulator.[14] For reference, an MNL (without latent variables) was also considered. Details about the rest of the joint HCM estimation, including the structural and measurement equations, are provided in **?**. In **?** a comparison with simultaneous estimation results is provided.

The signs of all the estimated parameters are consistent with microeconomic theory: the marginal utilities of the latent variables are positive[15], whereas the time-related attributes, cost, and transfers represent a disutility to individuals. Also, note that all the parameters in the HCM are statistically significant, at least at the 90% confidence level.

Even though men are more sensitive to travel time than women according to the HCM results, in the MNL case (without the latent variables) the opposite effect is observed. In fact, the difference in magnitude of the marginal disutility of travel time for women is remarkably high (nearly a 500% difference between the HCM and the MNL). These differences, in sign and magnitude, show a potential serious problem when using a model that neglects latent variables for forecasting.

To have a notion of the impact of these differences, willingness to pay (WTP) measures

[14]Note that since 15 latent variables are considered, this case study is a large-scale application of FISML for HCMs.

[15]A positive sign indicates that accessibility-comfort, reliability, and safety are desirable features of a travel mode.

Mode Choice Model	HCM		MNL	
	est	t-stat	est	t-stat
$\beta_{cost/w}$	-0.022	-7.32	-0.019	-8.13
β_{TT}	-0.006	-4.67	-0.033	-4.82
β_{TTw}	-0.001	-3.01	0.03	2.98
β_{wait}	-0.015	-1.69	-0.009	-0.53
β_{walk}	-0.022	-2.89	-0.016	-1.8
β_{transf}	-1.102	-8.21	-1.11	-8.2
$\Gamma_{acc\&com}$	0.622	3.79	-	
Γ_{rel}	0.441	2.7	-	
Γ_{saf}	0.613	1.87	-	
ASC_{car-dr}	0.733	2.03	1.22	5.84
$ASC_{car-pass}$	-0.889	-2.12	-0.8	-3.64
$ASC_{shared-taxi}$	-1.331	-1.78	-1.42	-4.6
ASC_{metro}	0.247	0.81	0.241	1.56
$ASC_{car-dr/metro}$	0.223	0.51	0.779	2.65
$ASC_{car-pass/metro}$	-0.882	-2.22	-0.309	-1.28
$ASC_{shared-taxi/metro}$	-0.913	-1.55	-0.078	-0.36
$ASC_{bus/metro}$	0.342	1.41	0.608	4.59
$ASC_{shared-taxi/bus}$	-1.005	-3.68	-0.473	-1.68

Table 3.6: Santiago case study, choice model results

were reported. Table **??** presents the subjective values of travel, waiting and walking time[16], as well as the valuations associated with transfers and the three latent variables. 95% confidence intervals for these WTP measures were calculated using the approach proposed by **?**. For all the valuations shown, an individual mean wage rate of \$5.35 [USD] per work hour was considered.[17]

Subjective Value of	HCM		MNL	
TT for men [\$/hr]	1.201	[0.11,2.73]	0.457	[0.03,1]
TT for women [\$/hr]	1.029	[0.14,2.76]	6.431	[0.81,16.48]
$WAITT$ [\$/hr]	2.544	[0.06,5.29]	1.758	[0.01,3.53]
$WALKT$ [\$/hr]	3.716	[0.26,8.33]	1.729	[0.09,6.55]
$TRANSF$ [\$/transfer]	3.087	[0.9,13.71]	3.644	[0.94,14.32]
$ACC\&COM$ [\$/unit]	1.744	[0.3,4.29]	-	-
REL [\$/unit]	1.229	[0.11,2.77]	-	-
SAF [\$/unit]	1.715	[0.11,3.66]	-	-

Table 3.7: Santiago case study, subjective valuations

Whereas the HCM produces reasonable valuations of travel time savings (on the order of 20 to 25% of the wage rate) it is clear that the subjective value of travel time obtained from the model without latent variables is problematic. In fact, for men it appears to be underestimated (it represents only 8.6% of the mean wage rate), while for women it is clearly overestimated (20% higher than the mean wage rate). This clear disadvantage of the model without latent variables reaffirms the importance of including latent variables in the formulation of the model: the decision about including or not latent variables in demand models may have significant consequences when forecasting or evaluating transportation policies.

3.5 Conclusions

In this chapter I have described the system of equations associated with hybrid choice models. This system is 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

[16]Subjective values of time represent the amount of money a traveler is willing to pay to reduce in 1 unit the time spent (traveling, waiting, or walking).

[17]This value was obtained from the socio-demographic data of the Santiago Panel.

Figure 3.3: Santiago de Chile and its transportation system

(continuous or discrete) indicators. I have shown that although classical HCM estimation requires the evaluation of complex multi-dimensional integrals, a simulation-based approximation can be successfully derived. This full information simulated maximum likelihood (SML) solution is a valid estimator of the true solution and offers an unbiased, consistent and smooth estimator of the true probabilities. However, classical estimation of HCMs is computationally very demanding in situations with numerous latent variables and large sets of potentially interrelated choices. In fact, the latent variables affect the behavior of the simulated likelihood function in such a way that a standard optimization algorithm may require a huge number of iterations to converge.

I also described an alternative estimation method, based on a sequential two step procedure. Although this method appears as conceptually simpler than the full information estimator, the limited information maximum likelihood (LIML) estimator is not efficient. Moreover, the sequential method as it is applied in practice does not offer consistent estimators either. The problems of inefficiency and inconsistency can be overcome by correcting the covariance matrix of the second step at the expense of a highly complex procedure. The results of a Monte Carlo simulation study show that a two step estimator may have problems recovering the parameters of the latent variables. Also, in the simulation experiments the full information SML estimates capture efficiency gains that are reflected by smaller standard errors than those obtained with the sequential LIML.

Finally, I presented a case study showing that the full information SML solution for HCMs is genuinely capable of adapting to practical situations. The travel mode choice results show differences in magnitude for certain parameters when latent variables are omitted. This omission translates into problems in the willingness to pay measures derived from the model without latent variables. A clear example of this are the low subjective value of time obtained for men and the high subjective value of time obtained for women (which is even larger than the mean wage rate), when latent variables are not considered. This empirical disadvantage of omitting latent variables reaffirms the importance of the HCM framework: the decision about including or not latent variables in demand models may have severe consequences when forecasting or evaluating policies.

References

P.A. Armstrong, R.A. Garrido, and J. de D. Ortúzar. Confidence intervals to bound the value of time. *Transportation Research Part E*, 37(2):143–161, 2001.

K. Ashok, W. Dillon, and S. Yuan. Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, 39(1):31–46, 2002.

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

M. Ben-Akiva, D. Bolduc, and J. Walker. Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. Working Paper, MIT, 2001.

M. Ben-Akiva, 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): 163–175, 2002.

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

D. Bolduc. Generalized autoregressive errors in the multinomial probit model. *Transportation Research B*, 26:155–170, 1992.

D. Bolduc, M. Ben-Akiva, J. Walker, and A. Michaud. Hybrid choice models with logit kernel: applicability to large scale models. In M. Lee-Gosselin and S. Doherty, editors, *Integrated Land-Use and Transportation Models: Behavioral Foundations*, pages 275–302. Elsevier, New York, 2005.

D. A. Bunch. Estimability in the multinomial probit model. *Transportation Research B*, 25: 1–12, 1991.

E. Cherchi and J. de D. Ortúzar. Empirical identification in the mixed logit model: Analysing the effect of data richness. *Networks and Spatial Economics*, 8(2):109–124, 2008.

B. R. Dansie. Parameter estimability in the multinomial probit model. *Transportation Research*, 19:526–528, 1985.

M. Vredin Johansson, T. Heldt, and P. Johansson. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40:507–525, 2006.

D. McFadden. The choice theory approach to market research. *Marketing Science*, 5(4):275–297, 1986. Special issue on consumer choice models.

D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57:995–1026, 1989.

M.A. Munizaga and R. Alvarez-Daziano. Testing mixed logit and probit models by simulation. *Transportation Research Record*, 1921:53–62, 2005.

K.M. Murphy and R.H. Topel. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, 3(1):370–379, 1985.

S. Raveau. *Estimación Simultánea de Modelos de Elección Discreta con Variables Latentes*. Civil Engineering Thesis, Pontificia Universidad Católica de Chile, Department of Transport Engineering and Logistics, 2009. (in Spanish).

S. Raveau, R. Alvarez-Daziano, M.F. Yáñez, D. Bolduc, and J.D. Ortúzar. Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record, 2010. (in press).

D. C. Stapleton. Analyzing political participation data with a MIMIC model. *Sociological Methodology*, 9:52–74, 1978.

M.L. Tam, W.H.K. Lam, and H.P. Lo. Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica*, 6(1):3–17, 2010.

K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, second edition, 2009.

K. Train, D. McFadden, and A. Goett. The incorporation of attitudes in econometric models of consumer choice. Working paper, Cambridge Systematics, 1986.

J. Walker. The mixed logit (or logit kernel) model: Dispelling misconceptions of identification. *Transportation Research Record*, 1805:86–98, 2002.

J. Walker and M. Ben-Akiva. Generalized random utility model. *Mathematical Social Sciences*, 43(3):303–343, 2002.

M.F. Yáñez, P. Mansilla, and J. de D. Ortúzar. The Santiago panel: measuring the effects of implementing Transantiago. *Transportation*, 2009. In press.

Chapter 4

Incorporating environmental preferences toward low-emission vehicles

Modeling private vehicle purchase decisions using discrete choice models has a long tradition; more recently this kind of model has been used to analyze choice among different automobile technologies, such as the use of alternative fuels. Although environmentally-conscious consumers should be more likely to choose low-emission vehicles, current demand models have a hard time representing this likelihood.

Using stated data on both vehicle purchase decisions and environmental concerns, I analyze the practical feasibility of a Bayesian estimator for hybrid choice models. I show that the Bayesian approach for HCMs is methodologically easier to implement than simulated maximum likelihood because the inclusion of latent variables translates into adding independent ordinary regressions; and also because forecasting and deriving confidence intervals for willingness to pay measures is straightforward. My empirical results coincide with a priori expectations, namely that environmentally-conscious consumers are aware of the dangers of climate change and oil dependency; their concerns about the role of transportation in global warming change their consuming behavior, and they are willing to pay more for sustainable solutions (low-emission vehicles) despite potential drawbacks (such as a reduced refueling availability). The model outperforms standard discrete choice models because it not only incorporates pro-environmental preferences but also provides tools to build a profile of eco-friendly consumers.

4.1 Environmental effects of individual travel behavior

Individual travel behavior is related to how people move over space and time. People travel to accomplish diverse activities that need to performed in locations that are distant, and therefore travel demand is derived from the activity system. In this sense, traveling is a necessity. Transportation is extremely important, since it is a key motor for the economy.[1] Because of the strong relationship between transportation, energy consumption, the environment, and the economy in general, the negative externalities produced by our individual travel behavior[2] need to be addressed and mitigated. We need a transportation system that is not only quick and efficient, but also cleaner and sustainable. Thus, the effects of individual travel behavior on the environment are of special interest.

Transportation, broadly defined, plays a major role in oil dependency. To illustrate this fact, in Figure **??** the US oil demand by sector is depicted. Note that the share associated with transportation (shown on the secondary vertical axis) reached 70% in recent years. Canada is the top supplier of both crude oil and total petroleum for the US. In the internal Canadian market, the estimated oil consumption for 2010 is 2,260,000 barrels per day.

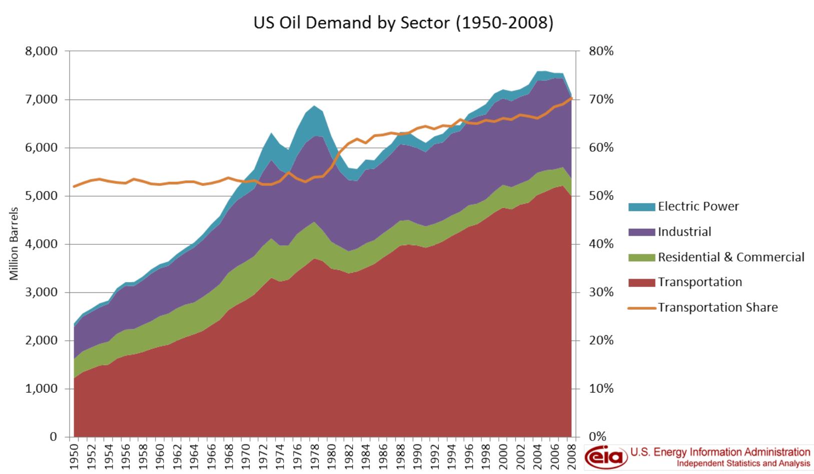

Figure 4.1: US oil demand by sector

Given the fact that the internal combustion engine is the current dominating technology

[1]For instance, we commute to work and we travel for shopping. In general, economic growth requires a transportation system that can move people, goods and services quickly and efficiently.

[2]Such as pollution, congestion, accidents, and road deterioration.

in the automobile industry, oil is nearly the sole energy source for transportation (See Figure **??**.) There are several issues associated with oil dependency. On the one hand,

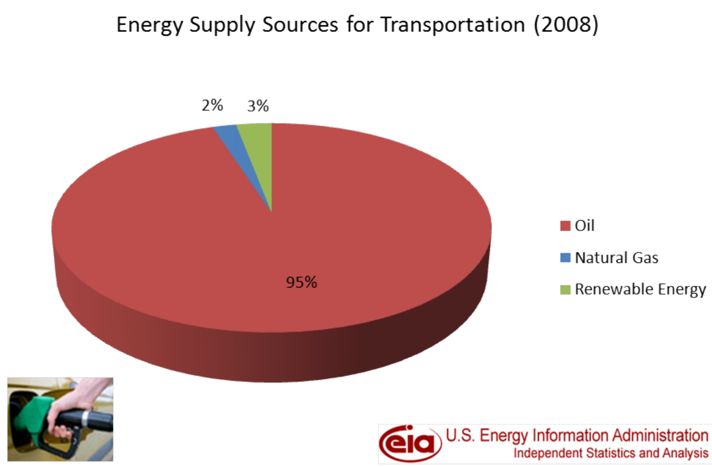

Figure 4.2: Energy Supply Sources for Transportation

dependence on foreign production is a potential problem for national and economic security. On the other hand, as revealed by the 2010 Gulf of Mexico oil spill, security concerns related to our oil dependency go beyond reliance on foreign imports.

In the case of fossil fuels, energy use is accompanied by emissions that may have an impact on climate change. As Figure **??** shows, in the last decade transportation became the principal source of carbon dioxide. Because of dependence on oil, the vast majority

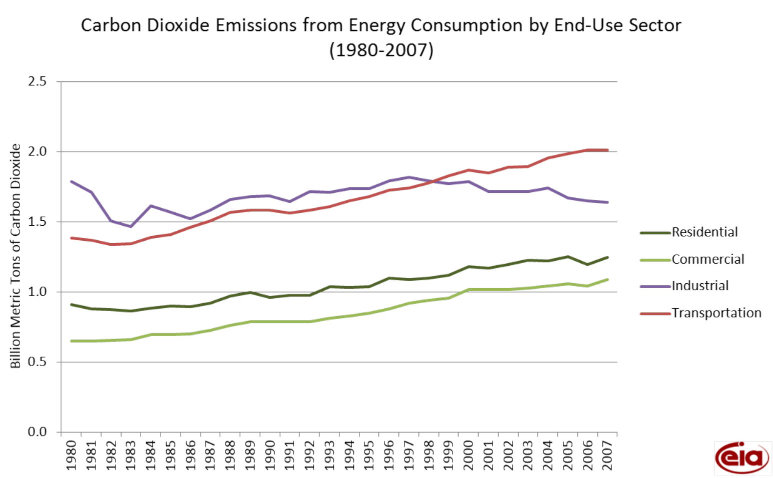

Figure 4.3: Carbon dioxide emissions from energy consumption by end-use sector

of emissions come from burning oil (See Figure **??**.) In Canada, total carbon dioxide

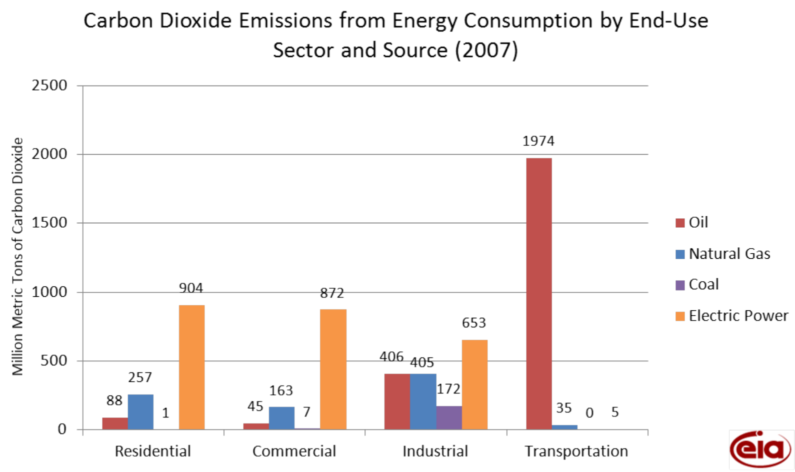

Figure 4.4: Carbon dioxide from energy consumption by end-use sector and source

emissions from consumption of fossil fuels achieved 29,195 million metric tons in 2008. Private transportation in Canada accounts for almost 15% of the total greenhouse gas emissions, including NO_x, CO, SO_2, COV, and PM_{10}.

To achieve a clean and more secure energy future, we need to study solutions that target the goals of energy independence and security, as well as reduce greenhouse gas emissions and other pollutants. Technological innovation and the use of cleaner alternative fuels have traditionally been proposed as sustainable solutions to transportation problems such as oil dependency and greenhouse gas emissions. Henceforth, a transition to sustainable transportation technologies (and behavior) will be desirable, including development of low emission cars[3] and alternative fuels. While automobile manufacturers have been working in research and development of cleaner and more efficient vehicles, it is not possible to forecast the future market conditions based on the supply side alone[4]. Effectively, consumer response is essential to understand and predict the reasons for a successful introduction and penetration of low emission vehicles in the market. The demand side is also relevant for planning investments related to fuel supply, as well as for planning government actions such as policies and economic incentives. Thus, to understand the effects of climate change and energy security concerns on travel behavior it is relevant to model how a consumer decides which vehicle to purchase.

[3]In particular, automobile use is associated with numerous problems including greenhouse gas emissions.

[4]The supply side corresponds to the different vehicles that will become available as a result of research and development programs looking at technical success and cost reductions from volume production of low emission vehicles.

4.2 Car purchase decisions and environmental preferences

New car purchasing is an example of discrete choice. Effectively, economic-preference models of discrete choice aim to explain the process of individual choice among a mutually exclusive, exhaustive and finite group of alternatives (**?**). According to consumer theory the decision process reflects rational preferences set by a utility-maximization behavior. In the case of standard consumption theory, the utility function representing the preference relation depends on the continuous quantities composing the consumption set. However, when the nature of a specific good is discrete, the preference relation is assumed to depend on a group of attributes (**?**) combined according to individual tastes. In the context of private vehicle purchase decisions, each vehicle is described by a group of attributes, such as make and model, purchase price, performance, reliability, durability, comfort, style, and safety. According to individual preferences, each consumer selects the alternative that has the highest level of satisfaction. Then, the market demand for private vehicles is determined by the market share of each alternative, which is constructed as the number of consumers choosing each particular alternative. In discrete choice modeling, the most common approach is based on random utility theory (**?**), which introduces the concept of individual choice behavior being intrinsically probabilistic. Whereas the Random Utility Maximization (RUM) framework recognizes the existence of a systematic component of individual behavior (the decision maker is capable of perfect discrimination), RUM also takes into account the incapacity of the analyst to observe all the variables that have an influence on the decision (incomplete information that implies the presence of uncertainty; **?**). Different RUM discrete choice models can be derived based on various assumptions on the distribution of the random term. The probabilistic nature of the choice behavior implied by the RUM framework leads to individual probabilities of each consumer selecting each available alternative.

Modeling the private vehicle purchase decision using RUM discrete choice models has a long tradition (for example **?????**). In these applications, the alternatives are different types of private vehicles, such as choice between car and SUV. Each alternative is described using attributes such as purchase cost, fuel economies, vehicle size and age of the vehicle. More recently – owing to the interest in studying sustainable solutions to the environmental problems created by personal transportation – this kind of model has been used to analyze choice among different automobile technologies, namely the use of alternative fuels (**?????**). In this context, the baseline is the use of standard attributes such as purchase price, operating costs (including both maintenance and fuel costs), and power (comprising motor power, performance, top speed, and acceleration). However, to characterize vehicles with alternative technologies and/or fuel types we must introduce new sets of attributes that take into account special features or requirements for the new alternatives (**??????**). These attributes include variables such as *service station availability* (stations selling the proper fuel – relevant in the case of new fuels), *driving range* (some new technologies such as electric cars suffer from a limited driving range between refueling), whether or not the vehicle would be granted certain priorities (such as *express lane access*), and *greenhouse gas emissions* GHG (CO_2 concentrations play a key role in global warming; and hence a reduction on this variable – which in fact is the leading objective of climate policies – determines how 'green' the new technology is). However, it is hard to maintain that these characteristics alone permit a full representation of consumer behavior that allow us to understand demand for 'green' (low-emission) personal vehicles. For instance using *greenhouse gas emissions*, environmental concerns are represented only by emission reductions without any consideration to other dimensions such as eco-friendly habits (cf. **?**).

Consumers' preferences for green vehicles must be understood first in a context where the new technologies often have a low market share (or even a zero market penetration in the case of the introduction of a new alternative) and hence the role of knowledge, experience and information is critical. Second, demand for low-emission vehicles is a decision-making process guided by environmental preferences, among other dimensions such as desires for energy independence and for advanced technologies. **?** finds that households with environmental knowledge and attitudes own fewer and more fuel-efficient vehicles; these households actually show an eco-friendly travel behavior because they drive their vehicles less. Environmentally-conscious consumers are aware of the dangers of climate change and oil dependency; their concerns about the role of transportation in global warming has a consequent change in their purchasing and travel behavior, and they are willing to pay more for sustainable solutions (low-emission vehicles) despite potential drawbacks (such as a reduced refueling availability). Although environmentally-conscious consumers should be more likely to choose vehicles that are good for the environment, current demand models have a hard time representing this likelihood. The key is then how to incorporate the consumer's environmental concerns into an economic model for private vehicle purchase decisions.

According to cognitive psychology, preferences and behavior (toward green technologies in the case of environmental psychology) are affected by perceptions and attitudes. On the one hand, *perception variables* measure the individual cognitive capacity to represent the attributes of different alternatives. Perceptions are relevant because the choice process depends on how attribute levels are perceived according to the individual beliefs of a specific consumer. On the other hand, *attitude variables* measure the evaluation of favor or disfavor assigned by the individual to the features of different alternatives. Attitudes influence behavioral intentions (e.g. to adopt a new technology), are related to individual heterogeneity (*taste variations*) and reflect individual tastes, needs, values, goals, and capabilities that develop over time and are affected by knowledge, experience and external factors, such as the socioeconomic characteristics of the decision-maker (**?**).[5]

Attitudinal research[6] for green vehicles has mainly been centered on public acceptance of hydrogen and fuel cell technologies using attitudinal/perceptual surveys (**???**). Consumers reveal highly positive attitudes toward green vehicles, although knowledge of the new technologies is low (for a more comprehensive review of the attitudinal approach see **?**). **?** also review the economic preference approach as well as the semiotic approach applied by **?**. An important problem with the attitudinal approach (if used independently from economic models of choice) is that it does not necessarily explain economic choice behavior, and in some cases the attitudes being measured are not even directly related to actual purchase intentions (for instance, **??**, measured general attitudes and knowledge toward hydrogen vehicles, as a concept for technological development not as a choice). In fact as noted by **?**, economic preference surveys usually provide lower acceptance levels for new technologies than those predicted by attitudes alone.

In sum, there are two important modeling tools to analyze vehicle choice, namely economic-preference models (discrete choice models) and attitudinal models. Although it is clear that both approaches should be integrated, the literature does not offer any example of such an expansion. Hybrid choice modeling provides the methodological framework to accomplish the task of integrating latent pro-environmental preferences into an economic model for private vehicle purchase decisions.

In this chapter I analyze stated choices made by Canadian consumers when faced with green personal vehicle alternatives (**?**). I seek to implement both theoretically and empirically a Bayesian approach to an HCM of personal vehicle choice (cf. **?**, where using the same data we analyzed classical estimation of HCMs). Specifically, I construct an HCM setting where I take perceptual indicator variables about transport policies and problems, and then define an environment-related latent variable which enters directly into the choice process. This paper expands on my previous research (**??**) by introduc-

[5] For a discussion related to transportation see **?**

[6] Psychometric studies, mostly focused on factor analysis.

ing Bayesian methods to analyze the data, not only for estimation of parameters and willingness to pay measures but also for forecasting policy scenarios.

4.3 Personal vehicle choice data

I use data from a survey conducted in 2002 by the EMRG (Energy and Materials Research Group, Simon Fraser University) of stated personal vehicle choices made by Canadian consumers when faced with technological innovations. Full details regarding the survey, including the design of the questionnaire, the process of conducting the survey, and analysis of the collected data can be found in **?**.

Survey participants were first contacted in a telephone interview used to personalize 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:* Personal vehicle choice (stated preference experiment);
- *Part 3:* Transportation mode preferences;
- *Part 4:* Views on transportation issues; and
- *Part 5:* Additional information (gender, education, income).

SP questions in Part 2 considered four vehicle types:

1. *Standard gasoline vehicle* (SGV): operating on gasoline or diesel,
2. *Alternative fuel vehicle* (AFV): natural-gas vehicle,
3. *Hybrid vehicle* (HEV): gasoline-electric, and
4. *Hydrogen fuel cell vehicle* (HFC).

For each of these alternative vehicle types, the attributes were defined as:

- *Purchase price*: capital cost associated with the purchase of a new car [CAD2002/10000],
- *Fuel cost*: monthly operating costs [CAD2002/100-month],
- *Fuel availability*: proportion of stations selling the proper fuel [ratio],
- *Express lane access*: whether or not the vehicle would be granted express lane access,
- *Emissions data*: emissions compared to a standard gasoline vehicle [ratio], and

- *Power*: horsepower of engine compared to current vehicle [ratio].

The sampled individuals were randomly drawn from households living in Canadian urban centers with populations of more than 250,000 people. Respondents have an average household income approximately equal to $62,000 CAD, and a high level of education (75% of the sample attained undergraduate degrees or completed graduate school). The sample has a 59% proportion of females, and 59% of the sampled individuals are 41 years or older. Each participant needed to either have access to a vehicle, or commute to work. The respondents who met these criteria were asked to make up to four consecutive virtual choices while the vehicle attribute values were modified after each round according to randomized blocks of an individual-customized labeled SP experimental design (**?**, see Table **??**). The sample has 866 completed surveys (of the total of 1150 individuals, 75% response rate). After a clean up where we retain only the individuals who answered the whole perceptual-attitudinal rating exercise, there remain 1877 usable observations (pseudo-individuals) for HCM estimation. This analytic sample consists mainly of workers (80%) who commute, mostly driving alone (68%).

	SGV	AFV	HEV	HFC
Purchase Price (PP)	100% PP 105% PP 110% PP 115% PP	105% PP 110% PP	105% PP 120% PP	110% PP 120% PP
Fuel Cost (FC)	100% FC 110% FC 120% FC 130% FC	110% FC 120% FC	75% SGV	110% FC 120% FC
Fuel Availability	100%	25% 75%	100%	25% 75%
Express lane access	No	No Yes	= AFV	No Yes
Emissions	Equal	10% less	25% less	100% less
Power	Equal	Equal 10% less	Equal 10% less	Equal 10% less

Table 4.1: Experimental attribute levels, **?**

According to my literature review, emission data is the standard way to describe choice behavior of environmentally-conscious consumers when using discrete choice models for vehicle purchase decisions. But in the EMRG survey the emission variable does not vary across choice situations in the SP design. This simplifying assumption was made to avoid an explosive number of choice situations (see discussion in **?**, a fractional factorial design was used – this problem could have been avoided using an efficient SP design). The consequence of this assumption is that the effects of environmental benefits related to emission reductions cannot be distinguished from the alternative specific constants of a discrete choice model. This is a major problem if we make the hypothesis that ecologically motivated consumers have a different purchase behavior. However, the introduction of a latent variable will solve this issue.

In fact, using this data I want to model the impact of environmental-related cognitive factors on the private vehicle purchase decision. The first step to address this issue through an HCM is to set the latent variables involved. My hypothesis here is that the private vehicle purchase decision is affected not only by the attributes of the different vehicles but also by the environmental awareness of the consumer. Ultimately, an environmentally-conscious consumer should prefer a cleaner automobile technology associated with less environmental impact. In my model, this effect is taken into account by introducing the latent variable ***Environmental Concern*** (EC), related precisely to transportation and its environmental impact.

I continue the analysis focusing on two different relevant questions of the survey that translates into both transport policies support and transport problems evaluation.

Transport Policies Support (TPS): Evaluation of 8 different policies or government actions that influence the transportation system - according to degree of support: 5 levels from *Strongly Opposed* to *Strongly Supportive* (see Figure **??**).

1. Improving traffic flow by building new roads and expanding existing roads.
2. Discouraging automobile use with road tolls, gas taxes, and vehicle surcharges.
3. Making neighborhoods more attractive to walkers and cyclists by using bike lanes and speed controls.
4. Reducing vehicle emissions with regular testing and manufacturer emissions standards.
5. Making carpooling and transit faster by giving them dedicated traffic lanes and priority at intersections.
6. Making transit more attractive by reducing fares, increasing frequency, and expanding route coverage.
7. Reducing transportation distances by promoting mixed commercial, residential, and high-density development.

8. Reducing transportation needs by encouraging compressed workweeks and working from home.

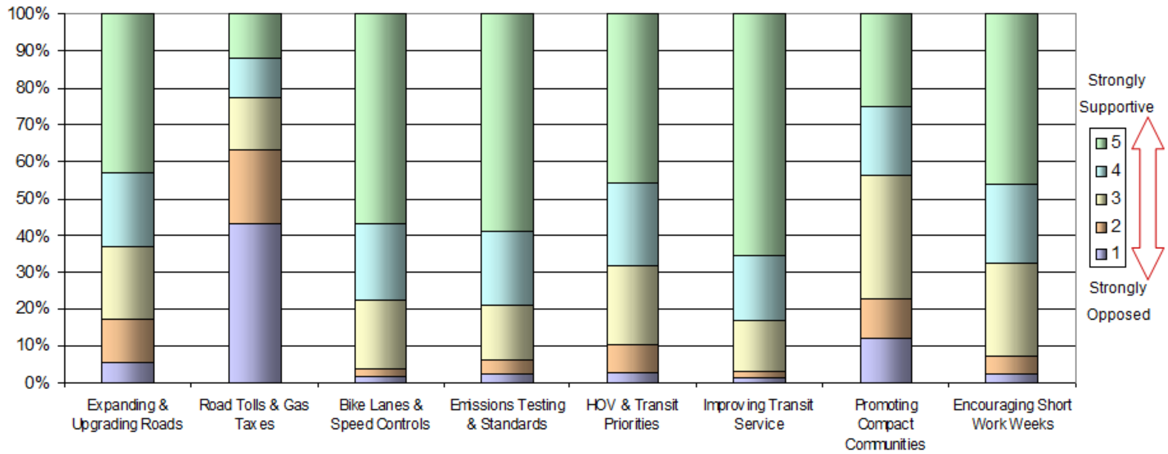

Figure 4.5: Transport Policies Support

Transport Problems Evaluation (TPE): Evaluation of 6 different problems related to transportation according to degree of seriousness: 5 levels from *Not a Problem* to *Major Problem* (see Figure **??**).

1. Traffic congestion that you experience while driving.
2. Traffic noise that you hear at home, work, or school.
3. Vehicle emissions that affect local air quality.
4. Accidents caused by aggressive or absent-minded drivers.
5. Vehicle emissions that contribute to global warming.
6. Unsafe communities because of speeding traffic.

The answers to these questions serve as perceptual indicator variables, which are used for construction of the environmental concern (EC) latent variable. This way, the EC variable measures consumers' concerns and awareness about transportation issues affecting the natural environment (e.g. possibility of reducing car emissions or the introduction of road tolls and gas taxes; problems related to poor local air quality, emissions and global warming) as well as the mobility environment (e.g. traffic congestion, traffic noise, safety concerns). The structure defining EC as a unidimensional construct – as opposed to a diverse structure differentiating, for instance, the natural and mobility environmental concerns – was tested among alternative relationships according to different MIMIC

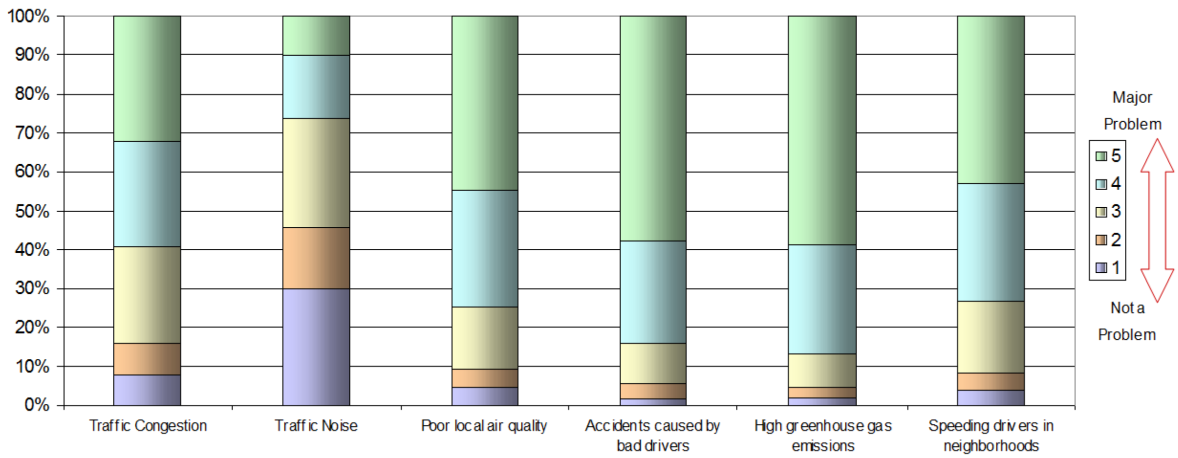

Figure 4.6: Transport Problems Evaluation

models. We did not use factor analysis (FA) basically because FA does not allow us to distinguish the effects of market segments that are taken into account when estimating the MIMIC models. We checked the reliability of the current structure for the EC variable getting an acceptable level of internal consistency (Cronbach's $\alpha = 0.7018$.)

4.4 Private vehicle Hybrid Choice Model

4.4.1 The HCM setting

In this particular choice context, I aim to explain the process of individual choice among the mutually exclusive, exhaustive and finite group of the personal vehicle alternatives: *standard gasoline vehicle* (SGV), *alternative fuel vehicle* (AFV), *hybrid electric vehicle* (HEV), and *hydrogen fuel cell vehicle* (HFC). At the same time, I postulate that the latent *environmental concern* (EC) variable, which reflects pro-environmental preferences, has a significant impact on the vehicle purchase decision. The whole behavioral process is represented by the following HCM group of structural and measurement equations (cf. **?**):

Structural equations

$$\mathrm{EC}_n = w_n b + \zeta_n, \quad \zeta_n \sim N(0,1) \tag{4.1}$$

$$U_n = X_n \beta + \Gamma \cdot \mathrm{EC}_n + \upsilon_n \tag{4.2}$$

Measurement equations

$$I_n = \Lambda \cdot \mathrm{EC}_n + \varepsilon_n, \quad \varepsilon_n \sim MVN(0, \mathrm{I}_{14}) \tag{4.3}$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

where EC is the latent environmental concern variable; w_n is a (1×10) vector of explanatory variables affecting the latent variable; b is a (10×1) vector of unknown parameters used to describe the effect of w_n on the latent variable. The choice model in equation (**??**) is written in vector form where we include the 4 private vehicle alternatives. Therefore, U_n is a (4×1) vector of utilities; υ_n is a (4×1) vector of error terms associated with the utility terms. X_n is a (4×8) attribute matrix – including 5 experimental attributes and 3 alternative specific constants (ASC) – with X_{in} designating the ith row of X_n. β is a (8×1) vector of unknown parameters. Γ is a (4×4) diagonal matrix of unknown parameters associated with the latent variable EC, with Γ_i designating the ith diagonal element of matrix Γ.

In the set of measurement equations (**??**), I_n corresponds to a (14×1) vector of the 14 indicators of latent variable EC associated with individual n; and Λ is a (14×1) vector of unknown parameters that relate the latent variable EC to the indicators. Even though the indicator variables are ratings (1-5), we treat them as being continuous according to standard practice in latent variable models. The term ε_n is a (14×1) vector of independent error terms with unitary variance – I_{14} being the identity matrix of size 14. Regarding independence of the measurement equations, we do recognize in our model that indicators are highly correlated variables, but we assume that the correlation structure is due to commonality of each indicator with the latent construct EC. Once we account for this commonality by modeling each indicator as a function of EC, then the residual of each indicator can reasonably be assumed to be uncorrelated with the other residuals. Additionally, a diagonal matrix is required for identification of the model.

Finally, we stack each individual choice indicator variable y_{in} into a (4×1) vector called y_n. For a full description of the equations and variables, see Appendix A.

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

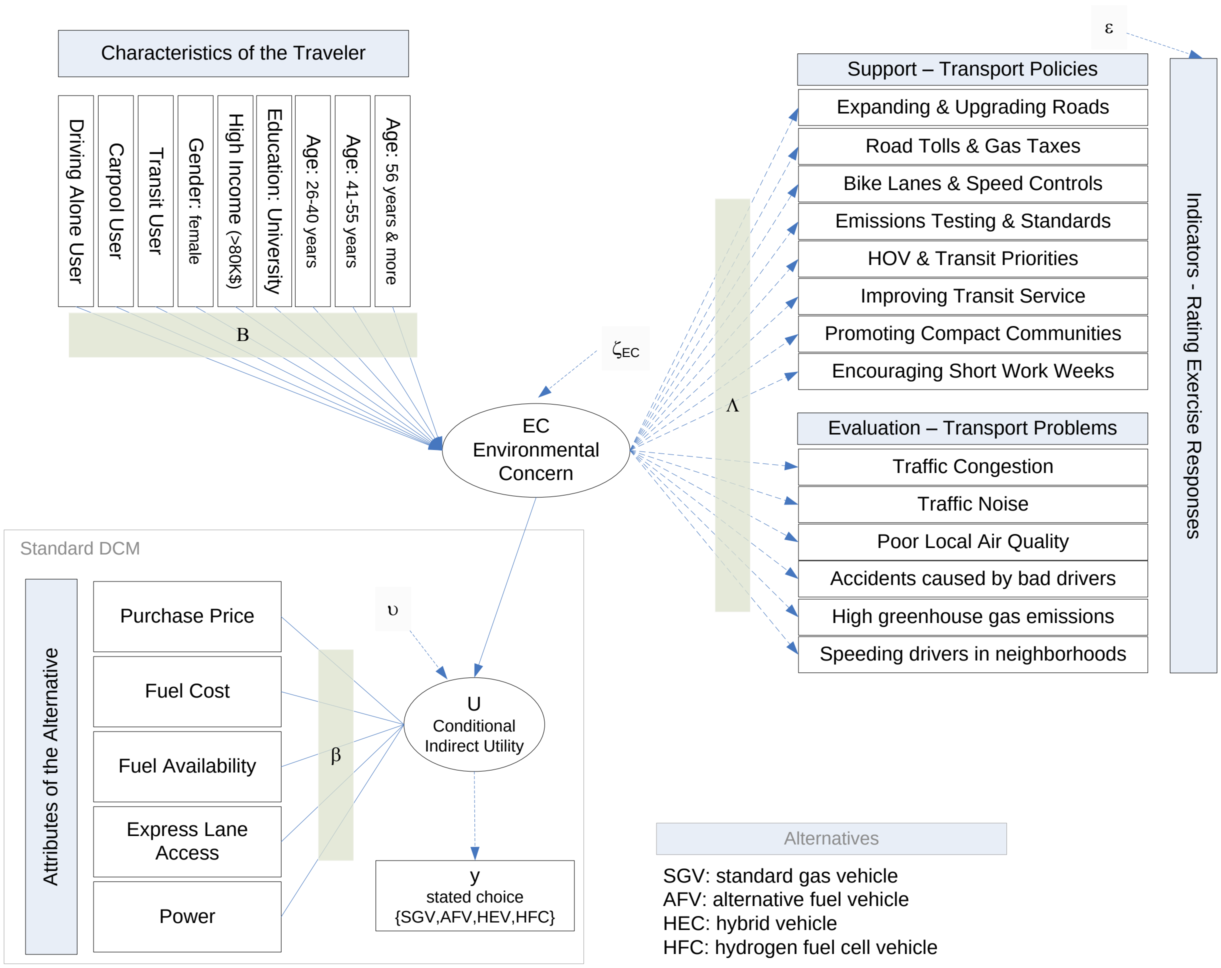

Figure 4.7: Private vehicle purchase HCM

If the latent variable EC were not present, the personal vehicle choice probability 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 EC variable, the choice probability would be represented by $P_n(i\,|\mathrm{EC}, X_n, \theta)$ where θ contains all the unknown parameters in the choice model of equation (**??**). Since EC is not actually observed, the choice probability is obtained by integrating the latter expression over the whole space defined by EC:

$$P_n(i|X_n, w_n, \theta, b) = \int_{\mathrm{EC}} P_n(i\,|\mathrm{EC}, X_n, \theta)g(\mathrm{EC}|w_n, b)d\mathrm{EC}, \qquad (4.5)$$

where $g(\mathrm{EC}_n|w_n, b)$ is the density of $N(w_n b, 1)$.

The indicators are manifest variables that permit identification of the parameters present in the distribution of the latent variable EC. Given our assumptions, the joint probability $P(y_{in} = 1, I_n) \equiv P_n(i, I)$ of observing jointly the choice y_n and the indicators I_n may be written as:

$$P_n(i, I|X_n, w_n, \delta) = \int_{\mathrm{EC}} P_n(i\,|\mathrm{EC}, X_n, \theta) f(I_n|\mathrm{EC}, \Lambda) g(\mathrm{EC}\,|w_n, b) d\mathrm{EC}, \tag{4.6}$$

where $f(I_n|\mathrm{EC}, \Lambda)$ is the density of I_n implied by equation (**??**). The term δ designates the full set of parameters to estimate (i.e. $\delta = \{\theta, b, \Lambda\}$).

As I discussed in chapter 2, including numerous attitudes in HCMs 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 a valid simulator for the choice probability through maximum simulated likelihood (MSL) estimation (**?**). HCM classical full information estimation requires maximizing the log likelihood function: $\sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i, I\,|X_n, w_n, \delta\,)$. In practice, with a large number of latent variables we need to replace the multidimensional integral with a smooth simulator with good statistical properties, leading to a maximum simulated likelihood (MSL) solution (**??**). Although feasible, 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. Even though classical estimation of HCMs is possible using a sequential approach (**???**), the results of this method are not efficient (**?**).

For these reasons, we propose to go beyond classical methods by introducing Bayesian techniques. Building on the rapid development of Markov Chain Monte Carlo (MCMC) techniques, and on the idea that Bayesian tools (with appropriate priors) can 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. This chapter represents the first step in developing a Bayesian estimator for HCMs, specifically for the vehicle purchase context I have introduced.

4.4.2 HCM Gibbs sampler implementation

The parameters to estimate in the private vehicle choice case we are analyzing are $\theta' = [ASC_{AFV}\ ASC_{HEV}\ ASC_{HFC}\ \beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \Gamma_{\mathrm{EC},AFV}\ \Gamma_{\mathrm{EC},HEV}\ \Gamma_{\mathrm{EC},HFC}]$, b and Λ. Bayes estimation implementation for these parameters requires making draws from the joint posterior distribution:

$$P(\theta, b, \Lambda | y, I), \tag{4.7}$$

or, using data augmentation from:

$$P(\mathrm{EC}, \theta, b, \Lambda | y, I), \tag{4.8}$$

where $\mathrm{EC} = (\mathrm{EC}_1, \ldots, \mathrm{EC}_N)'$, $y = (y_1, \ldots, y_N)'$ 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:

$$\begin{aligned}
\mathrm{EC}_n^{(g)} &\sim \pi(\mathrm{EC}_n | \theta^{(g-1)}, b^{(g-1)}, \Lambda^{(g-1)}, y_n, I_n), \forall n && (4.9)\\
b^{(g)} &\sim \pi(b | \mathrm{EC}^{(g)}, \theta^{(g-1)}, b^{(g-1)}, y, I) && (4.10)\\
\Lambda^{(g)} &\sim \pi(\Lambda | \mathrm{EC}^{(g)}, \theta^{(g-1)}, b^{(g)}, y, I) && (4.11)\\
\theta^{(g)} &\sim \pi(\theta | \mathrm{EC}^{(g)}, b^{(g)}, \Lambda^{(g)}, y, I). && (4.12)
\end{aligned}$$

Since the latent variable EC is not observable, we need to incorporate the information provided by the indicator I_n on EC. This information is explicitly given by the conditional probability $\pi(\mathrm{EC}|I_n)$ whose expression depends on the assumptions we make. We assume then a multivariate normal distribution:

$$\begin{bmatrix} \mathrm{EC}_n \\ I_n \end{bmatrix} \sim N\left(\begin{bmatrix} w_n b \\ \Lambda w_n b \end{bmatrix}, \begin{bmatrix} 1 & \Lambda' \\ \Lambda & \Lambda\Lambda' + \mathrm{I}_{14} \end{bmatrix} \right), \forall n, \tag{4.13}$$

where I_{14} represents the identity matrix of size 14. Equation **??** implies

$$\pi(\mathrm{EC}|\theta, b, \Lambda, y_n, I_n) \sim MVN(\mu_{\mathrm{EC}_n|I_n}, \sigma^2_{\mathrm{EC}_n|I_n}), \forall n, \tag{4.14}$$

where

$$\begin{aligned}
\mu_{\mathrm{EC}_n|I_n} &= w_n b + \Lambda' \left[\Lambda\Lambda' + \mathrm{I}_{14}\right]^{-1} \left[I_n - \Lambda w_n b\right] && (4.15)\\
\sigma^2_{\mathrm{EC}_n|I_n} &= 1 - \Lambda' \left[\Lambda\Lambda' + \mathrm{I}_{14}\right]^{-1} \Lambda. && (4.16)
\end{aligned}$$

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

When using data augmentation, the latent variable EC becomes observable through $\pi(\mathrm{EC}|\theta, b, \Lambda, y_n, I_n)$. This fact implies that the conditional distributions for b and Λ simply correspond to ordinary Bayesian regressions (b and Λ are assumed independent):

$$\begin{aligned}
\pi(b|\mathrm{EC}, \theta, b, \Lambda, y, I) &\sim N(\bar{b}, \bar{V}_b) && (4.17)\\
\pi(\Lambda|\mathrm{EC}, \theta, b, y, I) &\sim N(\bar{\Lambda}, \bar{V}_\Lambda). && (4.18)
\end{aligned}$$

If prior beliefs for b and Λ are described by $p(b) \sim N(\check{b}, \check{V}_b)$ and $p(\Lambda) \sim N(\check{\Lambda}, \check{V}_\Lambda)$ respectively, then I can show that

$$\begin{aligned} \bar{V}_b &= (\check{V}_b^{-1} + w'w)^{-1}, \;\; \bar{b} = \bar{V}_b(\check{V}_b^{-1} + w'\mathrm{EC}) && (4.19) \\ \bar{V}_\Lambda &= (\check{V}_\Lambda^{-1} + \mathrm{EC}'\mathrm{EC})^{-1}, \;\; \bar{\Lambda} = \bar{V}_\Lambda(\check{V}_\Lambda^{-1} + \mathrm{EC}'I). && (4.20) \end{aligned}$$

4.4.3 The discrete choice kernel

The analytical form of the conditional distribution $\pi(\theta|\mathrm{EC}, b, \Lambda, y_n, I_n)$ for the discrete choice kernel depends on the assumptions regarding the distribution of the random term υ_n defined in equation (**??**).

We can derive a probit kernel if we make the assumption that the error terms υ_n are multivariate normal distributed. When using classical techniques, the burdensome classical multinomial probit estimation process reduces the practicability of the standard probit model. In fact, simulated maximum likelihood estimation (SML) of mixed logit models outperforms the SML estimation of probit because of the good statistical properties that can be derived for the former estimator (**?**). However, Bayesian methods breaks down the complexity of classical estimation of the probit model (**?**). HCM Bayesian estimation with a probit kernel is straightforward because the properties of the normal distribution allow us to exploit data augmentation techniques, basically because the utility function follows a normal distribution (**??**).

In discrete choice models, decisions are based on utility differences, so we can consider a utility difference model with respect to the base alternative SGV that, in our particular case, leads us to write the structural equation in a system of three equations:

$$\widetilde{U}_n = \widetilde{V}_n + \widetilde{\upsilon}_n, \quad \widetilde{\upsilon} \sim MVN(0, \mathrm{I}_3), \tag{4.21}$$

where I_3 is the identity matrix of size 3 – which is the number of alternatives for the utility difference model. We can get the vector form of the model by stacking the individual utilities into $\widetilde{U} = \widetilde{V} + \widetilde{\upsilon}$, where $\widetilde{\upsilon} \sim MVN(0, \mathrm{I}_{(N\times 3)})$.

Note that to obtain the differenced model in its estimable form (equation **??**) from equation (**??**), we use the matrix difference operator $\Delta_{SGV}(\cdot)_{jn} = (\cdot)_{jn} - (\cdot)_{SGVn}, j = \{AFV, HEV, HFC\}$:

$$\Delta_{SGV} U_n = \Delta_{SGV} V_n + \Delta_{SGV} \upsilon_n, \tag{4.22}$$

where $\Delta_{SGV}V_n$ denotes the (3×1) extended deterministic part of the differenced utility expression for individual n, composed by $\Delta_{SGV}V_{jn} = ASC_j + \beta_1\Delta_{SGV}X_{1jn} + \cdots + \beta_5\Delta_{SGV}X_{5jn} + \Gamma_{\mathrm{EC},j}\mathrm{EC}_n \equiv \widetilde{\mathrm{X}}_{jn}\theta, j = \{AFV, HEV, HFC\}$, where $\widetilde{\mathrm{X}}_{jn}$ is a row vector that contains the incremental specification of the attributes of alternative j (attribute changes with respect to the base alternative's values) and the latent variable EC, and where θ is a column vector of unknown parameters. The matrix $\widetilde{\mathrm{X}}$ is built by stacking the vectors $\widetilde{\mathrm{X}}_{jn}$ for each alternative j and each individual n. The assumption for a probit model is that $\upsilon_n \sim MVN(0_{4\times4}, \Sigma_{4\times4})$, and so we have $\Delta_{SGV}\upsilon_n \sim MVN(0_{3\times3}, \Omega_{3\times3} = [\Delta_{SGV}\Sigma\Delta'_{SGV}])$. Let L be the Cholesky root of Ω^{-1}. Then, the model can be reexpressed as:

$$L'\Delta_{SGV}U_n = L'\Delta_{SGV}V_n + L'\Delta_{SGV}\upsilon_n, \tag{4.23}$$

leading to equation (**??**) above.

Assuming that $\widetilde{U}$ is observable, that Ω is known, and that prior beliefs for θ are described by $p(\theta) \sim N(\check{\theta}, \check{V}_\theta)$, then the model becomes a simple Bayesian regression with known variance (**??**):

$$\pi(\theta|\widetilde{U}, \Omega, \mathrm{EC}, b, \Lambda, y, I) \sim N(\bar{\theta}, \bar{V}_\theta), \tag{4.24}$$

$$\bar{V}_\theta = (\check{V}_\theta^{-1} + \widetilde{\mathrm{X}}'\widetilde{\mathrm{X}})^{-1}, \;\; \bar{\theta} = \bar{V}_\theta(\check{V}_\theta^{-1} + \widetilde{\mathrm{X}}'\widetilde{U}). \tag{4.25}$$

In the conditional distribution of θ, $\pi(\theta|\widetilde{U}, \Omega)$ indicates the use of data augmentation techniques. This is possible if and only if the conditional distributions of $\widetilde{U}$ and Ω are easy to describe. Note that $\widetilde{U}_{AFVn} \sim N(\widetilde{V}_{AFVn}, 1)$, $\widetilde{U}_{HEVn} \sim N(\widetilde{V}_{HEVn}, 1)$, and $\widetilde{U}_{HFCn} \sim N(\widetilde{V}_{HFCn}, 1)$. However, since $y_n = i \iff U_{in} = max(U_{SGVn}, U_{AFVn}, U_{HEVn}, U_{HFCn})$ then conditional on y_n, $\widetilde{U}_n$ follows a truncated multivariate normal (TMVN) distribution:

$$\pi(\widetilde{U}_n|\mathrm{EC}, \Omega, \theta, b, \Lambda, y_n, I_n) \sim \mathrm{TMVN}_{\Re|y_n}\left(\widetilde{V}_n, \mathrm{I}_3\right), \forall n, \tag{4.26}$$

where the truncation region $\Re$ is defined by the measurement equation of y_n:

$$y_n = \begin{cases} SGV & \text{if } (\widetilde{U}_{AFVn} < 0) \wedge (\widetilde{U}_{HEVn} < 0) \wedge (\widetilde{U}_{HFCn} < 0) \\ AFV & \text{if } (\widetilde{U}_{AFVn} \geq 0) \wedge (\widetilde{U}_{AFVn} > \widetilde{U}_{HEVn}) \wedge (\widetilde{U}_{AFVn} > \widetilde{U}_{HFCn}) \\ HEV & \text{if } (\widetilde{U}_{HEVn} \geq 0) \wedge (\widetilde{U}_{HEVn} > \widetilde{U}_{AFVn}) \wedge (\widetilde{U}_{HEVn} > \widetilde{U}_{HFCn}) \\ HFC & \text{if } (\widetilde{U}_{HFCn} \geq 0) \wedge (\widetilde{U}_{HFCn} > \widetilde{U}_{AFVn}) \wedge (\widetilde{U}_{HFCn} > \widetilde{U}_{HEVn}) \end{cases}. \tag{4.27}$$

Finally, if prior beliefs for Ω are described by an inverted-Wishart $IW(\check{\nu}, \check{V})$ distribution, then it can be shown that

$$\pi(\Omega|\mathrm{EC}, \widetilde{U}_n, \theta, b, \Lambda, y_n, I_n) \sim IW(\bar{\nu}, \bar{V}), \tag{4.28}$$

where $\bar{\nu} = \check{\nu} + N$ and $\bar{V} = \check{V} + \sum_{n=1}^{N}(\Delta_{SGV}\upsilon_n)(\Delta_{SGV}\upsilon_n)'$. This conditional distribution completes the set of distributions needed for Gibbs sampling with a probit kernel.

Note that it is uncomplicated to make an extension of the probit-kernel Gibbs sampler we developed to a normal error component model – such as randomly normal distributed taste variations with a probit kernel.

On the other hand, when modeling a normal error component model with a multinomial logit kernel – which results in a mixed logit or MMNL model – we no longer have the advantageous properties that make implementing the probit-kernel Gibbs sampler straightforward. Since the MMNL distribution of $\widetilde{U}$ is hard to describe, i.e. there is no closed form full conditional distribution for $\widetilde{U}$, data augmentation implementation for the utility function is no longer as simple. Thus, MMNL Bayesian estimation does not allow us to use a simple regression for θ and, as we will show, the use of Metropolis-Hastings (MH) methods is needed. The Bayesian procedure for a standard MMNL – without an associated structure of latent variables – described by **?** can be simply plugged into the HCM Gibbs sampler. If we focus on the normal random taste parameters case, then

$$U_n = X_n\beta_n + \Gamma_n \cdot \mathrm{EC}_n + \upsilon_n \equiv \widetilde{X}_n\theta_n + \upsilon_n, \tag{4.29}$$

where $\widetilde{X}$ is the extended matrix of the alternative attributes, including the latent variable EC; $\theta_n \sim MVN(\theta, \Sigma_\theta)$ is the unknown vector of randomly distributed taste parameters, with θ representing the population mean; and υ_n is a vector of independent and identically distributed extreme value type 1 error terms.

Following **?**, if prior beliefs for θ and Σ_θ are described by $p(\theta, \Sigma_\theta) = p(\theta)p(\Sigma_\theta)$, where $p(\theta) \sim N(\check{\theta}, \check{\Sigma})$ with extremely large variance, and $p(\Sigma_\theta)$ is inverted-Wishart $IW(\check{\nu_\Sigma}, \check{V_\Sigma})$, then the mixed logit kernel HCM Gibbs sampler is completed considering the following conditional posteriors:

$$\pi(\theta|\mathrm{EC}, \theta_n, \Sigma_\theta, b, \Lambda, y_n, I_n) \sim MVN(\bar{\theta}, \Sigma_\theta/N) \tag{4.30}$$

$$\pi(\Sigma_\theta|\mathrm{EC}, \theta, \theta_n, b, \Lambda, y_n, I_n) \sim IW\left(\check{\nu_\Sigma} + N, \frac{\check{\nu_\Sigma}\check{V_\Sigma} + N\bar{\Sigma}}{\check{\nu_\Sigma} + N}\right), \tag{4.31}$$

where $\bar{\theta} = \sum \theta_n/N$, $\bar{\Sigma} = \sum(\theta_n - \theta)(\theta_n - \theta)'/N$; and

$$\pi(\theta_n|\mathrm{EC}, \theta, \Sigma_\theta, b, \Lambda, y_n, I_n) \propto \frac{e^{\widetilde{X}_{y_n n}\theta_n}}{e^{\widetilde{X}_{SGVn}\theta_n} + e^{\widetilde{X}_{AFVn}\theta_n} + e^{\widetilde{X}_{HEVn}\theta_n} + e^{\widetilde{X}_{HFCn}\theta_n}}\varphi(\theta_n|\theta, \Sigma_n), \forall n \tag{4.32}$$

where $\varphi(\theta_n|\theta, \Sigma_n)$ is the normal density function. Note that the Metropolis-Hastings algorithm is needed in order to draw θ_n from the distribution in equation **??**. We will describe the MH algorithm for the multinomial logit (MNL) model. The MNL model is a particular case of the mixed logit model, where the taste parameters are fixed to the population means. In the MNL case, we fail to find a closed form full conditional distribution for θ. However, we can use an asymptotic approximation to the posterior (**?**):

$$\pi(\theta|\text{EC}, b, \Lambda, y_n, I_n) \propto |H|^{\frac{1}{2}} \exp\left(\frac{1}{2}(\theta - \hat{\theta})' H(\theta - \hat{\theta})\right), \tag{4.33}$$

with $\hat{\theta}$ being the maximum likelihood solution for θ, and with H being the asymptotic variance obtained from the expected sample information matrix ($\otimes$ denotes the Kronecker product):

$$H = -E\left[\frac{\partial^2 \ln l}{\partial\theta\partial\theta'}\right] = -\sum_{n=1}^{N}(\text{diag}(P_n) - P_n P_n') \otimes \widetilde{X}_n \widetilde{X}_n', \tag{4.34}$$

which is the Hessian matrix of the observed MNL log-likelihood $\ln l = \sum_{n=1}^{N} \ln P_{y_n}$, where $P_n = (P_{SGVn}, P_{AFVn}, P_{HEVn}, P_{HFCn})$, and with P_{in} below being the standard MNL form of the choice probability of alternative i for individual n:

$$P_{in} = \frac{e^{\widetilde{X}_{in}\theta_n}}{e^{\widetilde{X}_{SGVn}\theta_n} + e^{\widetilde{X}_{AFVn}\theta_n} + e^{\widetilde{X}_{HEVn}\theta_n} + e^{\widetilde{X}_{HFCn}\theta_n}}. \tag{4.35}$$

For Metropolis-Hastings implementation, a candidate θ^{cand} is drawn from a given distribution depending on whether we are using a random walk chain or an independence chain (**?**). The candidate realization θ^{cand} is then compared to the current θ^{curr} through:

$$\alpha = \min\left\{1, \frac{l(\theta^{cand}|y, \tilde{X})\pi(\theta^{cand})}{l(\theta^{curr}|y, \tilde{X})\pi(\theta^{curr})} \times \frac{q(\theta^{cand}, \theta^{curr})}{q(\theta^{curr}, \theta^{cand})}\right\}, \tag{4.36}$$

where $q(i, j)$ is the probability of generating candidate j given i. The candidate is accepted as the new $\theta^{curr} = \theta^{cand}$ with probability α, while the old one is preserved $\theta^{curr} = \theta^{curr}$ with probability $1 - \alpha$. By plugging this MH procedure into the Gibbs sampler developed in the previous section for b and Λ, we obtain a Bayesian MNL solution for the full set of parameters to estimate.

4.5 Application to vehicle choice data

In the previous section I discussed how to methodologically implement Bayesian estimation using different models for the discrete choice kernel. Although the probit kernel formulation is analytically straightforward, taking draws from a truncated multivariate normal distribution is necessary. For our empirical application we implemented an MNL kernel to avoid convergence problems due to truncation. HCM estimation with an MNL kernel requires an MH-within-Gibbs algorithm that does not make use of draws from a multivariate normal distribution. Thus, we do not expect a slowed-down estimation process because of rejection methods for truncation. Additionally, implementation of an MNL kernel also facilitates the comparison of our results with models estimated previously using the same data, namely classical estimation of an HCM with an MNL kernel to incorporate environmental preferences (**?**), as well as the standard discrete choice model originally calibrated after the survey (**?**).

I will now present the results of the HCM Bayesian estimation process for the vehicle choice data. Using the R language, I implemented the MNL Kernel Gibbs sampling routine presented earlier:

$$\pi(\theta|\mathrm{EC}, b, \Lambda, y, I) \propto |H|^{\frac{1}{2}} \exp\left(\frac{1}{2}(\theta - \hat{\theta})' H (\theta - \hat{\theta})\right) \tag{4.37}$$

$$\pi(\mathrm{EC}_n|\theta, b, \Lambda, y_n, I_n) \sim N(\mu_{\mathrm{EC}_n|I_n}, \sigma^2_{\mathrm{EC}|I}), \forall n \tag{4.38}$$

$$\pi(b|\mathrm{EC}, \theta, b, \Lambda, y, I) \sim N(\bar{b}, \bar{V}_b) \tag{4.39}$$

$$\pi(\Lambda|\mathrm{EC}, \theta, b, y, I) \sim N(\bar{\Lambda}, \bar{V}_\Lambda). \tag{4.40}$$

To construct the reported results I considered 5,000 draws – or iterations of the Gibbs sampler sequence – with a burn-in period of the first 500 draws. The mean of the Gibbs sampler draws is a consistent estimator for the posterior mean of the parameters of interest. Recall that under fairly weak conditions (**?**), the Gibbs sampler sequence of random draws forms an irreducible and ergodic Markov chain representing the joint posterior distribution. For this application, we adopt diffuse priors. In addition, the standard deviations used for the calculation of t-statistics are simply the standard deviations of the artificial samples generated by the Gibbs sampler. 5,000 draws (4,500 without the burn-in period) appear to be enough to reproduce the maximum likelihood results with a fair degree of accuracy. In fact, to test whether we achieved convergence we made several trials, including broken MCMC chains (for instance we tried 25,000 draws with thinning parameter k=5); using more draws or breaking the MCMC chain we recovered the same results. To give an idea from a similar context, note that in the case of mixed logit, **?** find that 5,000 draws appear as a good number to assure convergence even in the presence of serial correlation. The total time taken for parameter estimation was 120 minutes in an ordinary PC (cf. 90 minutes for classical estimation; note however that the processing time for the Bayesian approach is for the whole distribution of the parameters and not for just the point estimates, as is the case of the classical approach. Also, the processing time for the Bayesian approach includes the calculations needed for prediction).

Although the estimation process implies that all the equations are calibrated simultaneously, I 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. Since this is the first application of MCMC methods to a hybrid choice model, I first focus on the results of the estimated parameters using diffuse priors. We also present the results of a classical HCM with an MNL kernel (**?**).

4.5.1 Car Choice Model

First, I present the results of the car choice model (Table **??**.) As explained before, the car choice corresponds to the (MNL) discrete choice kernel, where the parameters to estimate are described by the taste parameter vector θ of the utility function. The deterministic utility contains the experimental attributes purchase price, fuel cost, fuel availability, express lane access, and 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 variable EC. The latent variable related to environmental concern (EC) was not considered for the standard gasoline vehicle SGV.

Unsurprisingly – and yet, reassuringly – Gibbs sampling and classical maximum likelihood parameters have the same sign and magnitude. The environmental concern latent variable enters very significantly and positively into the choice model specification. Thus, environmental concern (EC) encourages the choice of green automobile technologies through a positive impact in the choice probability of those alternatives. 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 HFC represents the cleanest engine technology of the experimental alternatives. The fact that HEV still makes use of standard fuel could explain the lower EC impact.

It is important to mention that my results for both the Bayesian and classical HCM to some extent reproduce the results of an MNL (without latent variables **??**): common parameters with the standard multinomial logit model have the same sign and magnitude, except for alternative specific constants (which now are affected by the inclusion of the

Car Choice Model	Bayesian Estimates		Classical Estimates	
	estimates	**t-stat**	**estimates**	**t-stat**
ASC_AFV	-6.185	-7.52	-6.189	-9.73
ASC_HEV	-2.530	-3.67	-2.541	-4.43
ASC_HFC	-4.049	-5.66	-4.093	-7.82
Purchase Price	-0.895	-4.21	-0.894	-4.22
Fuel Cost	-0.852	-4.27	-0.854	-4.18
Fuel Availability	1.388	7.42	1.398	7.31
Express Lane Access	0.158	2.26	0.160	2.26
Power	2.729	4.01	2.752	4.13
Latent Variables				
EC on AFV	0.585	3.68	0.592	4.09
EC on HEV	0.411	4.88	0.420	4.45
EC on HFC	0.674	7.37	0.692	6.95
Number of pseudo-individuals	1877		1877	
Number of draws (Burn-in)	5000 (500)		-	
Number of Halton draws	-		500	
Number of iterations	-		326	
Loglikelihood	-1955.34		-1987.52	
Adjusted ρ^2	0.249		0.236	

Table 4.2: Car Choice Model Results

latent variable). It is especially interesting to note that convergence is assured for the maximum likelihood estimation of the standard MNL. Thus, because of the MNL kernel assumption I can take the MNL estimates as 'reference values' for an informal test not only for assuring that the global maximum is achieved (classical estimation), but also for convergence of the Gibbs sampler I have implemented. In fact, since I used diffuse priors, the informal test of convergence – set as reproducing the classical estimates with a certain degree of accuracy – seems appropriate. Also note that the starting values were not data-based. The results presented were calculated using starting values set to zero, and I checked independence of the results and the starting values used for both Bayesian and classical estimation (for the latter, in order to check that a global maximum was attained).

To give an idea of the posterior distribution of θ I also present the Bayesian estimates for the quantiles needed to construct confidence intervals for the parameters.

	Quantiles				
	2.5%	5%	50%	95%	97.5%
ASC_AFV	-7.88	-7.56	-6.18	-4.83	-4.53
ASC_HEV	-3.95	-3.69	-2.53	-1.38	-1.13
ASC_HFC	-5.47	-5.24	-4.06	-2.86	-2.59
Purchase Price	-1.34	-1.26	-0.89	-0.55	-0.47
Fuel Cost	-1.27	-1.19	-0.85	-0.53	-0.43
Fuel Availability	1.01	1.08	1.39	1.70	1.77
Express lane access	0.01	0.04	0.16	0.27	0.30
Power	1.34	1.58	2.74	3.86	4.16
Latent Variables					
EC on AFV	0.26	0.32	0.58	0.86	0.91
EC on HEV	0.24	0.27	0.41	0.55	0.59
EC on HFC	0.48	0.52	0.67	0.83	0.87

Table 4.3: Car Choice Model - Bayesian Quantiles

4.5.2 Structural Model

The structural equation links consumer characteristics with the latent variables through a linear regression equation based on the usual mode of transportation (driving, carpooling or public transportation) either to commute (in the case of workers) or for other main purposes for the rest of the sample, the individual's gender, age, education level, and household income. The estimation results are shown in Table **??**:

Structural Model	**Bayesian Estimates**		**Classical Estimates**	
	est	**t-stat**	**est**	**t-stat**
Intercept	1.840	4.63	2.067	7.08
Driving Alone User	-0.157	-2.23	-0.143	-1.86
Carpool User	0.236	2.15	0.241	1.72
Transit User	0.482	4.76	0.468	3.92
Female Indicator	0.344	6.06	0.342	5.52
High Income Indicator (>80K$)	0.046	0.77	0.050	0.75
University Indicator	0.274	4.62	0.285	4.41
Age level: 26-40 years	0.447	3.96	0.439	3.35
Age level: 41-55 years	0.544	4.81	0.538	4.07
Age level: 56 years & above	0.839	6.70	0.829	5.79

Table 4.4: Structural Model Results

From this model, I can conclude that environmental concern (EC) is more important for public transportation users than for carpool users. We in fact observe a negative parameter for those who mostly drive alone. The results are in line with the idea that regular drivers may be unaware of the environmentally adverse effects of private car use (air pollution and congestion). Good public transportation service has been proposed as an alternative for car use reduction; our results show that transit users are more green with regard to the adoption of new transportation technologies.

I also find that concern about environmental issues in the car purchase choice context is more developed in women, older people and more educated people (cf. **?**). The effect of the high income variable is positive but not significant.

Since each respondent offers up to four SP vehicle-choices, we have repeated individuals in the sample. The structural equation implies that the problem of correlation between observations is addressed indirectly by the individual-specific latent variable via the socio-demographic variables. In effect there is no variation in these socio-demographic variables for a single individual's choice exercise, but there is variation among different groups of individuals. Only people who belong to the same cluster (defined by equal socio-demographic characteristics) will have a common variable (the deterministic part of the latent variable) that does not vary across choice situations. However I do recognize that in this application the structural model for the latent variable assumes independent error terms, even for different responses of the same individual. To address this issue it is possible to assume a common latent variable parameter that varies across individuals. This approach translates into incorporating exactly the same random draw of the latent variable for each choice exercise of a same individual. I tested this specification and the results were not significantly different from zero (implying that the underlying cluster classification was enough to address the problem of repeated choices).

Finally, I present the Bayesian quantile estimation:

4.5.3 Measurement Model

Lastly, several indicators were considered in the latent variable measurement model, which links the latent psychometric environmental concern variable to answers to attitudinal/perceptual qualitative survey questions. The questions selected to define the indicator variables concern the respondent's level of support for or opposition to various transport policies (*Transport Policies Support*), and their opinions on various transport-related issues (*Transport Problems Evaluation*). The results are shown in Table **??**.

	Quantiles				
	2.5%	5%	50%	95%	97.5%
Intercept	1.04	1.15	1.87	2.44	2.52
Driving Alone User	-0.29	-0.27	-0.16	-0.04	-0.02
Carpool User	0.02	0.06	0.24	0.42	0.45
Transit User	0.28	0.31	0.48	0.65	0.68
Female Indicator	0.23	0.25	0.34	0.44	0.46
High Income Indicator (>80K$)	-0.07	-0.05	0.05	0.14	0.17
University Indicator	0.16	0.18	0.27	0.37	0.39
Age level: 26-40 years	0.23	0.26	0.45	0.63	0.67
Age level: 41-55 years	0.32	0.36	0.54	0.73	0.77
Age level: 56 years & above	0.59	0.63	0.84	1.05	1.08

Table 4.5: Structural Model - Bayesian Quantiles

Measurement Model	**Bayesian Estimates**		**Classical Estimates**	
	estimates	**t-stat**	**estimates**	**t-stat**
Transport Policies Support				
Expanding & Upgrading Roads	-0.358	-12.56	-0.375	-13.77
Road Tolls & Gas Taxes	0.541	17.40	0.547	20.05
Bike Lanes & Speed Controls	0.339	13.29	0.344	8.85
Regular Testing for Reducing Car Emissions	0.277	10.86	0.283	8.01
High Occupancy Vehicles & Transit Priorities	0.426	16.08	0.426	11.63
Improving Transit Service	0.278	11.49	0.278	7.08
Promoting Compact Communities	0.257	8.77	0.250	10.77
Encouraging Short Work Weeks	0.234	9.14	0.230	7.67
Transport Problems Evaluation				
Traffic Congestion	0.365	12.01	0.355	13.20
Traffic Noise	0.575	19.17	0.569	18.28
Poor Local Air Quality	0.649	23.26	0.655	15.04
Accidents Caused by Bad Drivers	0.313	11.99	0.305	8.71
Emissions & Global Warming	0.445	17.32	0.446	9.32
Speeding Drivers in Neighborhoods	0.472	18.02	0.466	13.57

Table 4.6: Measurement Model

As explained previously, this model measures the effect of the latent variable on each indicator. While indicator variables permit identification of the latent variables and provide efficiency in estimating the choice model with latent variables (indicators add information content), at the same time some interesting conclusions can be drawn from the estimations. For instance, 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 environmentally-conscious consumers negatively perceived the priority given to cars by policies aimed at raising road capacity because of the negative impact on the environment. Expansion of the road network is not environmentally sustainable for city development, and our results show that green consumers are aware of this problem.

In addition, we see that the effect of environmental concern EC on the indicator related to support for applying road tolls and gas taxes is positive, indicating a perceived positive environmental impact of measures allowing for a presumably more rational use of private vehicles. A similar analysis can be done for the remaining indicator variables – all of them with a significant positive impact – with the corresponding effect of encouraging sustainable transport. For example, the positive sign of the effect of EC on support for reducing vehicle emissions with regular testing and manufacturer emission standards; the perception of poor local air quality motivating the adoption of green vehicles; and the encouragement of the expansion of the bicycle path network.

Note that according to the results *poor local air quality* is a major problem (respondents' opinions about this issue in our model weigh higher than other transport problems). At the same time, the variable *road tolls and gas taxes* has the highest weight among transport policies. Considering both results we can identify carbon pricing as an efficient instrument to encourage the adoption of low-emission vehicles.

We can also see that the effect of other indicators that may seem conceptually unrelated to environmental preferences do not have lower coefficients when compared to more traditional indicators. For instance, the correlation between EC and *speeding drivers in neighborhoods*, clearly a concept related to safety, is almost the same as the correlation between EC and concerns about *emissions and global warming*. Even though the alternatives in our model are differentiated by their impacts on the natural environment, as we mentioned earlier, the EC latent variable reflects concerns about the adverse effects of personal transportation on both the natural (e.g. *emissions and global warming*) and the mobility (e.g. *speeding drivers in neighborhoods*) environments. Other concepts affect both, such as *traffic congestion* (reflecting indiscriminate car use with corresponding externalities such as higher emission levels produced at low speeds), *noise* (that can be viewed as an externality of traffic congestion), *promoting compact communities* (implying reduced distances and therefore less emissions) and *encouraging short work weeks* (through a reduction of transportation needs). The derived correlation structure is a posterior justification of the unidimensionality of the EC variable.

	Quantiles				
	2.5%	5%	50%	95%	97.5%
Transport Policies Support					
Expanding & Upgrading Roads	-0.42	-0.41	-0.36	-0.31	-0.30
Road Tolls & Gas Taxes	0.48	0.49	0.54	0.59	0.60
Bike Lanes & Speed Controls	0.29	0.30	0.34	0.38	0.39
Regular Testing for Reducing Car Emissions	0.23	0.24	0.28	0.32	0.33
High Occupancy Vehicles & Transit Priorities	0.37	0.38	0.43	0.47	0.48
Improving Transit Service	0.23	0.24	0.28	0.32	0.32
Promoting Compact Communities	0.20	0.21	0.26	0.30	0.31
Encouraging Short Work Weeks	0.18	0.19	0.23	0.28	0.29
Transport Problems Evaluation					
Traffic Congestion	0.31	0.32	0.37	0.42	0.42
Traffic Noise	0.52	0.53	0.57	0.63	0.64
Poor Local Air Quality	0.60	0.60	0.65	0.70	0.71
Accidents Caused by Bad Drivers	0.26	0.27	0.31	0.36	0.36
Emissions & Global Warming	0.40	0.40	0.45	0.49	0.49
Speeding Drivers in Neighborhoods	0.42	0.43	0.47	0.52	0.52

Table 4.7: Measurement Model - Bayesian Quantiles

In sum, using real data about virtual personal vehicle choices I have shown that HCM is genuinely capable of adapting to practical situations. HCM combines the direct effect of environment-related underlying latent variables on the private vehicle choice probabilities with the socio-demographic characteristics of the consumers that enter the choice probabilities through the environmental concern latent variable. HCM also takes into account opinions and attitudes through the consumer's response to attitudinal environment-related rating exercises. Finally, these responses are taken as indicators of the environmental concern latent variable.

4.5.4 Forecasting

For forecasting, we have to consider the results of both the discrete choice kernel and the structural model. The choice model explains behavior and the structural model not only serves to build clusters of consumers, but also to predict values of the unobserved

EC variable necessary for the choice model. Even though indicators are necessary for identification of the latent variable, for forecasting there is no need for the latent variable measurement model.

Forecasting with discrete choice models is a question of consumers' trade-offs produced by changes in the values of the attributes. The first step in understanding these trade-offs is to derive willingness to pay (WTP) values from the estimates of the discrete choice kernel. Although the parameters associated with each attribute in the discrete choice kernel represent marginal utilities, since the utility function is ordinal it is hard to interpret the estimates of the model. However, the ratios of the parameters represent marginal rates of substitution that provide information about the trade-offs being made. For instance, WTPs correspond to marginal rates of substitution of some characteristics and price, in this case how much additional money the consumer is willing to pay to purchase a particular car given the increase (decrease) of an attribute that provides a higher (dis-)utility level while keeping the same level of satisfaction. In Table **??** I report the WTPs obtained from the model. A negative sign represents the amount of money [CAD/10000] that the consumer is willing to pay for the increase of one unit of an attribute that raises the general utility level, while a positive sign indicates the expected reduction in price for the increase of an attribute that decreases the utility level (or the willingness to pay for a reduction in one unit of that particular attribute). For instance, on average a consumer is willing to pay $166 for an increase of 1% of the service network density (cf. **?**). Note that I am presenting not only the mean WTPs but also the WTPs' standard deviations and quantile estimates. The distribution of the WTPs is a direct result of Bayesian estimation, whereas the estimation of confidence intervals for WTP is particularly tricky when using classical techniques (see **?**).

WTP [CAD/10000-unit]	mean	s.e.	**Quantiles** 2.5%	5%	50%	95%	97.5%
Fuel Cost	1.018	0.42	0.45	0.53	0.96	1.72	2.00
Fuel Availability	-1.660	0.59	-3.01	-2.63	-1.56	-1.02	-0.92
Express lane access	-0.189	0.11	-0.44	-0.37	-0.18	-0.05	-0.02
Power	-3.258	1.35	-6.39	-5.43	-3.08	-1.63	-1.35
Latent Variables							
EC on AFV	-0.701	0.32	-1.42	-1.21	-0.66	-0.33	-0.27
EC on HEV	-0.492	0.19	-0.97	-0.82	-0.46	-0.27	-0.23
EC on HFC	-0.807	0.29	-1.49	-1.29	-0.76	-0.50	-0.45

Table 4.8: Willingness to pay - Bayesian Quantiles

An interesting exercise is to derive the capital-cost equivalency from the results of the WTPs, i.e. how much (or less) of each attribute would be equal to an increase of $1000 in purchase price (see Table **??**, where I also present the original equivalencies based on the MNL results by **?**). According to my results, if the cost of fuel is reduced in $9.82 per month, the consumer is willing to buy a new vehicle costing $1000 more. This measures a trade-off that is important for policy making: a reduction in taxes on alternative fuels (or an increase in taxes on fossil fuels) can compensate for higher prices of green technologies.

WTP [CAD/10000-unit]	Change equal to 1000 CAD increase in capital cost	
	HCM	**?**
Fuel Cost	-9.82 [CAD/month]	-19.59 [CAD/month]
Fuel Availability	6.02%	8.00%
Express lane access	53.00%	56.00%
Power	3.07%	4.00%

Table 4.9: Capital-cost equivalency for vehicle attributes

Since the measurement scale for the EC latent variable is unknown, it is hard to interpret the values obtained for WTPs related to environmental preferences. However, from the Bayesian estimates we can describe the density function for EC. For instance, EC has a mean of 2.687 [units], a standard deviation equal to 0.908, and maximum and minimum values equal to 4.891 and -0.017, respectively. In addition, we can compare the different degrees of EC given by the different clusters obtained in the structural model. Women are more environmentally concerned than men; the mean value of EC for women is 2.816, while it is equal to 2.513 for men. Using the WTPs obtained from the model, this difference implies that on average women are willing to pay more for green technologies than men are. In Table **??** we show the derived average marginal WTPs for women and drive-alone commuters (the latter showing a environmentally indifference tendency according to our model). For example, women are willing to pay $2440 more than men for an HFC vehicle. Drive-alone commuters are willing to pay $2486 less for an AFV than commuters who carpool or do not use private vehicles.

Green Vehicles	Average marginal **WTP**	
	Women	Drive-alone commuters
AFV	2119 [CAD]	-2486 [CAD]
HEV	1487 [CAD]	-1745 [CAD]
HFC	2440 [CAD]	-2862 [CAD]

Table 4.10: Average marginal willingness to pay for low-emission vehicles

Finally, I simulate the impact of different policies. It is important to mention first that the experimental market shares obtained from the survey differ from current conditions in the automotive market (see **?**). In fact, actual market shares show that green vehicles still have a small penetration. HFC technologies have not even been introduced into the market yet. Thus, the hypothetical market conditions for the baseline scenario can be interpreted as a future market where green technologies have been introduced and where the attributes for the different alternatives have reached levels comparable to those considered in the experimental design. I consider the following scenarios:

- *Baseline scenario*: experimental situation presented in the survey.
- *Scenario 1*: 100% fueling network density for every alternative.
- *Scenario 2*: 25% increase in fueling network density for green vehicles.
- *Scenario 3*: 25% tax on fossil fuel costs.
- *Scenario 4*: 10% reduction in purchase price for green vehicles.
- *Scenario 5*: 50% increase in purchase price for new technologies (HEV and HFC).
- *Scenario 6*: Augmentation in EC equal to its mean value.
- *Scenario 7*: Baseline considering only women.

Whereas in the case of classical estimation extra simulation of the choice probabilities is required in forecasting, when using Bayesian techniques we make use of the sample of draws generated by the Gibbs sampler for estimating the model. (The Gibbs sampler generates simulations from the unconditional posterior distribution for the parameters.) For each draw a predicted policy outcome is calculated; what we obtain is a sample of simulations for the predictive distribution of the effects of each scenario (**?**). From the sample of draws for each policy simulation we obtain the point estimates – the predicted average market share for each scenario – with standard deviations provided in Table **??**.

First, the baseline scenario (simulated market shares) replicates the known market shares of the SP experiment: this can be statistically assessed through the Chi-squared index $\chi^2 = 4.63 < \chi^2_{c,(95\%,3)} = 7.815$ (**?**).

Limited fuel availability for green vehicles is an important concern for consumers (**??**). Scenario 1 represents an ideal situation where the fueling station network is expanded to its maximum (set by the SGV fueling network). In this context, important differences in the market shares are obtained. AFVs and HFCs are the alternatives that benefit from the increase in fuel availability (the increase being in the range of 25%-75%) and the model predicts that the market shares of both alternatives would increase significantly

	Market Shares			
	SGV	AFV	HEV	HFC
Observed	11.35%	3.73%	48.85%	36.07%
Baseline	12.88%	3.75%	48.45%	34.92%
s.d.	0.73%	0.44%	1.13%	1.07%
Scenario 1: 100% fuel net	9.53%	5.16%	35.32%	49.99%
s.d.	0.69%	0.62%	1.98%	2.25%
Percent Change	-25.95%	37.41%	-27.10%	43.14%
Scenario 2: ↑ 25% fuel net for AFV & HFC	10.62%	4.30%	45.31%	39.80%
s.d.	0.74%	0.52%	1.23%	1.33%
Percent Change	-17.55%	14.67%	-6.48%	13.97%
Scenario 3: 25% tax on fossil fuel	11.80%	4.10%	46.13%	37.97%
s.d.	0.72%	0.52%	1.48%	1.59%
Percent Change	-8.32%	9.17%	-4.78%	8.72%
Scenario 4: ↓ 10% price of green vehicles	11.41%	3.79%	49.26%	35.55%
s.d.	0.75%	0.45%	1.15%	1.09%
Percent Change	-11.42%	0.92%	1.67%	1.79%
Scenario 5: ↑ 50% price of HEV & HFC	20.67%	6.20%	41.73%	31.41%
s.d.	3.33%	1.48%	2.88%	1.93%
Percent Change	60.53%	65.14%	-13.88%	-10.07%
Scenario 6: Social marketing campaign	3.87%	6.01%	41.65%	48.46%
s.d.	0.77%	1.47%	3.01%	3.19%
Percent Change	-69.97%	60.29%	-14.03%	38.78%
Base (women)	12.08%	3.90%	48.55%	35.47%
s.d.	0.71%	0.46%	1.15%	1.09%
Percent Change	-6.16%	3.98%	0.20%	1.57%

Table 4.11: Policy Scenarios - Predicted Market Shares

to the extent that the market shares of both SGVs and HEVs decline (hybrid vehicles share the same network as SGVs). Because Scenario 1 represents an extreme situation, I simulate scenario 2 where the fueling network for alternative fuels is expanded by 25%. Both scenarios show that green alternatives become more attractive to consumers when the fueling infrastructure is competitive (?).

Because of the environmental externalities caused by gasoline consumption, carbon pricing is increasingly considered by policy makers as a valid instrument to reduce oil dependency and as an appropriate response to deal with the problems causing global warming (?). Scenario 3 considers an augmentation in fossil fuel costs by 25%, simulating the impact of a gas tax policy – which is equivalent to a carbon emission tax. As expected, both SGVs and HEVs reduce their market shares. The impact of the fuel tax is higher for SGVs (a reduction equivalent to 8.32% compared to 4.78% for HEVs), simply because hybrid vehicles require less fuel than standard vehicles do.

To encourage the adoption of new automobile technologies, certain Canadian provinces are considering providing tax incentives for buyers of low-emission vehicles. The impact of such a policy can be measured by reducing the purchase price of the green alternatives (scenario 4). A reduction by 10% of the capital cost of clean vehicles implies a reduction by 11.42% in the market share of SGVs. The resulting market share gains are bigger for HEVs, but small overall. According to ?, the attribute levels for the low-emission vehicles were set in the survey to values that seem particularly attractive, especially when compared with the actual market conditions. Thus I construct scenario 5, where we consider less attractive purchase prices for the most expensive technologies, namely HEVs and HFCs. The market shares of SGVs and AFVs rise dramatically; the overall penetration of green vehicles is however still high.

The previous scenarios can all be studied using standard discrete choice models (although the results will vary because of different ASCs and potentially different parameters). The innovation of my model results from incorporating environmental concerns through the latent construct EC. As discussed above, even though we do not know the measurement scale of the EC variable, once the model is estimated we can describe its distribution. EC reflects environmental preferences, and the higher its level the more likely consumers are to choose a low-emission vehicle. Scenario 6 seeks to represent a situation where through a social marketing campaign, environmentally unaware consumers are exposed to information on the benefits of reducing carbon emissions and the problems associated with the indiscriminate use of private cars (especially when using fossil fuels). Technically, this scenario is constructed by censoring the density function of the EC variable:

all consumers are constrained to have an EC level at least equal to the mean of the EC variable. In practical terms, the information campaign has successfully changed the concerns of the formerly environmentally unaware consumers. The impact of this simulated campaign is huge, reducing by 69.97% the number of consumers who decide to buy an SGV. In line with the magnitude of the estimated parameters for EC, the augmentation of the market shares is bigger for AFVs and HEVs.

The last scenario is built by considering the baseline but for female consumers only, making it easier to interpret the effect of the EC variable. (Note however that this is not a ceteris paribus analysis.) Our results show that women are more environmentally aware (they constitute a cluster of consumers with a higher level of EC), and so the expected result will be that women favor more low-emission vehicles. Even though the results here are not striking, the predicted market shares do show an increase in favor of green technologies.

4.6 An interesting extension of the model

In addition to the environmental concern variable, in the data I identified another latent dimension related to car purchase decisions and how important are the characteristics of the new alternative:

Appreciation of new car features (ACF): Evaluation of 7 different factors that influence the decision to purchase a new vehicle, according to degree of importance: 5 levels from Not Important to Very Important (see Figure **??**).

1. Purchase price.
2. Fuel economy.
3. Horsepower.
4. Safety.
5. Seating capacity.
6. Reliability.
7. Appearance and styling.

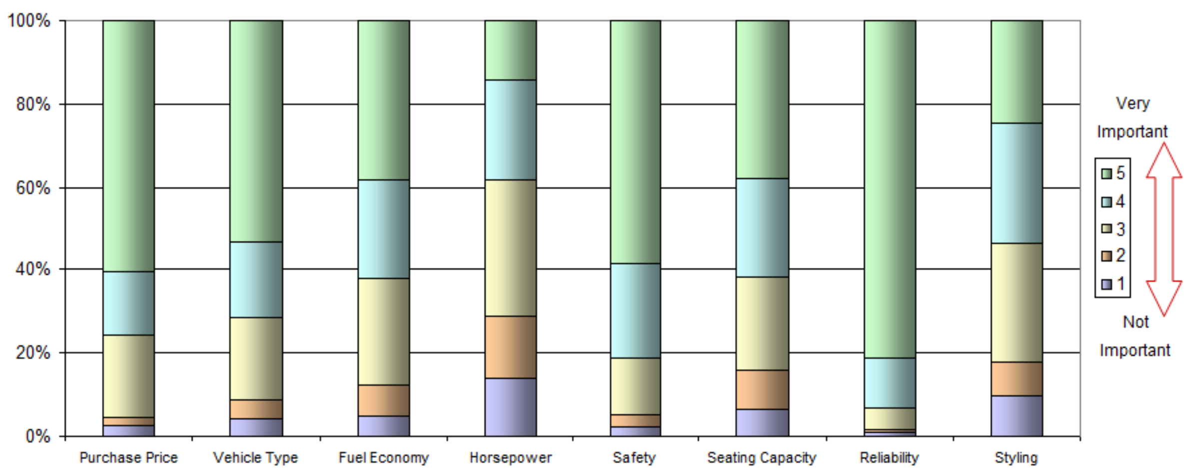

Figure 4.8: Appreciation of new car features

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

$$\begin{aligned} U_{SGVn} &= V_{SGVn} + \Gamma_{1,2}\mathrm{ACF}_n + \upsilon_{SGVn} \\ U_{AFVn} &= V_{AFVn} + \Gamma_{2,1}\mathrm{EC}_n + \Gamma_{2,2}\mathrm{ACF}_n + \upsilon_{AFVn} \\ U_{HEVn} &= V_{HEVn} + \Gamma_{3,1}\mathrm{EC}_n + \Gamma_{3,2}\mathrm{ACF}_n + \upsilon_{HEVn} \\ U_{HFCn} &= V_{HFCn} + \Gamma_{4,1}\mathrm{EC}_n + \Gamma_{4,2}\mathrm{ACF}_n + \Gamma_{4,3}\mathrm{REV}_n + \upsilon_{HFCn}, \end{aligned} \tag{4.41}$$

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

The path diagram of the extended model is sketched in Figure **??**. The results of this extension will be quickly described.

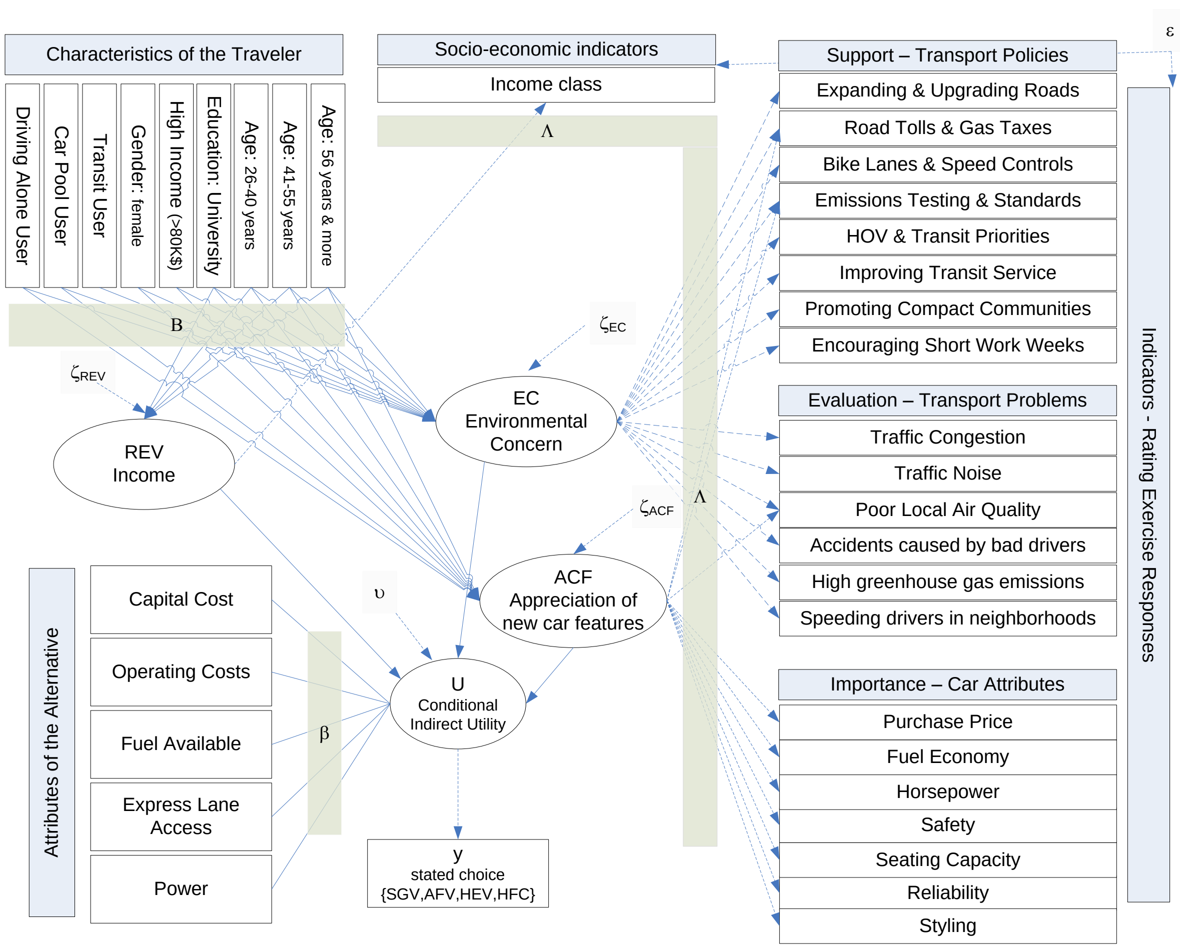

Figure 4.9: Extended private vehicle purchase HCM

In Table **??** the results of the extended choice model are presented. Note that both

EC and ACF have a significant positive impact on the probability of choosing one of the green technologies. However, ACF also had a positive impact on the probability of selecting a standard vehicle.

	Hybrid Choice Model	
	estimates	**t-stat**
ASC_AFV	-6.626	-5.162
ASC_HEV	-4.383	-5.688
ASC_HFC	-6.403	-8.750
Capital Cost	-0.943	-4.369
Operating Cost	-0.849	-3.917
Fuel available	1.384	7.096
Express lane access	0.162	2.229
Power	2.710	3.985
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	
Choice Model adj. rho-square	0.236	
Number of draws (burn-in)	5000 (500)	

Table 4.12: Extended Car Choice Model Results

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 **??**.

As in the original model, environmental concerns appear to be more developed in women and more educated people, as well as for transit users. Vehicle features are more appreciated by people who drive alone than by those who carpool. They are also more valued by women and by older and more educated people.

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*. The effect of age on 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).

	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 4.13: Extended Structural Model Results

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

$$TPS_2 = \alpha_{TPS_2} + \lambda_{EC,TPS_2}\mathrm{EC}_n + \lambda_{AF,TPS_2}\mathrm{ACF}_n + \varepsilon_{TPS_2}. \tag{4.42}$$

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 **??** and the results are displayed in Table **??**.

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. 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.

	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 4.14: Extended Structural Model Covariates

In sum, some interesting conclusions can be made from this extended model, particularly with regard to sustainable transportation. As in the original model, environmental concerns would discourage expansion of the road network, but would support expansion of the bicycle path network, implementation of tolls and increase of fuel taxes. Conversely, the appreciation of vehicles would have a negative impact on supporting sustainable transportation policies, such as the imposition of fees and emission standards.

4.7 Variations on a theme: taking account for safety

Beyond the role of environmental concerns that I have addressed throughout this chapter, in this section, I use the same data and tools to specifically analyze the role of safety-related attributes on pro-environmental choice.

When looking at the list of attributes frequently used in previous research in vehicle purchase modeling, a question arises on the role of safety in the vehicle purchase decision. Is safety an attribute that consumers do not consider when they are evaluating the different alternatives before buying a new car?

? is the first application of vehicle safety – through collision rates and the probability of a severe accident – as a relevant attribute in the choice of a new vehicle to purchase. Using a multinomial logit model (MNL), **?** defines vehicle safety as a subset of the attributes of the vehicle. Based on **?** and **?**, **?** considers the following related-to-safety attributes: head stroke (distance from the windshield to the seat back), passive restraint (airbag or passive seat belt), crashworthy index (according to model-year), and vehicle weight. In general, the results show that consumers do prefer safer vehicles, with significant parameters for the safety-related attributes. However, how well those safety attributes reflect individual preferences for safety is debatable.

The problem is that safety is a qualitative variable that is cognitively built using different dimensions. When we define measurable attributes for safety, one modeling possibility is to work with a quantitative approach of proxy variables, just as **?** did in his article. However, the use of proxy variables for safety should involve dealing not only with vehicle attributes (such as the presence of air bags, ABS, or other technical specifications related to vehicle safety) but also with attributes that reflect the environment where the vehicle is used: road conditions (road safety), number of accidents, presence of speeding drivers, and winter driving conditions and requirements. If the external factors define a less safe driving environment, then the likelihood of a consumer choosing a safer vehicle should be higher.

The problem with this approach is that, first, it is very complex to have an exhaustive list of measurable safety proxy attributes. Additionally, the safety proxy attributes together explain the effect of one variable only. This fact suggests that it is likely that these proxy variables will be correlated, with the corresponding modeling problems associated with collinearity.

Recognizing its qualitative nature, a better modeling approach is to consider safety as a perceptual/attitudinal latent construct. Different studies have been centered on the attitudinal dimension of safety, measuring the relative importance of vehicle safety in the vehicle purchase decision (**????????**). What these studies do is, in a real/hypothetical situation of buying a new car, ask consumers to rate the importance of different vehicle attributes (see **?**, for a quick but comprehensive discussion of most of those studies). Usually, the conclusion is that safety is an important attribute but is repeatedly outranked by other attributes, namely purchase price, appearance and reliability. If consumers are asked to mention the most important factor for a particular cluster of vehicles defined by price range and vehicle type, then safety appears as the main answer (**?**).

However, in the interesting work of **?**, the authors found that safety is the the most important factor in the purchase process for new vehicle consumers, whereas purchase price was ranked third. The authors also found that preference for safety shows cultural patterns, vehicle safety being more appreciated by Swedish participants than Spanish participants. Finally, they also showed that consumers conceptualize vehicle safety using safety-related features (such as air bags or ABS) rather than crash/safety test results or a crashworthy index. This particular finding, also mentioned in other works such as **?** and **?**, questions the use of crashworthiness as an attribute to model vehicle safety – which was used, for example, in the work of **?**. Once again, the problem with the attitudinal approach alone is that it does not explain choice behavior.

4.7.1 Safety in Canadian consumer's response to green vehicles

Whereas safety was not considered directly as an experimental attribute in the vehicle choice data I analyze throughout this chapter, there are safety-related factors appearing in two different questions of the survey. First, using 5 different levels, from not at all important (level 1) to very important (level 5), the participants were asked to rate the importance that a group of different factors had in the decision to purchase their current vehicle.[7] The factors to be evaluated were: purchase price, vehicle type, fuel economy, horsepower, vehicle safety, seating capacity, reliability and styling (see Figure **??**). While reliability tops the list, with 81% of the participants rating reliability as a very important factor, both purchase price and vehicle safety are almost as likely to be rated very important, reaching a very important evaluation of 61% and 59%, respectively. In fact, considering together the two highest levels of importance (grouping levels 4 and 5),

[7] I will take the responses to this rating exercise as attitudinal indicator variables for a latent variable related to *appreciation of car features* (ACF) of the consumer, just as I did in section **??**.

then vehicle safety comes second with 81%, after reliability (93%), and before purchase price (76%), which ends up closer to vehicle type (72%).

Additionally, participants were asked to evaluate both different policies or government actions that would influence the transportation system according to degree of support, as well as different problems related to transportation according to degree of seriousness. I used these indicators to identify the environmental concern variable in section **??**. However, note that among the policies and problems being evaluated, 4 of them are related to experienced safety while driving: bike lanes and speed controls, traffic congestion, accidents caused by bad drivers, and presence of speeding drivers in neighborhoods. The answers to these questions are related to *safety environment* and, together with the vehicle safety variable, they serve as perceptual indicator variables for a general latent construct related to *safety* (SAF).

4.7.2 An HCM of vehicle choice including safety

As **?** points out, since vehicle safety was not included as an attribute in the SP exercise of virtual vehicle choice, it is not possible to use a standard discrete choice model to consider the answers related to safety we just described. However, using an appropriate Hybrid Choice Model we can quantify the impact of the safety-related cognitive factors on the private vehicle purchase decision. Our hypothesis is that the decision of a new private vehicle purchase is affected not only by the experimental attributes of the different vehicles, but also by both the attitudinal *appreciation of car features* (ACF) of the consumer, including vehicle safety, and the latent construct related to *safety* (SAF). This version of the model is presented in Figure **??**.

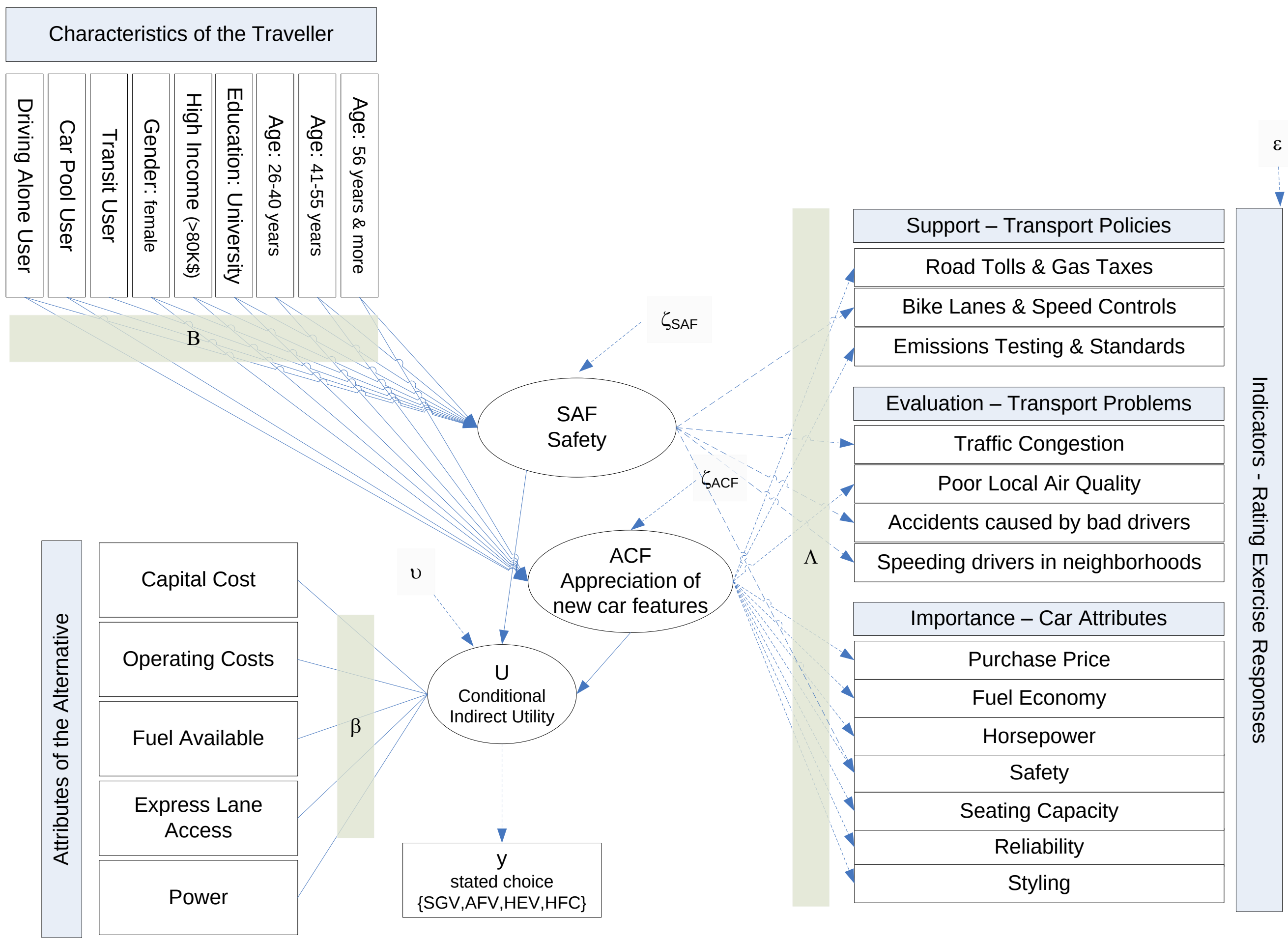

Figure 4.10: Private vehicle purchase HCM centered on safety

Expanding on the results of section **??**, the Bayesian Gibbs sampling routine for this case with two latent variables is:

$$\pi(\theta|\mathrm{ACF}, \mathrm{SAF}, b_{\mathrm{ACF}}, b_{\mathrm{SAF}}, \Lambda, y, I) \quad \propto \quad |H|^{\frac{1}{2}} \exp\left(\frac{1}{2}(\theta-\hat{\theta})' H(\theta-\hat{\theta})\right) \tag{4.43}$$

$$\pi(\mathrm{ACF}_n|\theta, b_{\mathrm{ACF}}, \Lambda, y_n, I_n) \quad \sim \quad N(\mu_{\mathrm{ACF}_n|I_n}, \sigma^2_{\mathrm{ACF}|I}), \forall n \tag{4.44}$$

$$\pi(\mathrm{SAF}_n|\theta, b_{\mathrm{SAF}}, \Lambda, y_n, I_n) \quad \sim \quad N(\mu_{\mathrm{SAF}_n|I_n}, \sigma^2_{\mathrm{SAF}|I}), \forall n \tag{4.45}$$

$$\pi(b_{\mathrm{ACF}}|\mathrm{ACF}, \theta, b_{\mathrm{ACF}}, \Lambda, y, I) \quad \sim \quad N(\bar{b_{\mathrm{ACF}}}, \bar{V}_{b_{\mathrm{ACF}}}) \tag{4.46}$$

$$\pi(b_{\mathrm{SAF}}|\mathrm{SAF}, \theta, b_{\mathrm{SAF}}, \Lambda, y, I) \quad \sim \quad N(\bar{b_{\mathrm{SAF}}}, \bar{V}_{b_{\mathrm{SAF}}}) \tag{4.47}$$

$$\pi(\Lambda|\mathrm{ACF}, \mathrm{SAF}, \theta, b_{\mathrm{ACF}}, b_{\mathrm{SAF}}, y, I) \quad \sim \quad N(\bar{\Lambda}, \bar{V}_{\Lambda}). \tag{4.48}$$

First, I present the results of the discrete choice model. As explained before, the private vehicle choice corresponds to the discrete choice kernel, in this case a multinomial logit, where the parameters to estimate are described by the taste parameter vector θ of the utility function. 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 HCM utility specification also contains the effect of the latent variables. Thus, the set of equations for the mode choice model alone are given by:

$$U_{SGVn} = V_{SGVn} + \upsilon_{SGVn} \tag{4.49}$$

$$U_{AFVn} = V_{AFVn} + \Gamma_{AFV,\mathrm{SAF}}\mathrm{SAF}_n + \Gamma_{AFV,\mathrm{ACF}}\mathrm{ACF}_n + \upsilon_{AFVn} \tag{4.50}$$

$$U_{HEVn} = V_{HEVn} + \Gamma_{HEV,\mathrm{SAF}}\mathrm{SAF}_n + \Gamma_{HEV,\mathrm{ACF}}\mathrm{ACF}_n + \upsilon_{HEVn} \tag{4.51}$$

$$U_{HFCn} = V_{HFCn} + \Gamma_{HFC,\mathrm{SAF}}\mathrm{SAF}_n + \Gamma_{HFC,\mathrm{ACF}}\mathrm{ACF}_n + \upsilon_{HFCn}, \tag{4.52}$$

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

Car Choice Model	**HCM**	
	estimates	**t-stat**
ASC_AFV	-0.824	-0.936
ASC_HEV	1.509	2.276
ASC_HFC	0.397	0.576
Capital Cost	-0.931	-4.366
Operating Cost	-0.811	-3.931
Fuel available	1.334	7.000
Express lane access	0.145	2.026
Power	1.509	2.026
Latent Variables		
SAF on AFV	0.715	2.767
ACF on AFV	0.152	0.525
SAF on HEV	0.488	3.640
ACF on HEV	-0.197	-1.293
SAF on HFC	0.369	2.773
ACF on HFC	0.230	1.445
Number of individuals	1877	
Choice Model adj. rho-square	0.237	
Number of draws (burn-in)	5000 (500)	

Table 4.15: Car Choice Model Results, preferences for safety

As shown in Table **??**, the significance of the latent variable parameters shows a relevant effect on the individual utilities. Through its positive sign, the safety latent variable SAF has a pro-environmental impact encouraging the choice of green technologies. In fact, SAF has the highest effect on the *alternative fuel vehicle* AFV, followed by the *hybrid vehicle* HEV and then the *hydrogen fuel cell vehicle* HFC. Interestingly, the order appears to be reversed when compared with the results when an environmental concern variable is included (cf. Table **??**).

Note that we obtain an unexpected negative sign for the effect of the ACF latent variable on the probability of choosing the HEV alternative. This unexpected sign can be explained by the switch in sign of the HEV specific constant, which was negative in the standard MNL case and then changed sign in the HCM case. Thus, to adjust to the observed market shares, a negative sign for the effect of ACF on HEV is obtained.

The structural equation links consumer characteristics with the latent variables through a linear regression equation based on the usual mode of transportation to get to work or school (driving, carpooling or trasit), the individual's gender, age, education level, and household income. The estimation results are shown in Table **??**:

Structural Model	**SAF**		**ACF**	
	est	**t-stat**	**est**	**t-stat**
Intercept	-3.900	-10.67	-0.597	-7.74
Driving Alone User	0.069	1.10	-0.153	-3.14
Car Pool User	0.205	1.99	0.089	0.94
Transit User	0.136	1.40	-	-
Female Dummy	0.402	8.52	0.280	6.95
High Income Dummy (>80K$)	0.029	0.58	0.001	0.02
Education: University	-0.112	-2.26	0.058	1.34
Age level: 26-40 years	0.053	0.59	0.334	4.41
Age level: 41-55 years	0.147	1.60	0.624	7.45
Age level: 56 years & more	0.410	3.98	0.490	5.54

Table 4.16: Structural Model Results, preferences for safety

Interestingly, from this model we can conclude that safety (SAF) is more important for car pool users than for driving alone users. We also find that concern about safety issues in the vehicle purchase choice context is more developed in women and older people. These results show a risk-aversion behavior and are in line with previous findings. For example **?** find that older participants were more likely to list safety as their most important consideration than younger or middle-aged participants. In addition, the effect of the high income variable on safety is positive but not significant.

Whereas I obtain an unexpected negative sign for the variable indicating university studies on safety, the effect of this variable on the appreciation of car features (ACF) is positive (although with a low significance). In addition, the effect of age on ACF is the highest for people between 41 to 55 years.

Lastly, the measurement model identifies the effect of the latent variable on each indicator variable. Several indicators were considered in the latent variable measurement model, which links the latent variables SAF and ACF to answers to attitudinal/perceptual qualitative survey questions. As explained before, the questions selected to define the indicator variables concern the respondent's level of support for or opposition to transport policies (Transport Policies Support), their opinions on various transport-related issues (Transport Problems Evaluation), and their evaluation of car attribute importance. The results are shown in Table **??**.

Measurement Model	**estimates**	**t-stat**
Transport Policies Support		
ACF on Road Tolls & Gas Taxes	0.075	1.161
SAF on Bike Lanes & Speed Controls	0.174	4.460
ACF on Emission Testing & Standards	0.442	10.804
Transport Problems Evaluation		
SAF on Traffic Congestion	0.461	8.648
ACF on Poor Local Air Quality	0.200	3.801
SAF on Accidents Caused by Bad Drivers	0.832	15.427
SAF on Speeding Drivers in Neighbourhoods	1.000	-
Car Attribute Importance		
ACF on Purchase Price	0.107	2.276
ACF on Fuel Economy Importance	0.385	7.163
ACF on Horsepower Importance	0.530	9.065
SAF on Vehicle Safety Importance	0.213	6.909
ACF on Vehicle Safety Importance	1.000	-
ACF on Seating Capacity Importance	0.784	14.558
ACF on Reliability Importance	0.532	15.483
ACF on Styling	0.347	6.151

Table 4.17: Measurement Model, preferences for safety

Some interesting conclusions can be drawn from the estimated parameters, especially about perceptions related to safety. For example, we confirm that the latent variable SAF is related to safety driving conditions. In this context, the factor with the highest weight is *speeding drivers in neighborhoods* confirming the idea that a major worry for individuals are drivers who do not respect speed limits, risking fatal crashes. Next comes *accidents caused by bad drivers*, which could be interpreted as an outcome of the presence of speeding drivers, among other attributes defining *bad drivers*. Thus, to favor consumers' perception of a safer driving environment some policies in this regard could be implemented, such as effective speeding reduction interventions. Although reducing speed limits can be cited as one of these interventions, the estimated parameter for *speed controls* is low when compared to the other safety indicators. However, it is essential to note that the actual indicator in our database is *bike lanes & speed controls*. This definition does not allow us to identify a specific weight only for *speed controls*. As for vehicle safety, the lower weight is explained by a lower correlation with the concept of *safety environment*.

The results of the indicator variables for the relative importance of different car attributes show the highest weight for the *vehicle safety indicator*. This result must be understood in a context where some attributes that are being evaluated also appear in the utility function of the discrete choice model, namely purchase price (as capital cost), fuel economy (as operating cost), power, and type of car (through the set of available alternatives). Our results should be interpreted then as evidence from a more general framework for the result we find in the literature stating that for a particular cluster of vehicles, safety appears as the major concern when choosing a new car to purchase.

Understanding consumers' preferences for safety is essential for encouraging safe driving behavior and for developing a safer driving environment: if we can model consumers' demand for safety, we can also test the impact of both different public policies related to safety as well as private improvements to vehicle safety. However, rating the importance of safety variables for vehicle consumers has been the dominant modeling tool in those few studies which have so far been done on the role of safety in the new vehicle purchase process. Although this approach allows us to determine the current relative importance of different safety-related features even when compared to other vehicle features (such as reliability or comfort), the use of rating techniques lacks predictive power because there is no representation of the behavioral choice process.

In this application, using empirical data, I developed, implemented and applied a hybrid choice model (HCM) to explain consumers' preferences for safety in the private vehicle choice context. Although the data were flexible enough to construct the first application of an HCM with a safety latent variable, it is important to mention that the database was not conceived to evaluate the role of safety in the vehicle purchase decision. However, my results introduce the elements for future applications focused on safety. Stated or revealed preference data about vehicle purchase decisions using standard measurable vehicle attributes (such as make and model, purchase price, performance, reliability, and durability) can be combined with attitudinal rating exercises that will be used to construct latent variables related to safety, just as I did in this application. To have a more complete description of factors, I suggest including, first, the assessment of the importance of different vehicle safety technologies (such as airbags, ABS, seat belts, seat belt reminders, speed alerts, parking assistance). This evaluation will permit us to construct a latent variable related specifically to vehicle safety (in the current database, vehicle safety appears as a general concept without a clear definition). Then, the attitudinal evaluation of a general and more exhaustive list of vehicle features (such as purchase price, performance, style, comfort, size) can be completed with the resulting vehicle safety latent variable. From a modeling perspective, this can be done by permitting simultaneity among the latent variables, an extension that can be easily incorporated using the HCM framework. To quantify the safety environment perceptions, I also suggest the evaluation of different safety policies related to road safety, such as traffic calming programs (for example speed cushions and speed bumps), and other traffic safety measures (speed controls, pedestrian and bicycle plans, educational programs for young drivers, improved safer road networks, and traffic engineering measures such as protective road fencing for vehicles). The evaluation of those policies, together with the perception of unsafe driving conditions (traffic congestion, speeding drivers, high accident rates, poor road conditions), will provide indicator variables to build a latent variable related to a safe driving environment.

This more general framework will permit us to understand even better the role of safety as a cognitive factor of vehicle consumers, with the concomitant gain in predictive power useful not only for manufacturers – who will have information about clusters of consumers according to their safety perceptions –, but also policy makers: if we understand better consumers' perceptions of safety it will be easier to plan more effective safety policies and safety campaigns. An example of this would be to reinforce the importance of an eventually low appreciated safety vehicle feature that has proven to avoid causalities in crash tests.

4.8 Conclusions

In this chapter, using survey data, I have developed, implemented and applied a hybrid choice model (HCM) to explain environmental preferences in a private vehicle choice context. My specification is consistent with the new trend in discrete choice modeling toward incorporating perceptual/attitudinal factors into the behavioral representation of the decision process. The HCM formulation offers an attractive improvement in modeling private vehicle choice behavior, because the choice model is only a part of the whole behavioral process in which we now incorporate individual attitudes, opinions and perceptions, thus yielding a more realistic econometric model. This improved representation outperforms standard discrete choice models because now it is possible to build a profile of Canadian consumers and their ability to adapt to technological innovation with regard to sustainable private vehicle alternatives. Indeed, a latent environmental concern EC variable enters very significantly and positively in the discrete choice kernel of my model, favoring the adoption of green automobile technologies through a positive impact in the choice probabilities.

In addition, I can summarize some of the practical results of my research: environmental concerns (ECs) are more important for public transportation users than for car pool users; car drivers may be unaware of the environmentally adverse effects of private car use; concern about environmental issues is more developed in women, older people and more educated people; environmentally-conscious consumers negatively perceived the priority given to cars by policies aimed at raising road capacity, and hence there is a perceived positive environmental impact of measures allowing a rational use of private vehicles as well as a measurable positive effect of encouraging sustainable transport. Whereas the discrete choice kernel and the structural model are used for policy simulations, the measurement model serves to infer which policies can be more effective in encouraging the adoption of green technologies. For instance, my results support fossil fuel taxation: road tolls and a tax on vehicle carbon emissions are the transport policies that show the highest correlation level with the EC variable. Then, if fuel taxes are applied, my policy simulation consistently predicts deeper market penetration for low-emission vehicles.

Beyond the interesting results showing that environmental concern encourages the choice of green automobile technologies, this application also proves the practical feasibility of the Gibbs sampler I have developed for HCM estimation, exploiting data augmentation techniques for the latent variables. To my knowledge, this is the first empirical application of the HCM Gibbs sampler. Whereas Gibbs sampling for a probit kernel is analytically straightforward because it also admits the use of data augmentation, in the case of both a multinomial logit (MNL) kernel and a mixed logit (MMNL) kernel one fails to find a closed form full conditional distribution for the taste parameters of the utility function. However, I showed that it is possible to exploit Metropolis-Hastings (MH) methods for both the MNL and MMNL cases. In fact, my numerical application concerns an MNL kernel. Even though the probit kernel formulation breaks down the methodological complexity of the model, the data augmentation step for the utility function is very demanding in computational terms, and eventually could be outperformed by a logit-based kernel – even with the additional MH step required by logit models. In addition, classical estimation of HCMs is very demanding in situations with a large number of latent variables – each additional latent variable sums another dimension in the joint choice probability. Thus, Bayesian HCM estimation clearly outperforms simulated maximum likelihood: the inclusion of additional latent variables under the Bayesian approach implies simply working with ordinary regressions (i.e. sampling additional draws from a normal distribution). Another advantage of the Bayesian approach is that it allows us to forecast using the same sample generated for estimation. In fact, the Bayesian estimates describe the posterior distribution, permitting a direct calculation of confidence intervals for WTPs as well as standard deviations for both the choice probabilities and market shares.

References

M. Achtnicht, G. Bühler, and C. C. Hermeling. Impact of service station networks on purchase decisions of alternative-fuel vehicles. Discussion paper no. 08-088, ZEW Zentrum für Europäische Wirtschaftsforschung GmbH - Centre for European Economic Research, 2008.

I. Ajzen. Nature and operation of attitudes. *Annual Reviews of Psychology*, 52:27–58, 2001.

J.H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

P.A. Armstrong, R.A. Garrido, and J. de D. Ortúzar. Confidence intervals to bound the value of time. *Transportation Research Part E*, 37(2):143–161, 2001.

K. Ashok, W. Dillon, and S. Yuan. Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, 39(1):31–46, 2002.

Charles River Associates. Consumer acceptance of automotive crash avoidance devices. A report of qualitative research, Charles River Associates Incorporated, Department of Transportation, Boston, Mass, 1998.

S.D. Beggs and N.S. Cardell. Choice of smallest car by multi-car households and the demand for electric vehicles. *Transportation Research*, 14A:389–404, 1980.

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

M. Ben-Akiva, 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): 163–175, 2002.

A. Bento, L. Goulder, M. Jacobsen, and R. Von Haefen. Distributional and efficiency impacts of increased U.S. gasoline taxes. *American Economic Review*, 99(3), 2009.

D. Bolduc and R. Alvarez-Daziano. On estimation of hybrid choice models. In S. Hess and A. Daly, editors, *Choice Modelling: the state-of-the-art and the state-of-practice, Proceedings from the Inaugural International Choice Modelling Conference*. Emerald, England, 2010.

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

D. Bolduc, N. Boucher, and R. Alvarez-Daziano. Hybrid choice modeling of new technologies for car choice in Canada. *Journal of the Transportation Research Board*, 2082:63–71, 2008. Transportation Research Board of the National Academies, Washington, D.C.

D. Brownstone and K.E. Train. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89:109–129, 1999.

D. Brownstone, D.S. Bunch, T.F. Golob, and W. Ren. A transactions choice model for forecasting demand for alternative-fuel vehicles. *Research in Transportation Economics*, 4:87–129, 1996.

D. Brownstone, D.S. Bunch, and K. Train. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research*, 34B:315–338, 2000.

D.S. Bunch, M. Bradley, T.F. Golob, R. Kitamura, and G.P. Occhiuzzo. Demand for alternative-fuel vehicles in California: a discrete-choice stated preference pilot project. *Transportation Research*, 27A:237–253, 1993.

J.K. Dagsvik and G. Liu. A framework for analyzing rank-ordered data with application to automobile demand. *Transportation Research Part A*, 43(1):1–12, 2009.

J.K. Dagsvik, T. Wennemo, D.G. Wetterwald, and R. Aaberge. Potential demand for alternative fuel vehicles. *Transportation Research B*, 36(4):361–384, 2002.

G. Dinse. Akzeptanz von wasserstoffbetriebenen Fahrzeugen. Eine Studie über die Verwendung eines neuen und ungewohnten Kraftstoffs. Working paper, Institut für Mobilitätsforschung (ifmo-studien), Berlin, 2000.

G. Ewing and E. Sarigöllü. Assessing consumer preferences for clean-fuel vehicels: a discrete-choice experiment. *J. Public Policy Mark*, 19:106–118, 2000.

S.A. Ferguson and A.H. Williams. What safety means to consumers and its role in the purchase decisions. *J. Traffic Med.*, 24(3):83–89, 1996.

B. Flamm. The impacts of environmental knowledge and attitudes on vehicle ownership and use. *Transportation Research Part D*, 14:272–279, 2009.

A. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

G. Godoy and J.D. Ortúzar. On the estimation of mixed logit models. In P.O. Inweldi, editor, *Transportation Research Trends*, pages 215–235. Nova Science Publishers, New York, 2009.

J. Gould and T.F. Golob. Clean air forever? A longitudinal analysis of opinions about air pollution and electric vehicles. *Transportation Research Part D: Transport and Environment*, 3(3):157–169, 1998.

H.F. Gunn and J.J. Bates. Statistical aspects of travel demand modelling. *Transportation Research*, 16A:371–382, 1982.

R.R. Heffner, K.S. Kurani, and T.S. Turrentine. Symbolism in California's early market for hybrid electric vehicles. *Transportation Research D: Transport and Environment*, 12(6):396–413, 2007.

M. Horne. *Incorporating preferences for personal urban transportation technologies into a Hybrid Energy-Economy Model.* Master's degree thesis, Simon Fraser University, School of Resource and Environmental Management, 2003.

M. Horne, M. Jaccard, and K. Tiedman. Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. *Energy Economics*, 27:59–77, 2005.

M. Vredin Johansson, T. Heldt, and P. Johansson. The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40:507–525, 2006.

O. Johansson-Stenman and P. Martinsson. Honestly, why are you driving a BMW? *J. Econ. Behav. Organ.*, 60:129–146, 2006.

S. Koppel, J. Charlton, B. Fildes, and M. Fitzharris. How important is vehicle safety in the new vehicle purchase process? *Accident Analysis and Prevention*, 40:994–1004, 2008.

K. Lancaster. A new approach to consumer theory. *Journal of Political Economy*, 74:132–157, 1966.

C. Lave and K. Train. A disaggregate model of auto-type choice. *Transportation Research A*, 3(1):1–9, 1979.

U. Lossen, M. Armbruster, S. Horn, P. Kraus, and K. Schich. *Einflussfaktoren auf den Markterfolg von wasserstoffbetriebenen Fahrzeugen*. Expert Verlag, Renningen, 2003.

C. Manski. The structure of random utility models. *Theory and Decision*, 8:229–254, 1977.

C.F. Manski and L. Sherman. An empirical analysis of household choice among motor vehicles. *Transportation Research A*, 14, 1980.

P. McCarthy. Market price and income elasticities of new vehicle demands. *The Review of Economics and Statistics*, 78(3):543–547, 1996.

P.S. McCarthy. Seat belt usage rates: a test of Peltzman's hypothesis. *Accident Analysis and Prevention*, pages 425–438, 1986.

P.S. McCarthy. Consumer demand for vehicle safety: an empirical study. *Economic Inquiry*, XXVIII:530–543, 1990.

R.R. McCulloch, N.G. Polson, and P.E. Rossi. Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99:173–193, 2000.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontier in Econometrics*. Academic Press, New York, 1974.

MORI. Euroncapconsumer car buying survey. Technical report, Market and Opinion Research International (MORI), 2005.

General Motors. Continuous automotive market information program buyer behavior. Technical report, General Motors Corporation, Detroit, 1994.

M.A. Munizaga and R. Alvarez-Daziano. Testing mixed logit and probit models by simulation. *Transportation Research Record*, 1921:53–62, 2005.

T. O'Garra, P. Pearson, and S. Mourato. Public acceptability of hydrogen fuelcell transport and associated refuelling infrastructure. In R. Flynn and P. Bellaby, editors, *Risk and the public acceptance of new technologies*. Palgrave MacMillan, Hampshire, 2007.

D. Potoglou and P.S. Kanaroglou. Household demand and willingness to pay for clean vehicles. *Transportation Research D: Transport and Environment*, 12(4):264–274, 2007.

J.D. Power. Automotive consumer profile study. Technical report, J.D. Power and Associates, Agoura Hills, CA, 1993.

DesRosiers Automotive Reports. *The Need for Dialogue: A Perspective on Vehicle Safety*, volume 16. 2002. ISSN 0841-9957.

M. Yetano Roche, S. Mourato, M. Fischedick, K. Pietzner, and P. Viebahn. Public attitudes towards and demand for hydrogen and fuel cell vehicles: A review of the evidence and methodological implications. *Energy Policy*, 2009. (in press).

P.E. Rossi, G. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. John Wiley and Sons, Chichester, 2006.

S. Scott. Data augmentation for the bayesian analysis of multinomial logit models. Proceedings of the american statistical association section on bayesian statistical science, 2003.

M.L. Tam, W.H.K. Lam, and H.P. Lo. Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica*, 6(1):3–17, 2010.

K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, second edition, 2009.

K. Train, D. McFadden, and A. Goett. The incorporation of attitudes in econometric models of consumer choice. Working paper, Cambridge Systematics, 1986.

J. Walker. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. PhD dissertation, MIT, Massachusetts, 2001.

J. Walker. The mixed logit (or logit kernel) model: Dispelling misconceptions of identification. *Transportation Research Record*, 1805:86–98, 2002.

C. Winston and F. Mannering. Consumer demand and automobile safety. Working Paper, Massachusetts Institute of Technology, 1983.

Appendix A: Variable description

In the following tables we give the details of the different components of the vectors of equations **??**, **??**, **??** and **??**.

Variable	Description
U_{SGV}	Utility associated with a Standard Gas Vehicle (SGV)
U_{AFV}	Utility associated with an Alternative Fuel Vehicle (AFV)
U_{HEV}	Utility associated with a Hybrid Electric Vehicle (HEV)
U_{HFC}	Utility associated with a Hydrogen Fuel Cell vehicle (HFC)
EC	Environmental Concern latent variable
I_1	Expanding & Upgrading Roads - Support Indicator
I_2	Road Tolls & Gas Taxes - Support Indicator
I_3	Bike Lanes & Speed Controls - Support Indicator
I_4	Reducing Car Emissions - Support Indicator
I_5	High Occupancy Vehicles & Transit Priorities - Support Indicator
I_6	Improving Transit Service - Support Indicator
I_7	Promoting Compact Communities - Support Indicator
I_8	Encouraging Short Work Weeks - Support Indicator
I_9	Traffic Congestion - Evaluation Indicator
I_{10}	Traffic Noise - Evaluation Indicator
I_{11}	Poor Local Air Quality - Evaluation Indicator
I_{12}	Accidents Caused by Bad Drivers - Evaluation Indicator
I_{13}	Emissions & Global Warming - Evaluation Indicator
I_{14}	Speeding Drivers in Neighborhoods - Evaluation Indicator

Table 4.18: Dependent Variables

Parameter	Variable	Description
b_1	w_1	Intercept
b_2	w_2	Driving Alone User
b_3	w_3	Car Pool User
b_4	w_4	Transit User
b_5	w_5	Female Indicator
b_6	w_6	High Income Indicator (>80K$)
b_7	w_7	Education: University
b_8	w_8	Age level: 26-40 years
b_9	w_9	Age level: 41-55 years
b_{10}	w_{10}	Age level: 56 years & more
ASC_{AFV}	$X_{AFV,1}$	Alternative Fuel Vehicle (AFV) constant
ASC_{HEV}	$X_{SGV,2}$	Hybrid Electric Vehicle (HEV) constant
ASC_{HFC}	$X_{SGV,3}$	Hydrogen Fuel Cell Vehicle (HFC) constant
β_1	$X_{\cdot,4}$	Purchase Price
β_2	$X_{\cdot,5}$	Fuel Cost
β_3	$X_{\cdot,6}$	Fuel Availability
β_4	$X_{\cdot,7}$	Express lane access
β_5	$X_{\cdot,8}$	Power
$\Gamma_{AFV,\mathrm{EC}}$	EC	EC effect on AFV
$\Gamma_{HEV,\mathrm{EC}}$	EC	EC effect on HEV
$\Gamma_{HFC,\mathrm{EC}}$	EC	EC effect on HFC
λ_1	EC	EC effect on Expanding & Upgrading Roads
λ_2	EC	EC effect on Road Tolls & Gas Taxes
λ_3	EC	EC effect on Bike Lanes & Speed Controls
λ_4	EC	EC effect on Reducing Car Emissions
λ_5	EC	EC effect on High Occupancy Vehicles & Transit Priorities
λ_6	EC	EC effect on Improving Transit Service
λ_7	EC	EC effect on Promoting Compact Communities
λ_8	EC	EC effect on Encouraging Short Work Weeks
λ_9	EC	EC effect on Traffic Congestion
λ_{10}	EC	EC effect on Traffic Noise
λ_{11}	EC	EC effect on Poor Local Air Quality
λ_{12}	EC	EC effect on Accidents Caused by Bad Drivers
λ_{13}	EC	EC effect on Emissions & Global Warming
λ_{14}	EC	EC effect on Speeding Drivers in Neighborhoods

Table 4.19: Independent Variables and Parameters

Appendix B: Posterior distributions and MCMC sequences

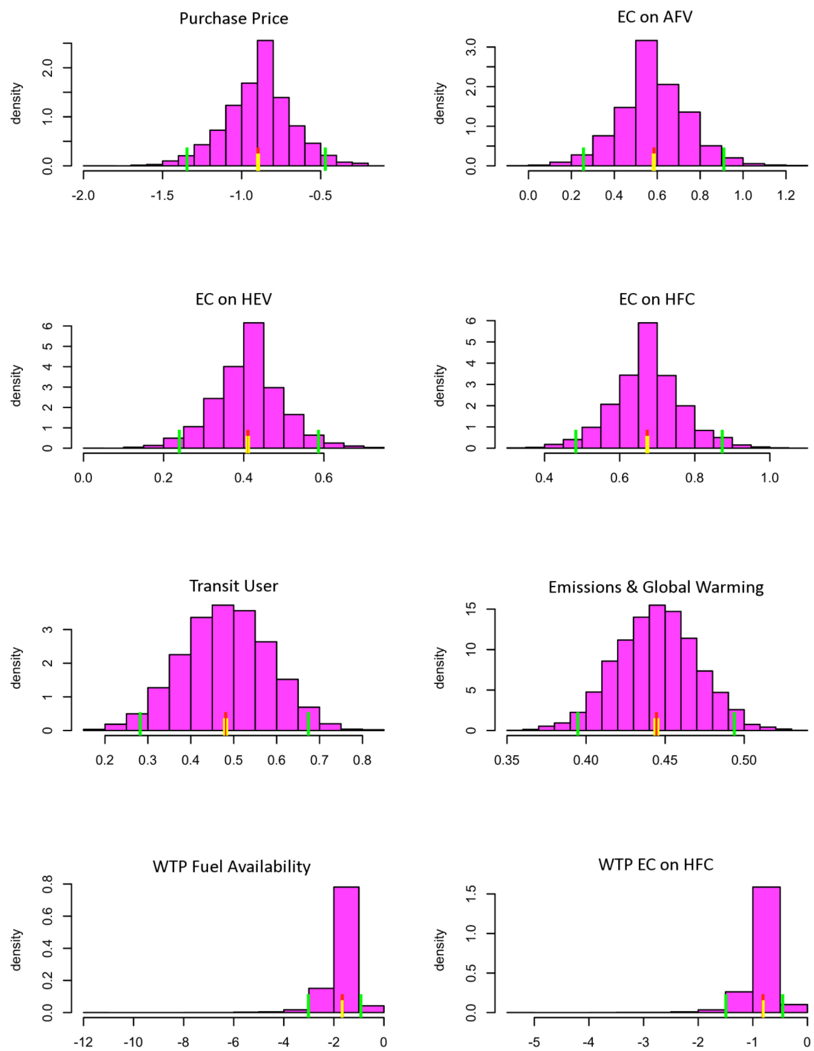

Figure 4.11: Posterior distribution of selected parameters

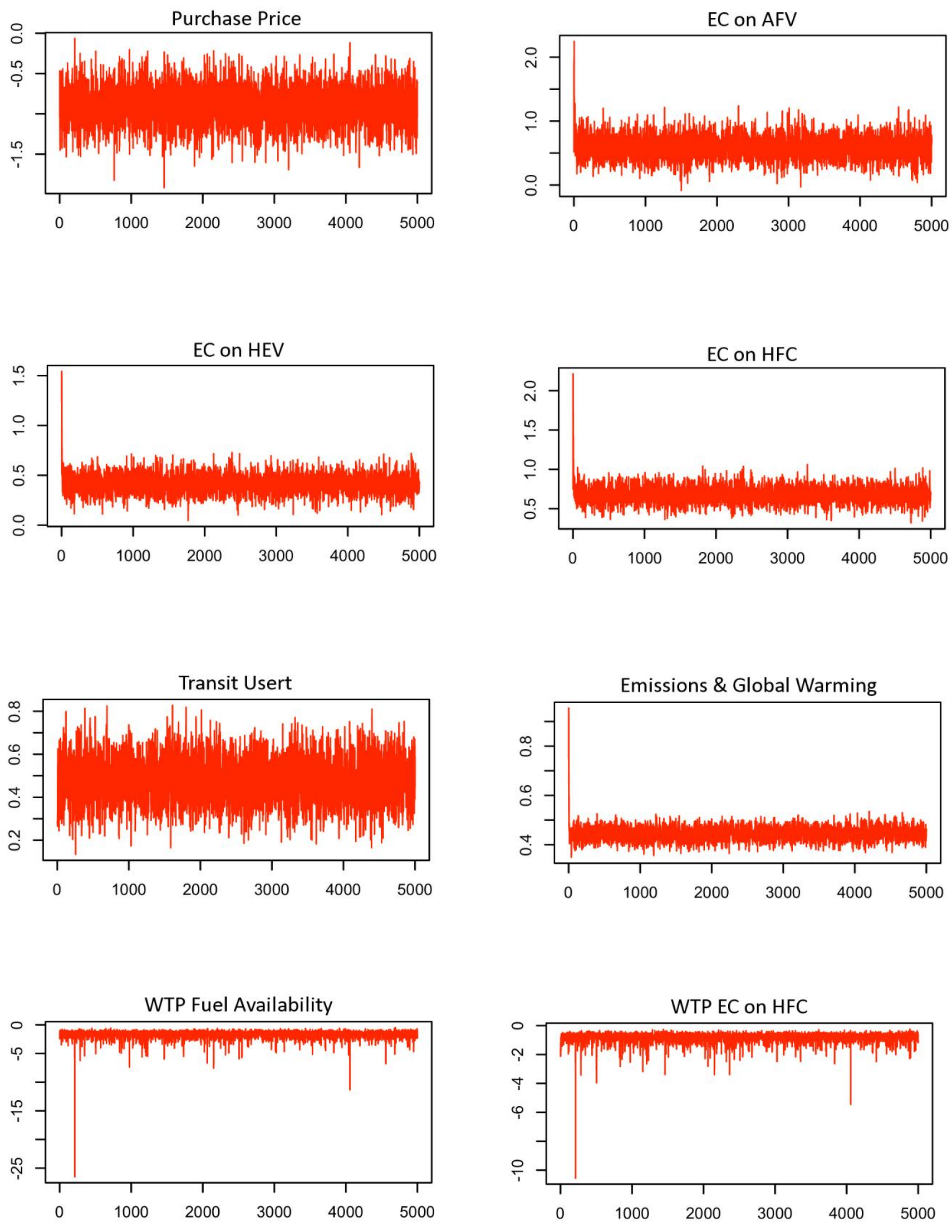

Figure 4.12: MCMC sequence of selected parameters

Chapter 5

Generalization of an MCMC method for Bayesian estimation of HCMs

In this chapter I introduce a general Gibbs sampler for Bayesian estimation of hybrid choice models. Then, using a virtual case of travel mode choice, I discuss the specification, estimation, and point estimate analysis of a hybrid choice model that allows one to include qualitative attributes in a standard discrete choice setting in a way that avoids problems of inconsistency. In particular, I set up a Monte Carlo experiment where I compare the point estimation results of two alternative methods of estimation, namely frequentist full information simulated maximum likelihood and Bayesian Metropolis Hastings-within-Gibbs sampling. Even though the two estimation methods I analyze are based on different philosophies, both the frequentist and Bayesian methods provide estimators that are asymptotically equivalent. The results show that both methods are feasible and offer comparable results with a large enough sample size. However, the Bayesian point estimates outperform maximum likelihood in terms of accuracy, statistical significance, and efficiency when the sample size is low.

5.1 Introduction: Hybrid Choice Modeling

As discussed in previous chapters, HCMs generalize standard DCMs for assessing how perceptions and attitudes, introduced as latent variables, affect choice (?). Whereas perceptual and attitudinal data can be modeled through standard structural equation modeling[1], discrete choice is a latent variable model where the indicator variable corresponds to the alternative that maximizes an unobserved conditional indirect utility. Hence, the econometric representation of an HCM corresponds to the following system of simultaneous equations[2]:

Structural equations

$$\underset{(L\times 1)}{z_n^*} = \Pi z_n^* + Bw_n + \zeta_n = [\mathrm{I}_L - \Pi]^{-1} Bw_n + [\mathrm{I}_L - \Pi]^{-1}\zeta_n, \zeta_n \sim MVN(0, H_{\Psi}^{-1}) \quad (5.1)$$

$$= \underset{(L\times M)}{\tilde{B}} \underset{(M\times 1)}{w_n} + \underset{(L\times 1)}{\tilde{\zeta}_n}, \tilde{\zeta}_n \sim MVN(0, H_{\tilde{\Psi}}^{-1})$$

$$\underset{(J\times 1)}{U_n} = \underset{(J\times K)}{X_n} \underset{(K\times 1)}{\beta} + \underset{(J\times Q)}{W_n^*(X_n, z_n^*)} \underset{(Q\times 1)}{\varrho} + \underset{(J\times L)}{\Gamma} \underset{(L\times 1)}{z_n^*} + \underset{(J\times F)}{P_n} \underset{(F\times F)}{T} \underset{(F\times 1)}{\xi_n} + \underset{(J\times 1)}{\nu_n} \quad (5.2)$$

Measurement equations

$$\underset{(R\times 1)}{I_n} = \underset{(R\times 1)}{\alpha} + \underset{(R\times L)}{\Lambda} \underset{(L\times 1)}{z_n^*} + \underset{(R\times 1)}{\varepsilon_n}, \quad \varepsilon_n \sim MVN(0, H_{\Theta}^{-1}) \quad (5.3)$$

$$\underset{(1\times 1)}{y_n} = i \in C_n \text{ iff } U_{in} - U_{jn} \geq 0, \forall j \in C_n, j \neq i. \quad (5.4)$$

where z_n^* is an endogenous random vector of latent variables; the matrix Π allows for 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 vector of explanatory variables affecting the latent variables; B is a matrix of K unknown regression coefficients used to describe the global effect of $(I_L - \Pi)^{-1}Bw_n$ on the latent variables; and H_{Ψ}^{-1} is a covariance matrix which describes the relationship among the latent variables through the error term. To simplify notation, we define $\tilde{B} = (I_L - \Pi)^{-1}B$, $\tilde{\zeta}_n = (I_L - \Pi)^{-1}\zeta_n$, and $H_{\tilde{\Psi}}^{-1} = [(I_L - \Pi)^{-1}]H_{\Psi}^{-1}[(I_L - \Pi)^{-1}]'$.

The choice model in equation (??) is written in vector form where we assume that there is a total of J_n available alternatives in the set C_n. Hence, U_n is a vector of indirect

[1]Without loss of generality, I adopt a MIMIC model (?).

[2]This system of equations represents a case that is slightly more general than the one introduced in Chapter 3.

utility functions; X_n is a matrix with X_{in} designating its i^{th} row; and β is a vector of unknown parameters. $W_n^*(X_n, z_n^*)$ is a matrix of Q interactions between the observable X_n and the latent z_n^* as well as interactions within the latent variables; ϱ is a vector of unknown parameters associated with these interactions. Γ is a matrix of unknown parameters associated with the latent variables present in the utility function, with Γ_i designating the i^{th} row of matrix Γ. The analytical form of the discrete choice kernel depends on the assumptions regarding the distribution of the random term ν_n.

In the set of measurement equations, I_n corresponds to a vector of manifest variables that serve as indicator responses for the latent variables z_n^*; α is an intercept vector and Λ is a matrix of G unknown factor loadings. The term ε_n is a vector of error terms with covariance matrix H_Θ^{-1}. Finally, we stack the choice indicators y_{in}'s into a vector called y_n.

5.2 HCM Gibbs sampler

If we call θ the joint set of parameters of the choice model[3], the parameters to estimate in an HCM are θ, $\tilde{B}$, α, Λ, $\tilde{\Psi}$, and Θ. Bayes estimation implementation for these parameters requires making draws from the joint posterior distribution:

$$P(\theta, \tilde{B}, \alpha, \Lambda, \tilde{\Psi}, \Theta | y, I), \tag{5.5}$$

or, using data augmentation from:

$$P(z^*, \theta, \tilde{B}, \alpha, \Lambda, \tilde{\Psi}, \Theta | y, I), \tag{5.6}$$

where $z^* = (z_1^*, \ldots, z_N^*)'$, $y = (y_1, \ldots, y_N)'$ and $I = (I_1, \ldots, I_N)'$ capture the information for the full group of individuals.

Using Gibbs sampling[4], the estimators are obtained from draws inside an iterative process

[3]We include in θ the parameters of the deterministic part of the utility function as well as the parameters associated with the random terms, *i.e.* $ASCs, \beta, \varrho, \Gamma, T$, and Σ.

[4]The Gibbs sampler discussed in this chapter is a generalization of the specific Gibbs sampler introduced in Chapter 4.

involving the set of *full conditional distributions*. Namely, at the g-*th* iteration:

$$\begin{aligned}
z_n^{*(g)} &\sim \pi(z_n^*|\theta^{(g-1)}, \tilde{B}^{(g-1)}, \alpha^{(g-1)}, \Lambda^{(g-1)}, \tilde{\Psi}^{(g-1)}, \Theta^{(g-1)}, y_n, I_n), \forall n \\
\theta^{*(g)} &\sim \pi(\theta|z^{*(g)}, \tilde{B}^{(g-1)}, \alpha^{(g-1)}, \Lambda^{(g-1)}, \tilde{\Psi}^{(g-1)}, \Theta^{(g-1)}, y, I) \\
\tilde{B}^{*(g)} &\sim \pi(\tilde{B}|z^{*(g)}, \theta^{(g)}, \alpha^{(g-1)}, \Lambda^{(g-1)}, \tilde{\Psi}^{(g-1)}, \Theta^{(g-1)}, y, I) \\
\alpha^{*(g)} &\sim \pi(\alpha|z^{*(g)}, \theta^{(g)}, \tilde{B}^{(g)}, \Lambda^{(g-1)}, \tilde{\Psi}^{(g-1)}, \Theta^{(g-1)}, y, I) \\
\Lambda^{*(g)} &\sim \pi(\Lambda|z^{*(g)}, \theta^{(g)}, \tilde{B}^{(g)}, \alpha^{(g)}, \tilde{\Psi}^{(g-1)}, \Theta^{(g-1)}, y, I) \\
\tilde{\Psi}^{*(g)} &\sim \pi(\tilde{\Psi}|z^{*(g)}, \theta^{(g)}, \tilde{B}^{(g)}, \alpha^{(g)}, \Lambda^{(g)}, \Theta^{(g-1)}, y, I) \\
\Theta^{*(g)} &\sim \pi(\Theta|z^{*(g)}, \theta^{(g)}, \tilde{B}^{(g)}, \alpha^{(g)}, \Lambda^{(g)}, \tilde{\Psi}^{(g)}, y, I)
\end{aligned}$$

Although the Gibbs sampler is performed simultaneously for the whole HCM, I present the conditional distributions separately for both the MIMIC and discrete choice sub-models.

5.2.1 Conditional distributions of the MIMIC model

From a frequentist point of view[5], the unknown parameters of equations **??** and **??** cannot be estimated using standard regression methods. However, whereas the latent variables are unobservable by definition, Bayesian estimation allows one to augment the observed data by simulating the random latent variables z_n^* through MCMC methods. With the parameter set augmented by z_n^* we obtain a normal linear model posterior such that inference becomes straightforward.

First the data augmentation step for z_n^* will be described. The indicator variables I_n provide information on z_n^* that one needs to take into account in the conditional probability $\pi(z_n^*|I_n)$ to make use of data augmentation techniques. In fact, the indicator variables serve for parameter identification of z_n^*. The joint distribution of z_n^* and I_n is then a key element for deriving the conditional distribution of z_n^*. It is possible to show that

$$\begin{bmatrix} z_n^* \\ I_n \end{bmatrix} \sim MVN\left(\begin{bmatrix} \tilde{B}w_n \\ \alpha + \Lambda\tilde{B}w_n \end{bmatrix}, \begin{bmatrix} \tilde{\Psi} & \tilde{\Psi}\Lambda' \\ \Lambda\tilde{\Psi} & \Lambda\tilde{\Psi}\Lambda' + \Theta \end{bmatrix}\right), \forall n, \tag{5.7}$$

where I_{14} represents the identity matrix of size 14. Equation **??** implies

$$\pi(z_n^*|\theta, \tilde{B}, \alpha, \Lambda, \tilde{\Psi}, \Theta, y_n, I_n) \sim MVN(\mu_{z_n^*|I_n}, \Sigma^2_{z_n^*|I_n}), \forall n, \tag{5.8}$$

[5]Classical estimation of HCMs is introduced and discussed in Chapter 3.

where

$$\mu_{z_n^*|I_n} = \tilde{B}w_n + [\tilde{\Psi}\Lambda'][\Lambda\tilde{\Psi}\Lambda' + \Theta]^{-1}[I_n - \alpha - \Lambda\tilde{B}w_n] \quad (5.9)$$

$$\Sigma^2_{z_n^*|I_n} = \tilde{\Psi} - [\tilde{\Psi}\Lambda'][\Lambda\tilde{\Psi}\Lambda' + \Theta]^{-1}[\Lambda\tilde{\Psi}]. \quad (5.10)$$

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

To estimate the remaining parameters of the latent variable model, we can apply equation **??** to simulate observations of the latent variable z_n^*. Moreover, using the simulated values for z_n^*, equations **??** and **??** become linear regression models with general covariance matrices. First, we rewrite these equations considering the regression coefficients in vector form[6] and the explanatory variables as a design matrix. Then we stack the N observations together. For the structural equation we obtain

$$\underset{(LN\times 1)}{z^*} = \underset{(LN\times K)}{W}\underset{(K\times 1)}{\tilde{b}} + \underset{(LN\times 1)}{\tilde{\zeta}}, \quad \tilde{\zeta} \sim MVN(0, H^{-1}_{\tilde{\Psi}_N}), \quad (5.11)$$

where W is a design matrix containing the elements in w_n, $\forall n$; $\tilde{b}$ is the vector of free parameters in $\tilde{B}$; and Ψ_N is a $LN \times LN$ covariance matrix. For instance, if $\tilde{\zeta}_n$ are assumed to be independent across individuals, then $H^{-1}_{\tilde{\Psi}_N}$ would be a block-diagonal matrix given by

$$H^{-1}_{\tilde{\Psi}_N} = \begin{bmatrix} H^{-1}_{\tilde{\Psi}} & 0_{L\times L} & \cdots & 0_{L\times L} \\ 0_{L\times L} & H^{-1}_{\tilde{\Psi}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{L\times L} \\ 0_{L\times L} & \cdots & 0_{L\times L} & H^{-1}_{\tilde{\Psi}} \end{bmatrix}.$$

For the measurement equation, we get

$$\underset{(RN\times 1)}{I} = \underset{(RN\times 1)}{\alpha} + \underset{(RN\times G)}{Z^*}\underset{(G\times 1)}{\lambda} + \underset{(RN\times 1)}{\varepsilon}, \quad \varepsilon \sim MVN(0, H^{-1}_{\Theta_N}), \quad (5.12)$$

where Z^* is a specification matrix formed by appropriately using the elements in z_n^*, $\forall n$; λ is the vector of free factor loadings in Λ; and $H^{-1}_{\Theta_N}$ is a $LN \times LN$ covariance matrix. For instance, if ε_n is assumed to be independent across individuals, then $H^{-1}_{\Theta_N}$ would be

[6] Note that in equations **??** and **??**, the unknown regression coefficients are written as matrices. This corresponds to the standard SEM notation.

a block-diagonal matrix given by

$$H_{\Theta_N}^{-1} = \begin{bmatrix} H_{\Theta}^{-1} & 0_{R\times R} & \cdots & 0_{R\times R} \\ 0_{R\times R} & H_{\Theta}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{R\times R} \\ 0_{R\times R} & \cdots & 0_{R\times R} & H_{\Theta}^{-1} \end{bmatrix}.$$

If prior beliefs for b and λ are described by $p(b) \sim MVN(\check{b}, \check{V}_b)$ and $p(\lambda) \sim MVN(\check{\lambda}, \check{V}_\lambda)$ respectively, then it can verified that, conditional on the other parameters of the model, the posteriors of b and λ are multivariate normal:

$$\pi(b|Z^*, \theta, b, \lambda, y, I) \sim MVN(\bar{b}, \bar{V}_b) \tag{5.13}$$

$$\pi(\Lambda|Z^*, \theta, b, y, I) \sim MVN(\bar{\lambda}, \bar{V}_\lambda), \tag{5.14}$$

where

$$\bar{V}_b = (\check{V}_b^{-1} + W' H_{\tilde{\Psi}_N} W)^{-1}, \;\; \bar{b} = \bar{V}_b(\check{V}_b^{-1} + W' H_{\tilde{\Psi}_N} Z^*) \tag{5.15}$$

$$\bar{V}_\lambda = (\check{V}_\lambda^{-1} + Z^{*\prime} H_{\Theta_N} Z^*)^{-1}, \;\; \bar{\lambda} = \bar{V}_\lambda(\check{V}_\lambda^{-1} + Z^{*\prime} H_{\Theta_N} I). \tag{5.16}$$

The conditional posterior for a general covariance matrix, either for $H_{\tilde{\Psi}_N}^{-1}$ or $H_{\Theta_N}^{-1}$, does not have an easily recognized form. However, as for any linear model with general covariance matrix, it is possible to derive appropriate posterior simulators for particular covariance structures. For instance, if the error terms are assumed to be i.i.d., then the resulting block-diagonal structure, combined with Wishart prior beliefs $p(H_{\tilde{\Psi}}) \sim W(\check{\nu}_{\tilde{\Psi}}, \check{H}_{\tilde{\Psi}})$, and $p(H_\Theta) \sim W(\check{\nu}_\Theta, \check{H}_\Theta)$ allow us to obtain

$$\pi(H_{\tilde{\Psi}}) \sim W(\bar{\nu}_{\tilde{\Psi}}, \bar{H}_{\tilde{\Psi}}) \tag{5.17}$$

$$\pi(H_\Theta) \sim W(\bar{\nu}_\Theta, \bar{H}_\Theta), \tag{5.18}$$

where

$$\bar{\nu}_{\tilde{\Psi}} = \check{\nu}_{\tilde{\Psi}} + N, \;\; \bar{H}_{\tilde{\Psi}}^{-1} = \check{H}_{\tilde{\Psi}}^{-1} + \sum_{n=1}^{N} \tilde{\zeta}_n \tilde{\zeta}_n' \tag{5.19}$$

$$\bar{\nu}_\Theta = \check{\nu}_\Theta + N, \;\; \bar{H}_\Theta^{-1} = \check{H}_\Theta^{-1} + \sum_{n=1}^{N} \varepsilon_n \varepsilon_n'. \tag{5.20}$$

5.2.2 Conditional distributions of the discrete choice model

The HCM Gibbs sampler is completed by determining the posterior simulator for the discrete choice model, which is a particular case of structural equation modeling. Note that the vector of conditional indirect utility functions is an unobservable dependent variable. As discussed previously, having a latent dependent variable does not allow us to consider the estimation problem as a classical regression problem. However, once again the use of data augmentation allows us to treat equation **??** as a standard regression. Note that the analytical form of the choice model depends on the assumptions regarding the distribution of the random term ν_n defined in equation **??**. In fact, as I show below, the use of data augmentation is straightforward only when working with a multinomial probit model.[7]

Just as in the case of the MIMIC model, the measurement equation provides identification of the latent variable in the structural equation. However, because the choice indicator depends on a discrete maximization, in discrete choice models only utility differences can be identified. Hence, we work with utility differences with respect to an arbitrary base alternative.[8] Let $\Delta_1(\cdot)_{jn} = (\cdot)_{jn} - (\cdot)_{1n}$ be a matrix difference operator. For example, $\Delta_1 U_n$ takes each element of U_n and subtracts the base element U_{1n} such that

$$\underset{(J-1\times 1)}{\Delta_1 U_n} = \Delta_1 \begin{bmatrix} U_{1n} \\ U_{2n} \\ \vdots \\ U_{Jn} \end{bmatrix} = \begin{bmatrix} U_{2n} - U_{1n} \\ \vdots \\ U_{Jn} - U_{1n} \end{bmatrix}.$$

If we rewrite equation **??** in stacked form and consider the regression coefficients in vector form, we get a regression expression with an unobserved dependent variable, unobserved explanatory variables and interactions between the observed and unobserved attributes:

$$\begin{aligned} \Delta_1 U &= \Delta_1 X\beta + \Delta_1 W^*(X, Z^*)\varrho + \Delta_1 Z^*\gamma + \Delta_1 PT\xi + \Delta_1\nu, \quad \Delta_1\nu \sim MVN(0, \Sigma_N), \\ \Delta_1 U &= X_\Delta\theta + \Delta_1 PT\xi + \Delta_1\nu \end{aligned} \tag{5.21}$$

where $\theta' = (\beta', \varrho', \gamma')$ is the vector of regression coefficients of the utility function; X_Δ is an extended attribute matrix built by appropriately stacking the matrices $\Delta_1 X$,

[7]We can derive a probit kernel if we make the assumption that the error terms ν_n are i.i.d. multivariate normal, i.e. $\nu_n \sim MVN(0, \Sigma), \forall n$.

[8]Without loss of generality we take the first alternative as base.

$\Delta_1 W^*(X, Z^*)$ and $\Delta_1 Z^*$; and where

$$\underset{((J-1)N\times(J-1)N)}{\Sigma_N} = \Delta_1 \begin{bmatrix} \Sigma & 0_{J\times J} & \cdots & 0_{J\times J} \\ 0_{J\times J} & \Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{J\times J} \\ 0_{J\times J} & \cdots & 0_{J\times J} & \Sigma \end{bmatrix} \Delta_1'.$$

We get a standard discrete choice model in the case without unobserved attributes. By definition an HCM considers the presence of the endogenous set of latent variables. However, in the previous subsection we developed a simulator for the latent variable Z^*. Therefore, if we take the unconditional observations of the latent attributes, the terms Z^* and $W^*(X, Z^*)$ simply enter equation **??** as standard observable exogenous attributes. In addition, given $PT\xi$, if we simulate observations for the latent utility function then equation **??** transforms into a linear regression model with a block-diagonal covariance matrix. In the case of a probit kernel, the properties of the normal distribution make it straightforward to exploit data augmentation techniques for performing simulations for the latent utility function, basically because the utility function follows a normal distribution. However we need to describe the conditional distribution of the utility function taking into account the choice indicators y_n. Since

$$y_n = \begin{cases} 1 & \text{if } \max \Delta_1 U_{in} \leq 0 \\ i & \text{if } \Delta_1 U_{in} > \max\{0, \Delta_1 U_{-in}\} \end{cases}, \tag{5.22}$$

where U_{-in} represents the set of all utility functions except U_{in}, then conditional on y_n, $\Delta_1 U_n$ follows a truncated multivariate normal (TMVN) distribution:

$$\pi(\Delta_1 U_n | Z^*, \theta, T, \xi, \Sigma_\Delta, y_n) \sim \text{TMVN}_{\Re|y_n}(X_\Delta \theta + \Delta_1 PT\xi, \Sigma_\Delta), \forall n., \tag{5.23}$$

where $\Sigma_\Delta = \Delta_1 \Sigma \Delta_1'$ corresponds to the $(J-1 \times J-1)$ covariance matrix of the utility-difference error term $\Delta_1 \nu_n \sim MVN(0, \Sigma_\Delta)$; and where the truncation region $\Re$ is defined by the inequalities in the measurement equation **??**.

Although data augmentation transforms the estimation problem of the discrete choice kernel into a Bayesian regression, simulations required for Σ_Δ must address some identification issues. Because we are working with the equation that considers utility differences, using the information contained in the covariance matrix of the utility-difference model Σ_Δ it is not possible to identify the $J(J-1)$ elements in the original covariance matrix Σ. Standard practice is to set the scale of the model by fixing the first diagonal element of Σ_Δ such that $\sigma^2_{\Delta,11} \equiv V(\Delta_1 \nu_{1n}) = V(\nu_{2n} - \nu_{1n}) = 1$. Following **?**, it is possible to generate Wishart draws given a diagonal element if we assume the Wishart prior $p(\Sigma_\Delta^{-1}) \sim W(\check{\nu}_{\Sigma_\Delta}, \check{\Sigma}_\Delta^{-1}) \mid_{\sigma^2_{\Delta,11}=1}$ such that

$$\pi(\Sigma_\Delta^{-1}) \sim W(\bar{\nu}_{\Sigma_\Delta}, \bar{\Sigma}_\Delta^{-1}) \mid_{\sigma^2_{\Delta,11}=1}, \tag{5.24}$$

where

$$\bar{\nu}_{\Sigma_\Delta} = \check{\nu}_{\Sigma_\Delta} + N, \;\; \bar{\Sigma}_\Delta = \check{\Sigma}_\Delta + \sum_{n=1}^{N} \Delta_1 \nu_n \nu_n' \Delta_1'. \tag{5.25}$$

Finally, we take $p(\theta) \sim MVN(\check{\theta}, \check{V}_\theta)$ as prior belief, and the regression coefficients of the discrete choice kernel can be sampled from the following posterior conditional distribution:

$$\pi(\theta | Z^*, \Delta_1 U_n, T, \xi, \Sigma_\Delta, y_n) \sim MVN(\bar{\theta}, \bar{V}_\theta), \tag{5.26}$$

where

$$\bar{V}_\theta = (\check{V}_\theta^{-1} + X_\Delta' {\Sigma_\Delta}^{-1} X_\Delta)^{-1}, \;\; \bar{\theta} = \bar{V}_\theta (\check{V}_\theta^{-1} + {X_\Delta}' {\Sigma_\Delta}^{-1} (\Delta_1 U - PT\xi)). \tag{5.27}$$

In the literature, the usual way to incorporate flexible error structures into a discrete choice model is by assuming a mixed logit (MMNL) model. However, when considering an MMNL kernel we no longer have the advantageous properties that make implementing the probit-kernel HCM Gibbs sampler straightforward. Effectively, in the case of a MMNL kernel there is no closed form full conditional distribution for $\widetilde{U}$, and hence it is not possible to augment the data. Thus, MMNL Bayesian estimation does not allow us to use an ordinary regression for θ and, as I will show, the use of Metropolis-Hastings (MH) methods is needed. However, the Gibbs sampler for the MIMIC portion of the HCM is still valid, and hence the estimation problem reduces to an MH-within a Gibbs sampler. This MH-within-Gibbs algorithm is simpler to derive than it may sound, because the Bayesian procedure for a standard MMNL – without latent explanatory variables – described by **?** can simply be plugged into the HCM Gibbs sampler for the MIMIC model.

I now describe the MH-within-Gibbs algorithm for the multinomial logit (MNL) kernel. The MNL model is a particular case of the mixed logit model where the taste parameters are fixed to the population means. In the MNL case, since there is no closed form full conditional distribution for θ, we cannot implement a Gibbs sampler with data augmentation for the utility function. However, we can use an asymptotic approximation to the posterior (**?**) that will be used for deriving an MH algorithm for θ:

$$\pi(\theta | \mathrm{EC}, b, \Lambda, y_n, I_n) \propto |H|^{\frac{1}{2}} \exp\left(\frac{1}{2}(\theta - \hat{\theta}_{ML})' H (\theta - \hat{\theta}_{ML})\right), \tag{5.28}$$

with $\hat{\theta}_{ML}$ being the maximum likelihood solution for θ, i.e. the value $\hat{\theta}_{ML}(y)$ that maximizes the likelihood function $\ell(y;\theta)$ once y is observed, and with H being the asymptotic variance obtained from the expected sample information matrix ($\otimes$ denotes the Kronecker product):

$$H = -E\left[\frac{\partial^2 \ln \ell}{\partial\theta\partial\theta'}\right] = -\sum_{n=1}^{N}(\mathrm{diag}(P_n) - P_n P_n') \otimes \widetilde{X}_n \widetilde{X}_n', \tag{5.29}$$

which is the Hessian matrix of the observed MNL log-likelihood $\ln \ell = \sum_{n=1}^{N} \ln P_{y_n}$, where $P_n = (P_{1n}, \ldots, P_{J_n n})$, with P_{in} being the standard MNL form of the choice probability of alternative i for individual n:

$$P_{in} = \frac{e^{\widetilde{X}_{in}\theta_n}}{\sum_{j=1}^{J_n} e^{\widetilde{X}_{jn}\theta_n}}. \tag{5.30}$$

Thus, a candidate $\theta^{cand} \in \Theta$ is drawn from the transition probability $q(\theta^{cand}|\theta^{curr})$ of generating candidate θ^{cand} given $\theta^{curr} \in \Theta$, such that $\theta^{curr} \sim p(\theta, y)$. The candidate realization θ^{cand} is then compared to the current $\theta^{curr} \in \Theta$ through the acceptance ratio:

$$\alpha = \min\left\{1, \frac{p(y|\theta^{cand})p(\theta^{cand})}{p(y|\theta^{curr})p(\theta^{curr})} \cdot \frac{q(\theta^{cand}|\theta^{curr})}{q(\theta^{curr}|\theta^{cand})}\right\}.$$

Starting with an arbitrary value $\theta^{(0)}$, in the MH algorithm at the g^{th} iteration the candidate is accepted as the new $\theta^{(g)} = \theta^{cand}$ with probability α, while the old one is preserved $\theta^{(g)} = \theta^{curr}$ with probability $1-\alpha$. In a random-walk Metropolis chain, the candidate realization is defined as $\theta^{cand} = \theta^{curr} + \varepsilon$, where $\varepsilon \sim N(0, s^2 H^{-1})$ and s^2 is the precision. The candidate generating process is a Metropolis independence chain if $\theta^{cand} \sim MSt(\upsilon, \hat{\theta}_{ML}, s^2 H^{-1})$, i.e. θ^{cand} is drawn from a multivariate t distribution with mean $\hat{\theta}_{ML}$, dispersion $s^2 H^{-1}$, and υ degrees of freedom.

5.2.3 Bayesian estimates of the HCM

Under fairly mild conditions (**?**) and for a sufficiently large number of draws, the Gibbs sampler sequence of random draws forms an irreducible and ergodic Markov chain converging at a exponential rate to the joint posterior distribution. In practice, the Bayesian estimates are calculated taking the sample means of the Gibbs sampler draws, as in

$$\hat{\theta} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\theta^{(g)}, \hat{\tilde{B}} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\tilde{B}^{(g)}, \hat{\alpha} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\alpha^{(g)} \tag{5.31}$$
$$\hat{\Lambda} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\Lambda^{(g)}, \hat{\tilde{\Psi}} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\tilde{\Psi}^{(g)}, \hat{\Omega} = \mathcal{G}^{-1}\sum_{g=1}^{\mathcal{G}}\tilde{\Omega}^{(g)}.$$

The mean of the Gibbs sampler draws – the Bayesian estimates – are consistent estimators of the corresponding posterior means (**?**). Even though it is complex to derive analytic forms for the covariance matrices of the parameters, consistent estimates of these matrices can be obtained from the sample covariance matrices implied by the Gibbs sampler. In other words, the standard deviations used for the calculation of t-statistics are simply the standard deviations of the artificial samples generated by the Gibbs sampler:

$$\begin{aligned}
\hat{\mathbb{V}}(\theta|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\theta^{(g)}-\hat{\theta})(\theta^{(g)}-\hat{\theta})' \qquad (5.32)\\
\hat{\mathbb{V}}(\tilde{B}|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\tilde{B}^{(g)}-\hat{\tilde{B}})(\tilde{B}^{(g)}-\hat{\tilde{B}})'\\
\hat{\mathbb{V}}(\alpha|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\alpha^{(g)}-\hat{\alpha})(\alpha^{(g)}-\hat{\alpha})'\\
\hat{\mathbb{V}}(\Lambda|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\Lambda^{(g)}-\hat{\Lambda})(\Lambda^{(g)}-\hat{\Lambda})'\\
\hat{\mathbb{V}}(\tilde{\Psi}|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\tilde{\Psi}^{(g)}-\hat{\tilde{\Psi}})(\tilde{\Psi}^{(g)}-\hat{\tilde{\Psi}})'\\
\hat{\mathbb{V}}(\Omega|y,I) &= (\mathcal{G}-1)^{-1}\sum_{g=1}^{\mathcal{G}}(\Omega^{(g)}-\hat{\Omega})(\Omega^{(g)}-\hat{\Omega})'
\end{aligned}$$

5.3 Pre-posterior analysis

To test the Bayesian estimation of HCMs, in this section I carry out a pre-posterior analysis based on simulated data. To construct the choice situation, we take the example first outlined by **?**. In line with the development of discrete choice models, where travel behavior modeling has played a major role, this example considers a tri-modal choice.[9] Although for the sake of simplicity we study here a specific travel mode choice situation, the equations and analysis can be easily generalized. Consider three travel modes characterized not only by the standard attributes travel time and travel cost, but also by two alternative-specific qualitative attributes: *comfort* and *convenience*. Qualitative attributes, as opposed to quantitative attributes (such as travel time), do not have a natural order or an overt measurement scale. As discussed in Chapter 2, qualitative attributes are often introduced as categorical variables on a nominal scale. The nominal scale may be adequate if the qualitative attribute is discrete in nature. But if there is some continuity in the evaluation of quality, the nominal scale becomes a proxy variable that measures the true qualitative attribute with error. If we omit a relevant qualitative attribute, it is clear that the estimators will be biased. Including the qualitative attribute using a proxy variable does not offer much of a solution, because endogeneity is still a problem. The HCM system of equations accounts for the real nature of the qualitative attributes, and avoids biased and inconsistent estimators.

In my example, each of the three modes of transportation is characterized by the observable attributes *travel time* x_{i1n} and *travel cost* x_{i2n}.[10] In addition, two alternative-specific unobservable attributes are considered: *comfort* z^*_{in1} and *convenience* z^*_{in2}, as perceived by each individual n for mode i. Under these assumptions and without considering any interaction nor any additional error component, the utility function of individual n (from equation) corresponds to the following parametric linear-in-parameter specification:

$$\begin{bmatrix} U_{1n} \\ U_{2n} \\ U_{3n} \end{bmatrix} = \begin{bmatrix} 0 & 0 & x_{11n} & x_{12n} \\ 1 & 0 & x_{21n} & x_{22n} \\ 0 & 1 & x_{31n} & x_{32n} \end{bmatrix} \begin{bmatrix} \mathrm{ASC}_2 \\ \mathrm{ASC}_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \Gamma_1 & \Gamma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Gamma_3 & \Gamma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Gamma_5 & \Gamma_6 \end{bmatrix} \begin{bmatrix} z^*_{11n} \\ z^*_{12n} \\ z^*_{21n} \\ z^*_{22n} \\ z^*_{31n} \\ z^*_{32n} \end{bmatrix} + \begin{bmatrix} \nu_1 n \\ \nu_2 n \\ \nu_3 n \end{bmatrix}. \tag{5.33}$$

Travel time and travel cost of each alternative are represented through $x_{\dot{1}n}$ and $x_{\dot{2}n}$, respectively, with generic marginal utilities β_{1n} and β_{2n}. Instead of introducing individual attitudes (which have mostly motivated the study of HCMs, as discussed in Chapter 2), when modeling qualitative attributes we can consider alternative-specific latent variables. For instance, the unobservable-to-the-modeler comfort experienced by a particular individual for one alternative is not necessarily the same as the perceived comfort for a

[9]Since we are working with simulated data, the modal choice situation was chosen for no other than an illustrative purpose.

[10]The original example considered only one attribute. In this chapter we include a second attribute because we want to test how well parameter ratios are recovered.

competing alternative. Thus, each qualitative attribute (comfort and convenience) is operationalized in 3 latent variables, leading to the 6 variables $z^*_{11n}, z^*_{12n}, z^*_{21n} z^*_{22n}, z^*_{31n}, z^*_{32n}$. For example, z^*_{11n} is comfort of alternative 1, z^*_{31n} represents comfort of alternative 3, and z^*_{22n} is convenience of alternative 2. Conversely, attitudes usually are individual-specific latent variables. In this example we considered alternative-specific taste parameters for the latent variables (Γ), but generic marginal utilities are also possible for qualitative attributes.

Equation **??** represents the standard structural equation of a discrete choice model. Depending on the assumptions made for the distribution of the random error term, different HCM kernels can be derived (see Ben-Akiva and Lerman, 1985.) I assume $\nu_{(\cdot n)} \sim Gumbel(0, 1)$ leading to a normalized MNL kernel. Recognizing the latent nature of the qualitative attributes we need to introduce structural equation modeling (SEM), such as a multiple indicators multiple causes (MIMIC) model. Note that most recent applications of HCMs have considered the introduction of attitudes as the latent attribute (see for example Chapter 4, where I introduced environmental concerns as an attitudinal variable in a vehicle choice context.) When the latent attributes are attitudes, natural explanatory variables in the structural equation are socioeconomic variables. However, socio-demographic variables are not always causal, which is especially the case in latent constructs representing qualitative attributes. For instance, level of service variables such as frequencies and parking availability may serve as explanatory variables of the qualitative convenience of travel modes. In addition, measurable attributes of the alternatives (or proxies of the alternative attributes) as well as perceived levels of service serve as indicators of the qualitative attribute. For example, perceived levels of relaxation and ease can be taken as indicators for convenience. In our virtual case study, we consider structural equations for the qualitative attributes with two explanatory variables (individual-specific w_{1n} and w_{2n} in this example, but these variables can be alternative-specific as well), and with generic effects across alternatives 2 and 3 (in the parameters $b_\cdot$), as in

$$\begin{bmatrix} z^*_{11n} \\ z^*_{12n} \\ z^*_{21n} \\ z^*_{22n} \\ z^*_{31n} \\ z^*_{32n} \end{bmatrix} = \begin{bmatrix} b_1 & 0 \\ b_2 & b_3 \\ b_4 & b_5 \\ b_6 & b_7 \\ b_4 & b_5 \\ b_6 & b_7 \end{bmatrix} \begin{bmatrix} w_{1n} \\ w_{2n} \end{bmatrix} + \begin{bmatrix} \zeta_{11n} \\ \zeta_{12n} \\ \zeta_{21n} \\ \zeta_{22n} \\ \zeta_{31n} \\ \zeta_{32n} \end{bmatrix}, \tag{5.34}$$

where we assume that the error terms are $\zeta_{\cdot n} \sim N(0, 1)$. Since the measurement scale of the qualitative attributes is unknown, by setting a unit variance for each latent variable we normalize the model.

Since the latent variables are unobserved, we need to introduce indicator variables. For instance, let I_{i1n}, I_{i2n} and I_{i3n} denote *relaxation*, *reliability* and *ease* of mode i as perceived by individual n, respectively. We assume that both *relaxation* and *reliability* serve to measure *comfort*, whereas the indicator variables *reliability* and *ease* provide identification for *convenience*. Making the effect to be generic across modes 2 and 3, we get:

$$\begin{bmatrix} I_{11n} \\ I_{12n} \\ I_{13n} \\ I_{21n} \\ I_{22n} \\ I_{23n} \\ I_{31n} \\ I_{32n} \\ I_{33n} \end{bmatrix} = \begin{bmatrix} z^*_{11n} & z^*_{12n} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & z^*_{11n} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & z^*_{12n} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & z^*_{21n} & z^*_{22n} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & z^*_{21n} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & z^*_{22n} \\ 0 & 0 & 0 & 0 & z^*_{31n} & z^*_{32n} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & z^*_{31n} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & z^*_{32n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11n} \\ \varepsilon_{12n} \\ \varepsilon_{13n} \\ \varepsilon_{21n} \\ \varepsilon_{22n} \\ \varepsilon_{23n} \\ \varepsilon_{31n} \\ \varepsilon_{32n} \\ \varepsilon_{33n} \end{bmatrix}, \tag{5.35}$$

where $\varepsilon_{\cdot n} \sim N(0, \omega_{\cdot}^2)$. The variances $\omega_{\cdot}^2$ do not need to be the same. Even though a heteroscedastic general covariance matrix is possible for the measurement equations, because of identification restrictions the measurement equations cannot be correlated. If the indicator variables are continuous, no further identification restrictions are necessary.

The HCM of travel mode choice is given by the simultaneous system of equations **??**, **??**, and **??**. This system simultaneously permits identification of the qualitative attributes, of the effects of these on choice through the utility function effect, as well as identification of the parameters of the structural equation. The latter is particularly relevant for policy analysis, because once the model is estimated there is no need for the measurement equation and effects on choice produced by changes in the qualitative attributes come only via the structural relationship. In Figure **??** we sketch the path diagram of the HCM system of equations representing the virtual travel mode choice that we analyze in the present chapter.

I discuss now the parametric assumptions taken for implementation of the Monte Carlo experiment. I consider a fixed value for θ. The parameters that define θ were calibrated to assure both that the choice process is neither completely deterministic nor completely random and that the choices are relatively balanced. The population was set to $N = 100,000$ individuals. The attributes were generated as independent random terms with continuous uniform distributions $x_{11} \overset{iid}{\sim} U(0,1)$, $x_{21} \overset{iid}{\sim} U(0,1)$, $x_{12} \overset{iid}{\sim} U(0,1)$, $x_{22} \overset{iid}{\sim} U(0,1)$, $x_{31} \overset{iid}{\sim} U(0,3)$, and $x_{32} \overset{iid}{\sim} U(0,1)$. Without loss of generality the explanatory variables for the latent variables were assumed following a discrete Bernoulli distribution, $w_1, w_2 \overset{iid}{\sim} Bernoulli(p = 0.5)$. Error terms of both the structural and measurement equations of the latent variables were drawn from standard normal distributions, $\zeta_{jk} \overset{iid}{\sim} N(0,1), j \in \{1,2,3\}, k \in \{1,2\}$. $\varepsilon_{jk} \overset{iid}{\sim} N(0,1), j \in \{1,2,3\}, k \in \{1,2,3\}$. Finally, the error term for the utility function was drawn according to $\nu_i \overset{iid}{\sim} Gumbel(0,1), i \in \{1,2,3\}$.

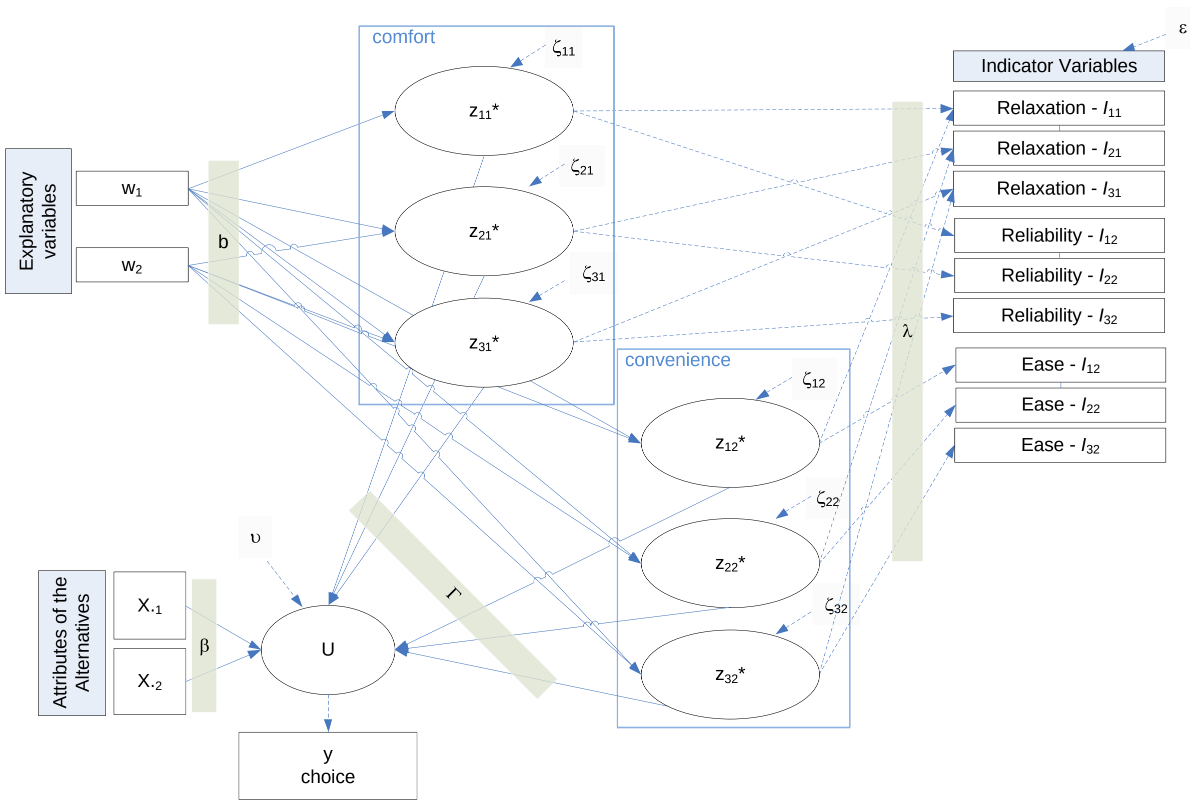

Figure 5.1: Path diagram of the virtual travel mode choice case

5.3.1 Classical and Bayesian estimation of an MNL kernel

In this section I perform a Monte Carlo study for studying point estimation of θ for both the frequentist[11] and Bayesian[12] estimators. For estimating the model, both frequentist and Bayesian estimation use simulation. In the frequentist case, the estimator $\hat{\theta}_{ML}$ is

[11]Frequentist estimation of HCMs is based on full information simulated maximum likelihood. See Chapter 2 for a discussion of this method.

[12]The Gibbs sampler derived in the previous section of this chapter.

given by the simulated maximum likelihood solution which depends on the number of repetitions of the simulator. For each iteration of the simulator, an average likelihood is calculated. In the case of the Bayesian simulator, each iteration defines a draw from the posterior distribution. Then, the Bayesian point estimator $\hat{\theta}$ is obtained from the posterior means. In both cases, as the number of repetitions increases we get a better approximation of θ. However, increasing the number of repetitions is highly resource consuming. In practice, we use a given number of repetitions that assures a certain level of parameter accuracy. Another critical factor in point estimation is sample size. Recall that the properties of frequentist estimators are asymptotic, in the sense that we know that frequentist estimators work in large samples, but it is not clear whether frequentist estimators will perform well for small samples. Considering both the number of repetitions and the sample size as experimental variables, I construct the simulation plan for a Monte Carlo study defined in Table **??** below.

Bayesian		Frequentist	
Sample Size	Repetitions	Sample Size	Repetitions
150	500	150	25
150	2500	150	100
150	10000	150	250
500	500	500	25
500	2500	500	100
500	10000	500	250

Table 5.1: Simulation Plan

For each sample size ($N = 150, 500$) we take 50 random samples from the simulated population, and then we estimate the model with each method, varying the number of repetitions used for the respective simulators. The number of FISML repetitions is lower because we seek to get a point estimate, but for Bayesian estimation we need a higher number because we are mapping a whole distribution. Note that the simulation results correspond to the average of the sampling process.

For assessing the presence of eventual bias I compare the true parameter recovery, specifically through a t-stat against the true δ_0, as well as a summary measure of distance between the point estimates and the true δ_0. Effectively, as a measure of general accuracy for each simulation case I report the Euclidean distance between both $\hat{\delta}$ and δ, as in

$$\mathcal{D}(\hat{\delta}, \delta_0) = \sqrt{(\hat{\delta} - \delta_0) \cdot (\hat{\delta} - \delta_0)} = \sqrt{\sum_{p=1}^{P} \left(\hat{\delta}_p - \delta_{0p}\right)^2}, \tag{5.36}$$

which provides a scalar measurement of the closeness between the average estimates and the true parameters used to generate the data.

The HCM point estimates using both estimation methods are presented in Tables **??**-tab5-5. A first important issue is that frequentist estimation with a relatively low number of repetitions (25 Halton draws) failed to converge for some of the samples in the repeated sampling process. Specifically, frequentist estimation with 25 Halton draws failed to converge in 33% of the samples for N=150, and in 20% of the samples for N=500. It is important to recall that the since we have 6 latent variables (2 qualitative attributes that are specific for each of the 3 alternatives), the Halton draws are being used to approximate a 6-dimensional integral. The calculated average results we provide consider only the samples where convergence was achieved. No convergence problems for the maximum likelihood simulator were observed for the higher numbers of Halton draws considered.

We are interested, first, in the accuracy of the estimates in the sense of the capacity of the model to reproduce the true parameters. To analyze how accurate the estimates are we look at the t-statistic under the null hypothesis $H_0 : \hat{\delta}_i = \delta_{0i}$. These individual statistics are reported as *t-target* in the tables. In the case of Bayesian estimation, all of the estimated parameters reproduce the true values at the 95% confidence level, independently of both the sample size and number of Bayesian repetitions. Frequentist estimates recover the true parameters at the 95% level when the number of Halton draws is high enough. However, when the number of repetitions is too low, problems are detected. In fact, for the larger sample size (N=500), using just 25 Halton draws yields serious problems in reproducing the true values of the measurement parameters for the qualitative attributes. As discussed above, to facilitate comparing the accuracy of the

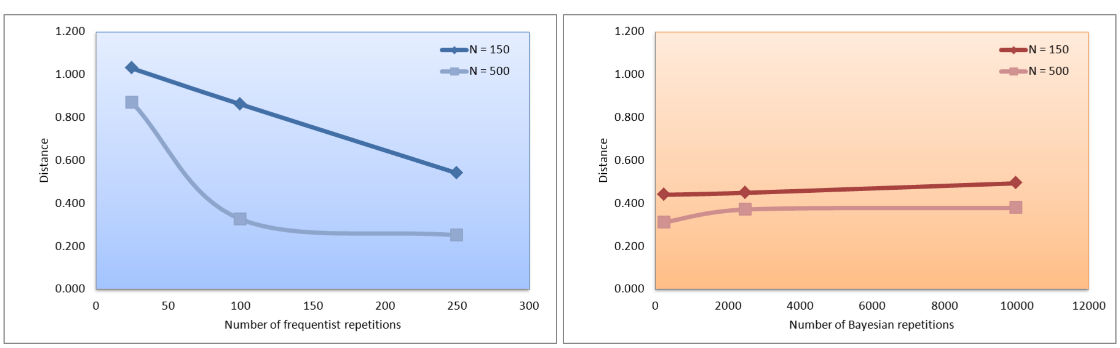

Figure 5.2: Accuracy of the frequentist and Bayesian estimates

δ		500 Bayes repetitions			2500 Bayes repetitions			10000 Bayes repetitions		
		est	t-stat	t-target	est	t-stat	t-target	est	t-stat	t-target
ASC_2	-0.6	-0.422	-1.24	0.52	-0.598	-1.71	0.01	-0.554	-1.63	0.14
ASC_3	0.8	0.768	1.60	-0.07	0.914	1.90	0.24	0.755	1.64	-0.10
β_1	-1.2	-1.242	-4.78	-0.16	-1.113	-4.45	0.35	-1.063	-4.43	0.57
β_2	-0.8	-0.774	-3.37	0.11	-0.832	-3.47	-0.13	-0.805	-3.35	-0.02
Γ_1	-0.4	-0.396	-1.58	0.02	-0.371	-1.48	0.12	-0.402	-1.68	-0.01
Γ_2	0.5	0.503	2.19	0.01	0.482	2.19	-0.08	0.406	1.93	-0.45
Γ_3	0.5	0.452	1.88	-0.2	0.546	2.10	0.18	0.449	1.80	-0.20
Γ_4	-0.6	-0.468	-1.67	0.47	-0.562	-1.96	0.13	-0.563	-1.98	0.13
Γ_5	0.5	0.543	2.17	0.17	0.459	1.91	-0.17	0.493	2.05	-0.03
Γ_6	0.6	0.566	2.10	-0.13	0.493	1.76	-0.38	0.517	1.91	-0.31
b_1	0.5	0.486	2.86	-0.08	0.525	3.09	0.15	0.592	3.29	0.51
b_2	0.8	0.988	3.53	0.67	1.030	3.12	0.70	1.103	2.45	0.67
b_3	1.5	1.772	5.06	0.78	1.744	4.15	0.58	1.742	4.58	0.64
b_4	0.3	0.363	2.59	0.45	0.303	2.33	0.02	0.318	2.27	0.13
b_5	1.8	1.852	9.75	0.27	1.750	9.72	-0.28	1.897	9.49	0.49
b_6	0.5	0.551	4.24	0.39	0.540	4.15	0.31	0.515	3.96	0.12
b_7	1.0	1.014	6.76	0.09	1.145	7.16	0.91	1.067	7.11	0.45
λ_1	0.7	0.659	5.49	-0.34	0.611	5.09	-0.74	0.687	5.73	-0.11
λ_2	0.5	0.459	5.10	-0.46	0.439	4.88	-0.68	0.392	4.36	-1.20
λ_3	1.0	0.996	8.30	-0.03	0.991	9.01	-0.08	0.971	8.83	-0.26
λ_4	0.7	0.612	6.80	-0.98	0.652	5.93	-0.44	0.586	5.33	-1.04
λ_5	0.5	0.461	6.59	-0.56	0.482	6.03	-0.23	0.478	5.98	-0.28
λ_6	0.8	0.757	9.46	-0.54	0.732	9.15	-0.85	0.763	8.48	-0.41
λ_7	1.0	0.959	13.70	-0.59	1.025	12.81	0.31	0.968	12.10	-0.40
λ_8	1.0	0.960	12.00	-0.50	0.925	11.56	-0.94	0.980	12.25	-0.25

Table 5.2: Bayesian point estimates, N=150

δ		500 Bayes repetitions			2500 Bayes repetitions			10000 Bayes repetitions		
		est	t-stat	t-target	est	t-stat	t-target	est	t-stat	t-target
ASC_2	-0.6	-0.585	-3.25	0.08	-0.512	-2.84	0.49	-0.546	-3.03	0.30
ASC_3	0.8	0.819	3.28	0.08	0.769	3.20	-0.13	0.731	2.92	-0.28
β_1	-1.2	-1.176	-8.40	0.17	-1.099	-8.45	0.78	-1.129	-8.68	0.55
β_2	-0.8	-0.816	-6.28	-0.12	-0.799	-6.15	0.01	-0.720	-6.00	0.67
Γ_1	-0.4	-0.376	-2.89	0.18	-0.271	-2.08	0.99	-0.363	-2.79	0.28
Γ_2	0.5	0.465	3.88	-0.29	0.371	3.09	-1.08	0.398	3.32	-0.85
Γ_3	0.5	0.491	3.51	-0.06	0.405	3.12	-0.73	0.465	3.58	-0.27
Γ_4	-0.6	-0.542	-3.61	0.39	-0.533	-3.55	0.45	-0.545	-3.63	0.37
Γ_5	0.5	0.472	3.63	-0.22	0.429	3.30	-0.55	0.397	3.31	-0.86
Γ_6	0.6	0.560	4.00	-0.29	0.494	3.53	-0.76	0.612	4.37	0.09
b_1	0.5	0.600	6.67	1.11	0.524	5.82	0.27	0.517	5.74	0.19
b_2	0.8	0.870	5.44	0.44	0.914	6.09	0.76	0.959	5.99	0.99
b_3	1.5	1.714	8.57	1.07	1.644	8.22	0.72	1.717	7.80	0.99
b_4	0.3	0.328	4.69	0.40	0.320	4.57	0.29	0.306	4.37	0.09
b_5	1.8	1.851	16.83	0.46	1.869	16.99	0.63	1.867	16.97	0.61
b_6	0.5	0.510	7.29	0.14	0.534	7.63	0.49	0.502	7.17	0.03
b_7	1.0	1.083	12.03	0.92	1.062	13.28	0.78	1.072	13.40	0.90
λ_1	0.7	0.647	9.24	-0.76	0.669	9.56	-0.44	0.675	9.64	-0.36
λ_2	0.5	0.434	8.68	-1.32	0.439	8.78	-1.22	0.432	8.64	-1.36
λ_3	1.0	0.964	16.07	-0.60	0.973	16.22	-0.45	0.990	16.50	-0.17
λ_4	0.7	0.641	10.68	-0.98	0.648	10.80	-0.87	0.613	10.22	-1.45
λ_5	0.5	0.434	10.85	-1.65	0.457	11.43	-1.08	0.458	11.45	-1.05
λ_6	0.8	0.808	16.16	0.16	0.778	15.56	-0.44	0.779	15.58	-0.42
λ_7	1.0	0.975	19.50	-0.50	0.972	24.30	-0.70	0.971	24.28	-0.73
λ_8	1.0	0.963	24.08	-0.93	0.966	24.15	-0.85	0.984	24.6	-0.40

Table 5.3: Bayesian point estimates, N=500

δ		25 Halton repetitions			100 Halton repetitions			250 Halton repetitions		
		est	t-stat	t-target	est	t-stat	t-target	est	t-stat	t-target
ASC_2	-0.6	-0.812	-1.98	-0.52	-0.853	-1.94	-0.58	-0.831	-1.89	-0.53
ASC_3	0.8	1.246	2.31	0.83	0.954	1.59	0.26	0.970	1.70	0.3
β_1	-1.2	-1.188	-3.96	0.04	-1.291	-3.91	-0.28	-1.273	-3.86	-0.22
β_2	-0.8	-1.049	-3.50	-0.83	-1.029	-3.32	-0.74	-1.024	-3.3	-0.72
Γ_1	-0.4	-0.625	-1.89	-0.68	-0.596	-1.66	-0.54	-0.491	-1.44	-0.27
Γ_2	0.5	0.329	1.32	-0.68	0.426	1.58	-0.27	0.428	1.48	-0.25
Γ_3	0.5	0.255	0.98	-0.94	0.490	1.53	-0.03	0.495	1.55	-0.02
Γ_4	-0.6	-0.401	-1.11	0.55	-0.527	-1.35	0.19	-0.584	-1.46	0.04
Γ_5	0.5	0.31	1.29	-0.79	0.460	1.59	-0.14	0.450	1.55	-0.17
Γ_6	0.6	0.396	1.24	-0.64	0.726	1.96	0.34	0.687	1.91	0.24
b_1	0.5	0.521	3.26	0.13	0.527	3.10	0.16	0.532	2.96	0.18
b_2	0.8	1.04	3.85	0.89	1.139	2.42	0.72	0.970	2.94	0.52
b_3	1.5	1.804	5.15	0.87	2.091	2.95	0.83	1.784	3.64	0.58
b_4	0.3	0.405	3.12	0.81	0.332	2.55	0.25	0.310	2.38	0.08
b_5	1.8	2.283	8.15	1.73	1.960	9.33	0.76	1.878	8.94	0.37
b_6	0.5	0.513	4.66	0.12	0.527	4.39	0.23	0.490	3.77	-0.08
b_7	1.0	1.076	8.28	0.58	0.974	6.96	-0.19	0.999	6.66	-0.01
λ_1	0.7	0.606	5.05	-0.78	0.620	4.77	-0.62	0.658	5.06	-0.32
λ_2	0.5	0.431	5.39	-0.86	0.434	4.34	-0.66	0.448	4.48	-0.52
λ_3	1.0	0.906	8.24	-0.85	0.943	8.57	-0.52	0.939	7.83	-0.51
λ_4	0.7	0.599	5.99	-1.01	0.581	5.28	-1.08	0.633	5.28	-0.56
λ_5	0.5	0.437	7.28	-1.05	0.487	6.96	-0.19	0.492	7.03	-0.11
λ_6	0.8	0.702	10.03	-1.40	0.768	9.60	-0.4	0.785	9.81	-0.19
λ_7	1.0	0.790	11.29	-3	0.916	11.45	-1.05	0.951	11.89	-0.61
λ_8	1.0	0.920	11.50	-1	0.986	12.33	-0.18	0.989	12.36	-0.14

Table 5.4: Frequentist point estimates, N=150

δ		25 Halton repetitions			100 Halton repetitions			250 Halton repetitions		
		est	t-stat	t-target	est	t-stat	t-target	est	t-stat	t-target
ASC_2	-0.6	-0.519	-2.73	0.43	-0.561	-2.81	0.20	-0.575	-2.88	0.13
ASC_3	0.8	0.782	3.01	-0.07	0.794	3.05	-0.02	0.843	3.24	0.17
β_1	-1.2	-1.121	-8.01	0.56	-1.168	-7.79	0.21	-1.184	-7.89	0.11
β_2	-0.8	-0.726	-5.58	0.57	-0.743	-5.31	0.41	-0.758	-5.41	0.30
Γ_1	-0.4	-0.283	-2.02	0.84	-0.364	-2.43	0.24	-0.400	-2.67	0.00
Γ_2	0.5	0.387	3.23	-0.94	0.448	3.45	-0.40	0.485	3.46	-0.11
Γ_3	0.5	0.447	3.19	-0.38	0.504	3.36	0.03	0.522	3.48	0.15
Γ_4	-0.6	-0.600	-3.53	0.00	-0.710	-3.74	-0.58	-0.724	-3.81	-0.65
Γ_5	0.5	0.410	3.15	-0.69	0.466	3.33	-0.24	0.475	3.39	-0.18
Γ_6	0.6	0.535	3.57	-0.43	0.562	3.51	-0.24	0.552	3.45	-0.30
b_1	0.5	0.535	6.69	0.44	0.503	6.29	0.04	0.490	5.44	-0.11
b_2	0.8	0.987	6.17	1.17	0.858	6.13	0.41	0.815	5.82	0.11
b_3	1.5	2.176	8.70	2.70	1.749	9.72	1.38	1.677	9.32	0.98
b_4	0.3	0.315	5.25	0.25	0.319	5.32	0.32	0.308	4.40	0.11
b_5	1.8	2.074	20.74	2.74	1.853	18.53	0.53	1.822	18.22	0.22
b_6	0.5	0.557	11.14	1.14	0.514	8.57	0.23	0.521	8.68	0.35
b_7	1.0	1.117	15.96	1.67	1.026	12.83	0.33	0.999	12.49	-0.01
λ_1	0.7	0.671	11.18	-0.48	0.715	11.92	0.25	0.724	12.07	0.40
λ_2	0.5	0.364	9.10	-3.40	0.441	11.03	-1.48	0.463	9.26	-0.74
λ_3	1.0	0.880	17.60	-2.40	0.963	16.05	-0.62	0.946	15.77	-0.90
λ_4	0.7	0.519	10.38	-3.62	0.632	12.64	-1.36	0.661	13.22	-0.78
λ_5	0.5	0.428	14.27	-2.40	0.467	11.68	-0.82	0.486	12.15	-0.35
λ_6	0.8	0.726	18.15	-1.85	0.794	19.85	-0.15	0.784	19.60	-0.40
λ_7	1.0	0.857	21.43	-3.58	0.956	23.90	-1.10	0.972	24.30	-0.70
λ_8	1.0	0.881	22.03	-2.98	0.971	24.28	-0.73	0.984	24.60	-0.40

Table 5.5: Frequentist point estimates, N=500

estimates we analyze the Euclidean distance $\mathcal{D}(\hat{\delta}, \delta)$ as in equation (13). In Figure **??** we present the calculated distance $\mathcal{D}(\hat{\delta}_B, \delta_0)$ for the Bayesian estimates (graph on the right), as well as the distance $\mathcal{D}(\hat{\delta}_{ML}, \delta_0)$ for the FISML estimates (graph on the left.)

For both estimation cases, the overall distance between the estimates and true values is smaller when the sample size is larger (reflecting consistency of the estimators.) Note that in the case of classical estimation, the distance $\mathcal{D}(\hat{\delta}_{ML}, \delta_0)$ between the estimate $\hat{\delta}_{ML}$ and the true δ_0 is decreasing with the number of repetitions. For N=500 it is clear that increasing the number of repetitions has a decreasing marginal effect on the accuracy of the estimator. Effectively, for N=500 we can see a dramatic effect on $\mathcal{D}$ when passing from 25 to 100 Halton draws for simulation of the likelihood function, but the gains when using 250 Halton draws are not striking.

For the lower sample size (N=150) it is not clear at what level of repetitions we will achieve a limiting point. (Even though the graph may suggest a constant effect on the accuracy of the parameters, we expect that after a certain number of repetitions the increase in accuracy will be marginal.) In the case of Bayesian estimation, even though the distance $\mathcal{D}(\hat{\delta}_B, \delta_0)$ presents a slight increase for a larger number of repetitions, $\mathcal{D}$ is almost flat. Note that for $\mathrm{N} = 500$, frequentist estimation with 100 Halton draws (which marks a sort of limiting point for important accuracy gains in our experiment) offers an accuracy level that is comparable with the one obtained with Bayesian estimation. With 250 Halton draws, frequentist accuracy outperforms Bayesian results. However, Bayesian accuracy outperforms frequentist results for the smaller sample size, independently of the number of frequentist repetitions being used. In fact, the Bayes estimates for the smaller sample size perform quite well when compared to the larger sample size in terms of the distance $\mathcal{D}(\hat{\delta}_B, \delta_0)$.

Even though both estimation methods recover the true parameters well (with an adequate number of frequentist repetitions), for small samples in the frequentist approach the parameters associated with the latent variables as qualitative attributes in the utility function are not statistically significant. At the 95% confidence level for the frequentist results when N=150, we cannot reject the null hypothesis that the latent variables do not affect choice, which is a result that evidently contradicts the assumptions of the model. Note that for N=150 the Bayesian estimates are statistically significant for every qualitative attribute, at least at the 90% confidence level (in fact for about half of the Γ's the estimates are significant at the 95% level.) The results of both estimation methods allow us to reject the null hypothesis, as the sample size becomes larger (N=500.)

Finally, we can examine efficiency and interval estimation results. In general, the standard errors obtained from averaging the precision of the Metropolis Hastings-within-Gibbs sampling estimates for θ (the taste parameters of the utility function) are always lower than those obtained in the case of FISML. In both cases, the standard errors are relatively constant within different numbers of repetitions for each simulator but show a precision increase for the larger sample size. Because of $\hat{\theta}_B$ having a higher precision than $\hat{\theta}_{ML}$, the resulting asymptotic Bayesian confidence intervals are tighter than those obtained for the frequentist approach. For instance, in Figures 4 and 5 (in Appendix B) we show the point estimates and confidence intervals at the 95% level, focusing on the marginal utilities of the qualitative attributes (Γ).

We observe that the precision of the estimates is higher when the sample size is bigger, as well as that the average Bayesian confidence intervals are tighter than their frequentist counterparts. Because of the latter, when we pass to a 90% confidence level for the smaller sample size, all Bayesian point estimates emerge as statistically significant, which is not necessarily the case for the frequentist estimates (as discussed above.) Note however that the higher precision in the Bayesian estimates of the marginal utilities is somewhat compensated for by a lower precision in the rest of the parameters for N=500. In general, for the larger sample size, the efficiency gains in the Bayesian approach are marginal (representing about 0.2% of reduction in the standard errors.) However, this compensation does not appear in the case of the smaller simple size; and for N=150 the Bayesian standard errors are on average 10% lower.

5.4 Conclusions

In this chapter I have introduced how to use a general hybrid choice model for incorporating relevant qualitative attributes in the context of travel mode choice, which is the most common discrete choice problem in travel behavior analysis.

In Chapter 3 I have studied and applied a full information maximum likelihood solution for the estimation of HCMs. In this chapter I have described how to use this simulator in a virtual travel mode choice situation, but I also have generalized a Bayesian Metropolis Hastings-within-Gibbs sampler that is valid for an MNL kernel and, in the context of repeated sampling, provides a solution that is equivalent to the maximum likelihood estimator.

Using a particular example of virtual travel mode choice, I presented the system of equations that allows one to introduce qualitative attributes, as well as the equations and methodological steps required for implementation of frequestist and Bayesian estimation. Then, using a Monte Carlo experiment I have analyzed the results of the point estimators for both estimation methods. In particular, I analyzed two sample sizes (small and large) as well as the effects of the number of repetitions needed for the simulators.

In theory, when using a repeated sampling process, the Bayesian and frequentist estimators are asymptotically equivalent. The results have shown that even though both procedures give similar results for the larger sample size in terms of accuracy, statistical significance, and efficiency, some specific problems are found for the frequentist approach when the sample size is small. In particular, the frequentist marginal utilities of the qualitative attributes are not statistically significant. This is a problem because based on the resulting t-statistics we cannot reject the null hypothesis that the qualitative attributes have no effect on choice (even though in this virtual choice experiment we know they do.) Additionally, the Bayesian confidence intervals are tighter than the frequentist intervals for the smaller sample size. That the Bayesian estimator outperforms the classical estimator for a reduced sample size is a direct result of the properties of the latter being asymptotic. On the one hand, properties of the maximum likelihood estimator are only valid for large enough samples, because frequentist inference is based on infinite hypothetical repeated experiments. On the other hand, the Bayesian point estimates represent the optimal estimator summarizing the whole posterior distribution. The posterior distribution corresponds to our beliefs about the unknown parameters after the sample realization of the data, and in this sense, Bayesian inference works independently of the size of the sample. Thus, the results are in line with statistical theory.

References

M. Ben-Akiva, 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): 163–175, 2002.

D. Bolduc, M. Ben-Akiva, J. Walker, and A. Michaud. Hybrid choice models with logit kernel: applicability to large scale models. In M. Lee-Gosselin and S. Doherty, editors, *Integrated Land-Use and Transportation Models: Behavioral Foundations*, pages 275–302. Elsevier, New York, 2005.

A. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

C.J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7:473 – 483, 1992.

K. Jöreskog 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, 2008.

A. Nobile. Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics*, 99:335–345, 2000.

S. Scott. Data augmentation for the bayesian analysis of multinomial logit models. Proceedings of the american statistical association section on bayesian statistical science, 2003.

K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, second edition, 2009.

Appendix A: Confidence Intervals

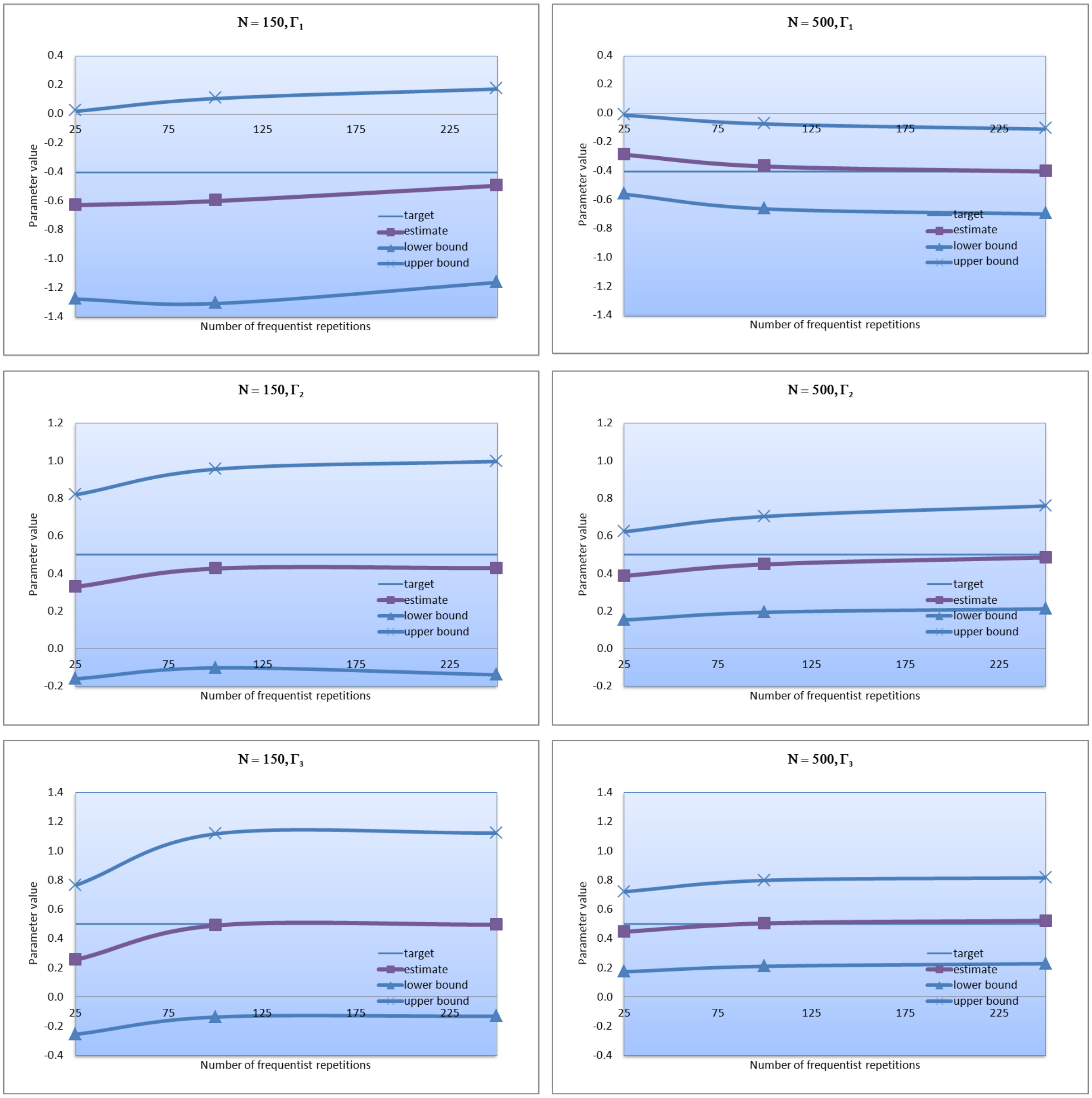

Figure 5.3: Frequentist confidence intervals for Γ_1, Γ_2, and Γ_3

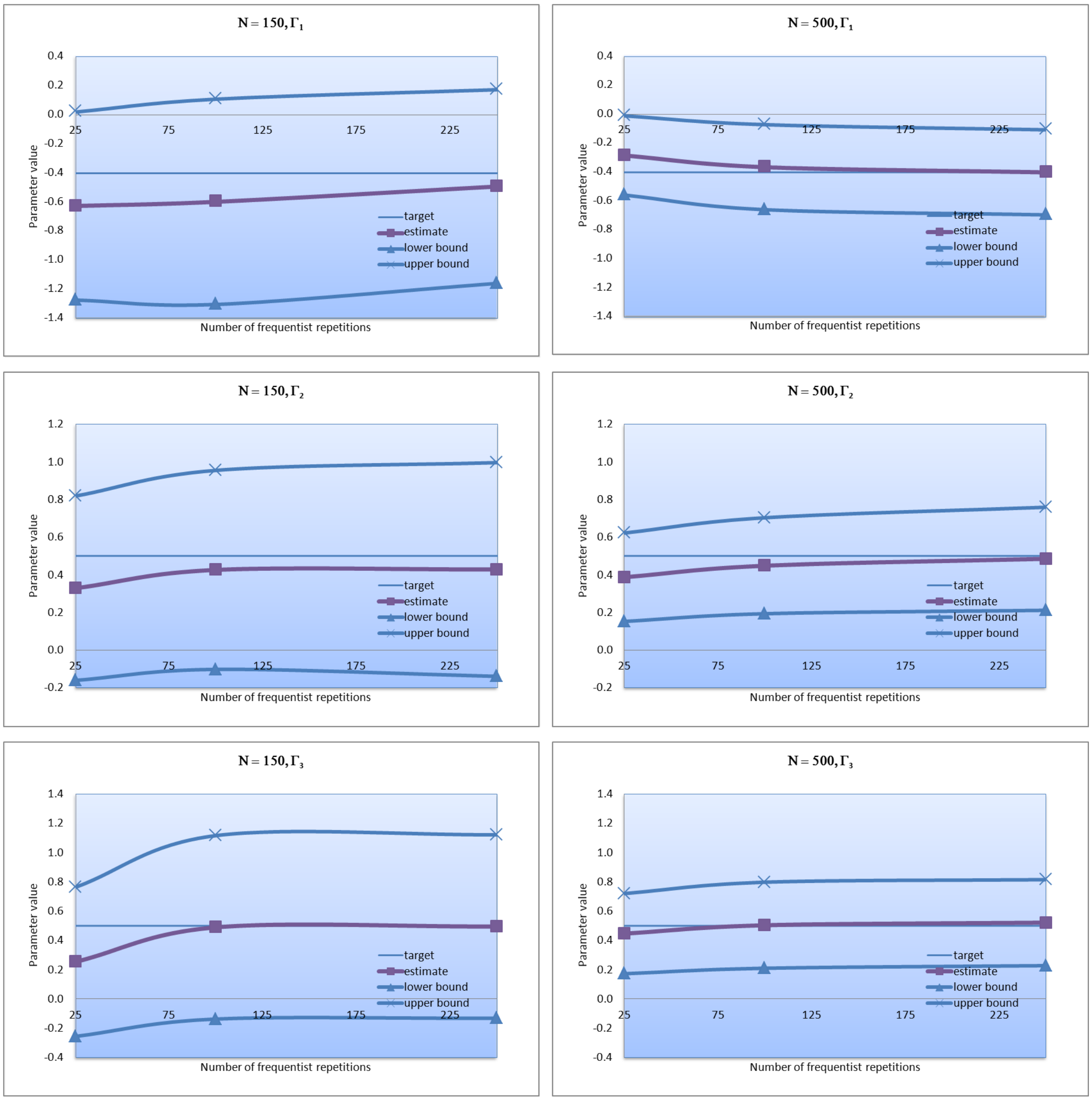

Figure 5.4: Frequentist confidence intervals for Γ_4, Γ_5, and Γ_6

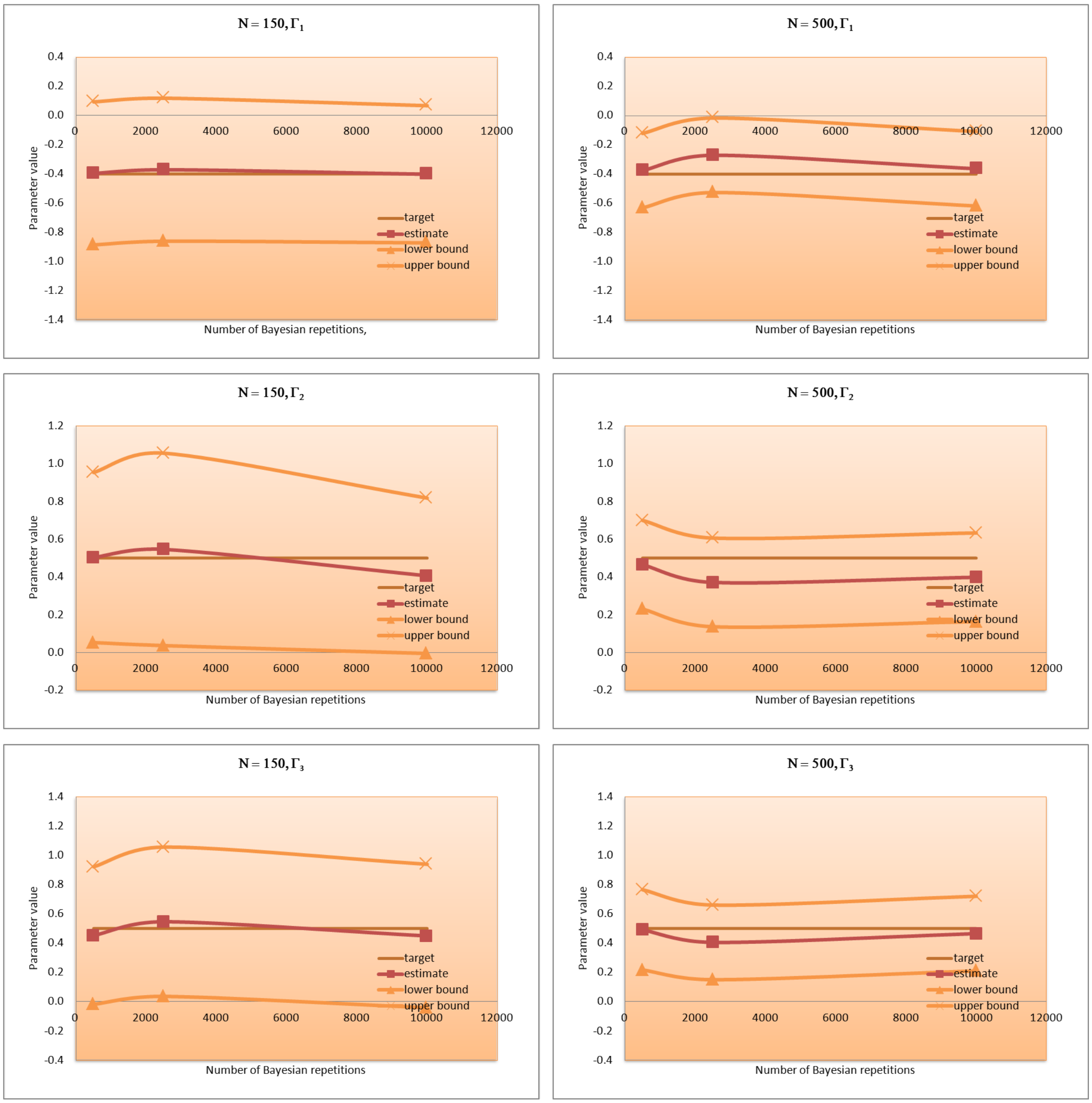

Figure 5.5: Bayesian confidence intervals for Γ_1, Γ_2, and Γ_3

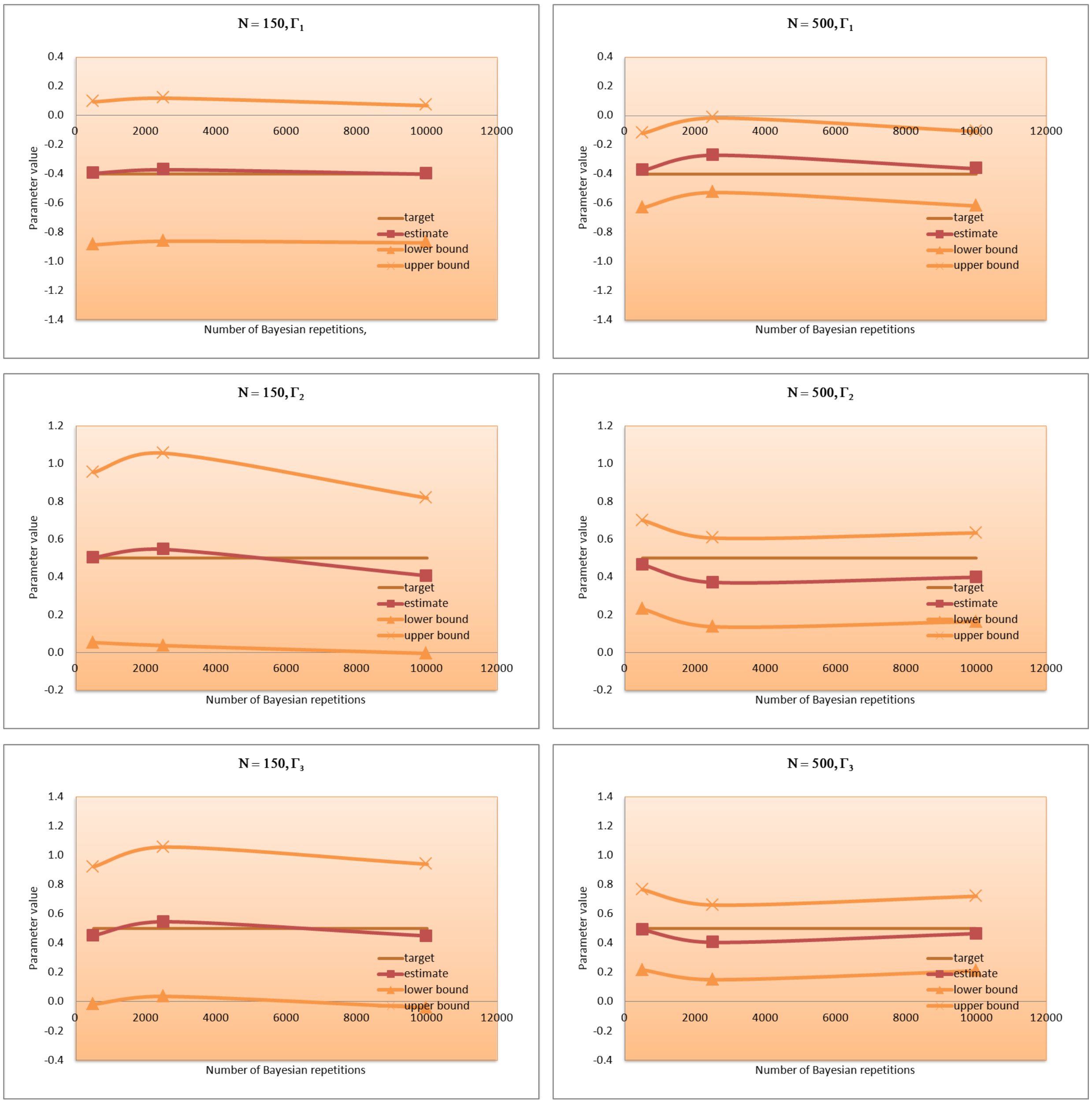

Figure 5.6: Bayesian confidence intervals for Γ_4, Γ_5, and Γ_6

Chapter 6

Consumer behavior toward IP telephony access in Japan

Knowledge and awareness are key elements in the decision to adopt new technologies, especially when these technologies are at an early stage of development or their overall penetration in the market is low. In this chapter, I analyze the role of knowledge, awareness and promptness to adopt new technologies as qualitative attributes in the choice of adopting IP telephony. IP telephony allows telephone calls to be made through an Internet connection. IP telephony requires consolidated broadband access to achieve sufficient quality of service to compete with traditional telephony. Because of the dynamics of the Japanese telecom market, it is particularly interesting to study the adoption of IP telephony in Japan. In this chapter I derive and apply a hybrid choice model with a probit kernel with dichotomous effect indicator variables to analyze consumer behavior toward IP telephony. The model allows one to measure the effect of qualitative attributes related to knowledge as well as the individual attitude toward the adoption of new technologies on the choice probabilities of adopting IP telephony. The main findings of this study suggest that consumers desire a quality of service of IP telephony that will assure some features that in best effort IP telephony are not guaranteed. In particular, according to the forecasts of the model, in order to increase the penetration of IP telephony it is essential to provide access to emergency calls. Additionally, the hybrid choice model yields the profile of those users who can be labeled as early adopters.

6.1 IP telephony and the Japanese telecom market

6.1.1 IP telephony

IP telephony (Internet Protocol telephony) or VoIP (voice over Internet Protocol) is a relatively new technology that allows telephone calls to be made through an Internet connection rather than through the public switched telephone network (PSTN). VoIP can be accessed through traditional telecom service providers, as well as through web telephony services such as Skype. In VoIP, voice data as well as other signals such as facsimile are digitally encoded as packets of data and sent via the Internet (Figure **??**) at costs far below normal long-distance telephone charges. Whereas PSTN carries data as a single packet over a dedicated circuit-switch connection, VoIP shares information over separate paths through a packet-switched Internetwork following a TCP/IP model. A VoIP phone-to-phone call typically requires converting the voice data from analog to digital, and then compressing and dividing the digital signal for transmission over the Internet after receiving the signal at an IP gateway; the process is reversed at the receiving end. To avoid packet loss and for a reliable and high-quality service a sufficient bandwidth is essential.

Figure 6.1: VoIP diagram

Note that VoIP is an emerging market, even though **?** argue that "unlike some services that have emerged out of the electronic revolution, VoIP does not involve a new good per se, but rather a new way of providing an existing good at possibly lower cost and in a possibly more convenient manner." However, the development of services such as mobile VoIP (boosted by the emergence of smartphones), VoIP micro-blogging, and videoconferencing may open new markets for VoIP, changing the nature of personal telecommunications. But VoIP market penetration faces some important challenges. In particular, since VoIP packets are considerably latency-sensitive, quality of service (QoS) has been criticized. QoS is related to the ability to guarantee in advance a certain level of performance for data transmission in terms of avoiding packet delay variation (jitter) and packet loss through latency and jitter bounds. Background noise, interrupts, call echo, robotic sound, general delays, and even dropped calls are typical problems of best effort VoIP, for which QoS depends on traffic load (i.e. the network capacity is insufficient.) Whereas quality of experience (QoE) of best effort VoIP is comparable to a cell phone call (or worse under high load conditions), customers of QoS guaranteed VoIP cannot distinguish from Plain Old Telephone Service (POTS). Even though conventional IP routers provide best effort service, taking advantage of current broadband characteristics modern IP routers are able to provide guaranteed QoS to specific data flows, including VoIP. In sum, VoIP can be perceived by consumers as a real alternative to the traditional telephone service only when QoS is guaranteed, and modern broadband networks are focusing on satisfying the requirements of QoS by maximizing bandwidth and minimizing delays on voice data. Since VoIP is becoming more and more popular, it is interesting to understand the consumers' response to VoIP services. In this chapter, I analyze choice of telephony access for Japanese consumers.

6.1.2 Overview of the Japanese telecom industry

In 1985 the Japanese telecommunications industry was deregulated[1] and the Nippon Telegraph and Telephone (NTT) corporation, a publicly traded company regulated by the Law Concerning Nippon Telegraph and Telephone Corporation, replaced the government monopoly. Since then, the incumbent NTT has faced a dynamic process of change to competition, including the introduction of new carriers. However, NTT still dominates the telecom market in Japan. In fact, acting as a policy agency Japan's Ministry of

[1] After World War II, the Japanese government created the Nippon Telegraph and Telephone Public Corporation as a government monopoly. This decision was taken arguing a reconstruction strategy.

Internal Affairs and Communications (MIC)[2] has advocated competition in the Japanese telecom market through reduction of NTT's oligopolistic market power. In 1999 NTT was established as a holding company with three subsidiary telecom companies (NTT East, NTT West, and NTT Communications.)

To exemplify how dynamic the Japanese telecom industry is with regard to the adoption of new technologies we can look at some figures describing the mobile telephone market. First, Figure **??** presents the total number of Japanese mobile telephone subscribers from 1998 to 2004. In this period the total number of subscribers more than doubled, with annual increases in the order of 20% by the late 1990s. Japan was the first country to introduce both Internet access (i-Mode or 2.5G made broadly available in 1999) and third generation 3G[3] (2001) for mobile telephones. Accessing the Internet using a cell phone became a desirable feature for Japanese consumers who rapidly subscribed to a service providing a mobile phone with Internet access.

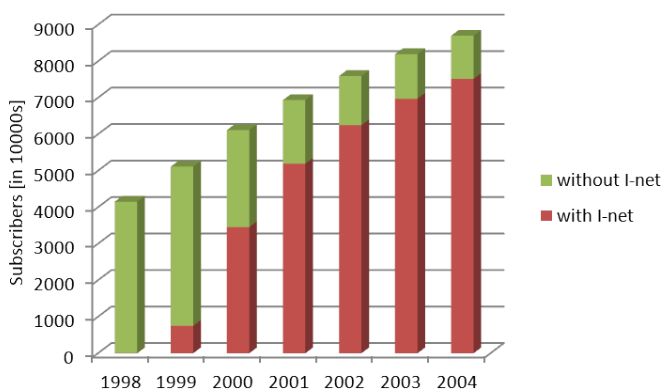

Figure 6.2: Mobile telephone subscribers in Japan. Source: MIC

Figure **??** shows the increase in the share of mobile phones with Internet access. In effect, in 5 years the penetration of mobile phones with Internet access achieved 85% of the market (see Figure **??**). In fact, as of 2006, the penetration of 3G represented more than 55% (51 million subscribers).

Mobile telephony is basically dominated by three companies: NTT DoCoMo, KDDI, and SoftBank (Figure **??**.)

[2]Japan's Ministry of Posts and Telecommunications (MPT), before 2001; Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT), from 2001 to 2004.

[3]First generation (1G) was first provided in 1979. Second generation (2G) mobile telephone services began in 1993.

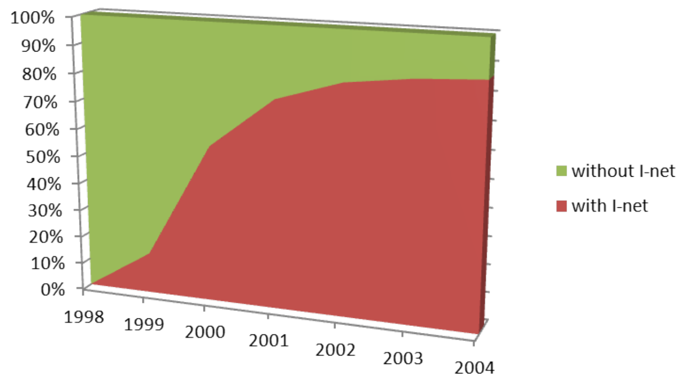

Figure 6.3: Penetration rate of mobile phones with Internet access. Source: MIC

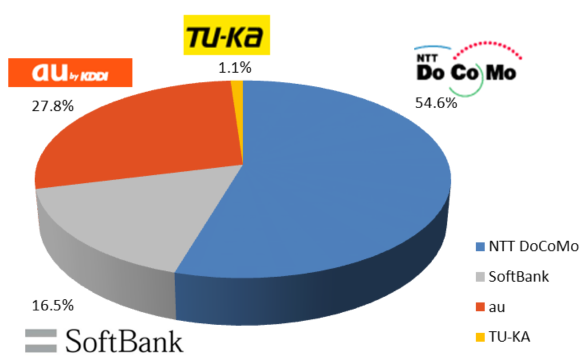

Figure 6.4: Market shares of mobile telephony providers. Source: NTT DoCoMo (Feb 2007)

Following a timid beginning in the late 1990s, after the e-Japan strategy was implemented in 2000 and during the first years of the 2000s the broadband (BB) penetration in the Japanese market experienced a remarkable increase (see Figure ??.) The huge development of BB services, which were accompanied by low prices[4] and very high quality in terms of speed, offered a perfect scenario for both an increasing demand for BB Internet access and the consequent deep penetration of IP telephony in Japan.

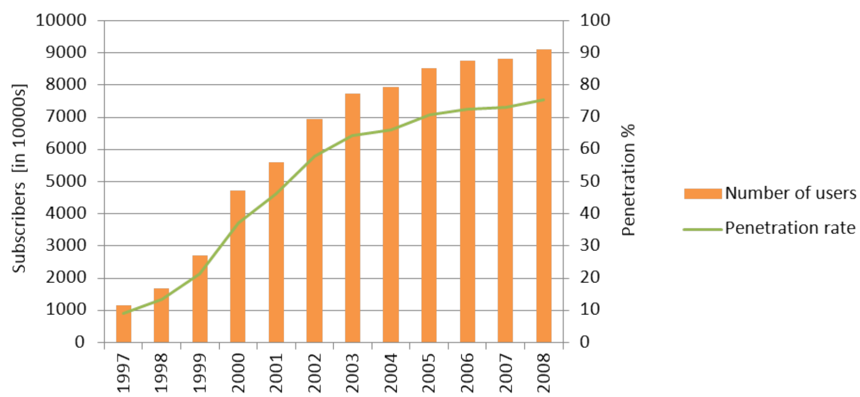

Figure 6.5: Internet access. Source: MIC

The dominating BB networks in Japan are asymmetric digital subscriber line (ADSL), cable television (CATV), and, more recently, fiber to home (FTTH)[5] (See Figure ??.) The main providers of BB services are NTT (through NTT East and West), KDDI, and Softbank (Figure ??.)

In Japan, IP telephony is regulated by the MIC. Figure ?? shows the expansion of VoIP in recent years. Calls between VoIP subscribers of a same provider or group are usually free of charge; calls from VoIP to POTS or personal handy-phone system (PHS[6]) are charged uniformly all over the country according to a fixed rate. When first introduced, IP phones did not carry a phone number and VoIP served only to make calls. Phone numbers were assigned to IP phones in 2002, and in 2005 local number portability was made available for FTTH subscribers. In fact, FTTH subscribers can switch from POTS to IP Telephony keeping the same phone number. To be more precise, currently in Japan both best-effort and QoS guaranteed VoIP are available. Best-effort VoIP, which is also known in Japan as 050-type, is mostly used by ADSL subscribers and has an associated

[4]Charges per bps are the lowest in the world.

[5]Typical maximum speed for these BB services in Japan are 50 Mbps (ADLS), 30 Mbps (CATV), and 100 Mbps (FTTH).

[6]Voice and data are sent through a narrow band service for a fixed rate.

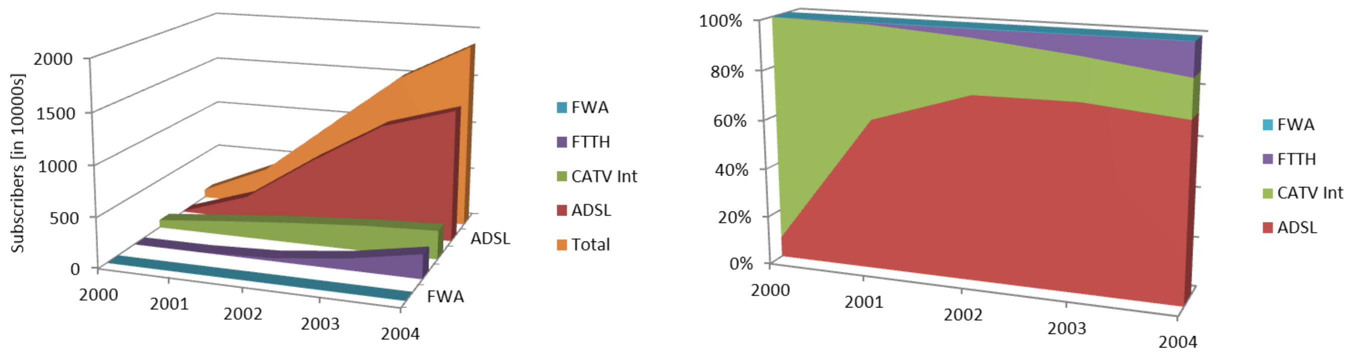

Figure 6.6: Evolution of broadband services in Japan: number of subscribers (left) and penetration rate (right). Source: MIC

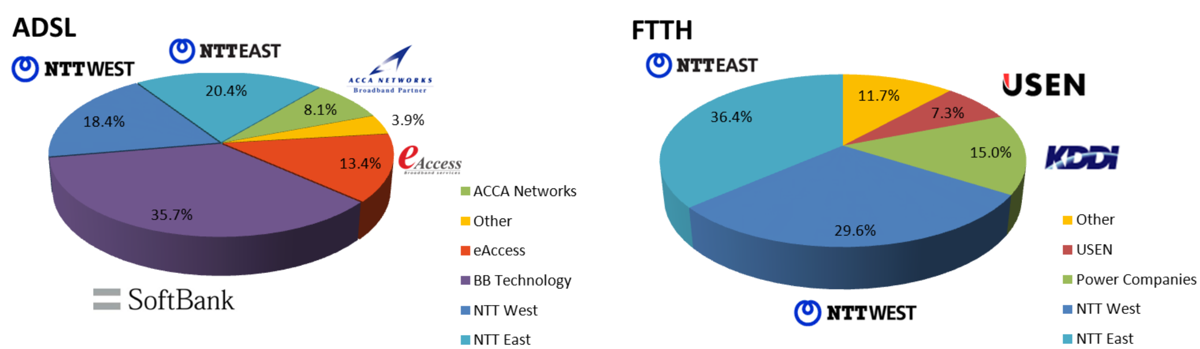

Figure 6.7: Market shares of ADSL and FTTH services. Source: MIC (Second quarter 2006)

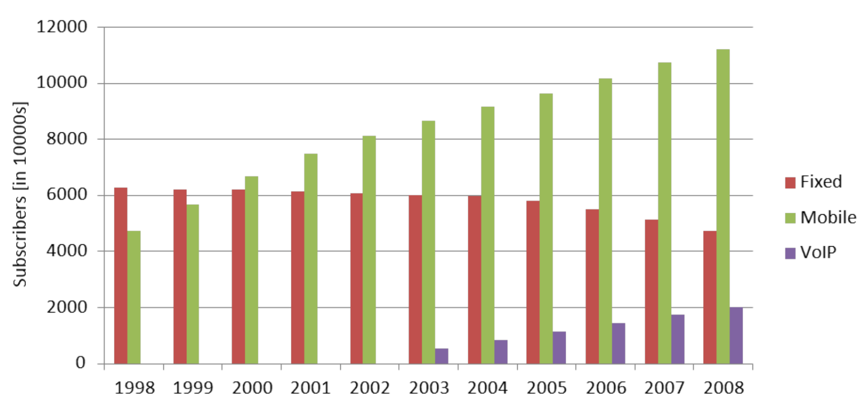

Figure 6.8: Evolution of the telephony market. Source: MIC

050 prefix for direct-dial calls from fixed telephones to VoIP terminals. QoS guaranteed IP telephony (0ABJ-type) is mostly used by FTTH subscribers and its voice quality is equivalent to normal telephony. 0ABJ-VoIP has a related location code that allows the provider to use the same numbering format used for fixed telephones (0AB-J); other features of 0ABJ-VoIP include the possibility of making emergency calls and fax usage. Ida et al. (2008) conclude that, in a market that is predominantly dominated by best-effort VoIP, IP telephnoy is perceived more as an add-on option of BB Internet access rather than as a perfect substitute for POTS. Effectively, most of IP telephony users show a parallel use of VoIP and POTS. Figure **??** shows the market shares for the main providers of VoIP.

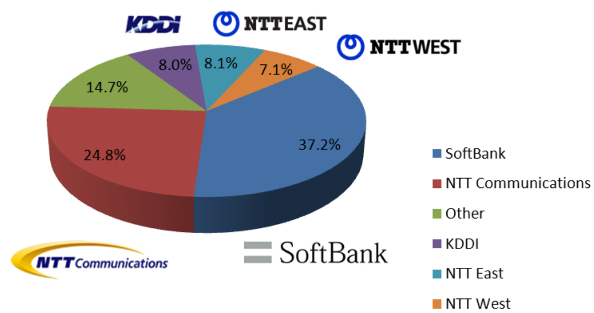

Figure 6.9: Market shares of VoIP providers. Source: MIC (Second quarter 2006)

6.1.3 Discrete choice models for telecom access demand

The discrete choice modeling framework is perfectly suited for the analysis of the demand for telecommunication access. Typically, different providers offer various options that the final consumer can choose from. For example, when searching for a new cell-phone we gather information about the different plans offered by each company. For a monthly charge, each plan offers a certain amount of voice minutes; a certain amount of data, including text messages and mobile browsing; long distance, roaming, and out-of-plan charges, as well as other features such as caller ID and call forwarding. Each plan-company pair presents an option or alternative that summarizes the associated cost and features, and final consumers choose the alternative that is most convenient for their needs. The same choice process applies to other telecom choice situations such as broadband choice and decisions regarding home telephony services. Discrete choice models (DCMs) consider choice to be a result of a utility maximizing behavior, where the final consumer evaluates the available alternatives and makes a decision according to individual tastes.

Even though there has been some interesting research on the application of DCMs for telecom disaggregated demand, this particular field has not achieved the same level of development as some other areas, such as travel behavior, where DCMs are the dominant tool for analyzing consumer response. In telecommunications, there has been a lot of analysis in terms of industrial organization concepts. Regarding the demand side, the typical econometric modeling is based on panel data analysis[7].

The earliest contributions in the application of DCMs to telecom access demand is the work of **???**. With U.S. data and using both logit and probit models (as well as linear probability models), Perl studied willingness to pay and access elasticities at the household level for residential phone services. In 1987, **?** used a nested logit to study aggregate price elasticities for different patterns of residential phone service demand. In Canada, one of the earliest applications of binary telephone subscription demand models is **?**. **?** uses the DCM framework to analyze policy implications of extended area service developments in the U.S. In **?**, the analysis is extended to include the effect of the local service rate on residential demand for fixed phones. **?** apply DCMs to the study of the relationship between consumption and self-selecting tariffs in the context of residential demand for local telephone services. **?** and **?** analyze long distance carrier choice in the United States and in Japan, respectively.

Following the trends of the telecom industry, more recent work has focused on mobile telephony access and Internet service choice, also incorporating recent advances in discrete choice modeling. For instance, **?** use a mixed logit model with Japanese data to analyze heterogenous substitution patterns and elasticity of demand of second (2G) and third (3G) generation mobile phone services. The same authors have also studied the market of broadband Internet in Japan (**?**). Using multinomial and nested logit models, the authors analyze choice between narrow and broadband services. They conclude that the ADSL market is independent of other BB services such as CATV and FTTH, and that within the ADSL submarkets, both low-speed and high-speed ADSL are highly elastic and compete with narrow brand (dial up) and BB, respectively. In the context of Internet service choice, **?** introduced discrete choice analysis with stated preference (SP) data. Using Australian consumers' revealed responses, **?** estimated a DCM of

[7] For a comprehensive review of telecommunication demand forecasting, see **?**. Note that this paper reviews only one application of DCMs to Internet access, and no model of IP telephony.

subscriber churn for Internet service providers. Other relevant work includes **???????**, and **?**. Of particular interest is the work of **?** who study consumers' Internet awareness, access, and use, as well as the willingness to pay for particular attributes by different clusters of consumers beyond the trade-off between subscription price and access speed. For instance, reliability of the service is, on average, the most valued attribute, followed by (with corresponding decreasing WTPs) service speed, always-on connectivity, and installation delays.

Again for the Japanese BB market, **?** focused on the explosive demand for FTTH. Using a mixed logit model with SP data the authors analyze the willingness to pay for public services over FTTH for urban and provincial areas (an important issue associated with the digital divide debate.) Expanding on the analysis of **?** and using a mixed logit with RP data, **?** study consumer shift within BB services (basically from ADSL to FTTH) and show that the migration is determined by variables such as income and service usage (motion-picture viewing and IP telephony.)

Finally, for the specific case of access demand for IP telephony, **?** provide an early discussion on the effects of substituting POTS by an Internet-based phone service in terms of the detriment to the profitable long distance market of telecom companies. **?** apply contingent valuation techniques to analyze demand for IP telephony. Using stated WTPs for VoIP services in the US, the authors conclude that the demand is elastic (over the range of current prices) and ascertain that the overall market for best-practice VoIP in the US is rather small. Using Japanese stated preference data gathered in 2005, **?** estimate an MNL for choice between fixed phone, IP phone, and parallel fixed and IP service. Each of these three alternatives were described in terms of fixed monthly charges, voice quality, number portability, emergency calling availability, FAX usage availability, and call charges. Based on the MNL estimates, a simulation of future scenarios was carried out to study the penetration of IP telephony. The authors conclude that the two most important features consumers are willing to pay for are access to emergency calls and a better voice quality. In fact, the overall quality of service is fundamental for the consolidation of VoIP in the market.

In this chapter I build on the model of telephony choice developed by **?** to explain IP telephony demand, which in fact is part of a more comprehensive model of Internet demand including I-net access and IP telephony choice. The original model of **?** considers a hybrid choice formulation (**?**) with a mixed logit kernel. Estimation of the model was performed using full information simulated maximum likelihood. In my model, I consider an MCMC-based Bayesian estimator for a hybrid choice model with a multivariate probit kernel.

6.2 A hybrid choice model with a probit kernel and dichotomous indicators

To model IP telephony choice in Japan, I consider a hybrid choice model with a probit kernel and dichotomous indicators. Hybrid choice models are a generalization of discrete choice models where endogenous latent variables enter the utility function of standard discrete choice as explanatory variables. The econometric representation of hybrid choice models consists of a simultaneous system of structural equation models. In the next subsection I present the structural model that constitutes the system of equations.

6.2.1 Structural model

Consider the following system

Structural equations

$$\underset{(L\times 1)}{z_n^*} = \underset{(L\times L)}{\Pi} z_n^* + \underset{(L\times M)}{B}\underset{(M\times 1)}{w_n} + \underset{(L\times 1)}{\zeta_n}, \zeta_n \sim MVN(0, H_\Psi^{-1}) \tag{6.1}$$

$$\underset{(J\times 1)}{U_{tn}^*} = \underset{(J\times K)}{X_{tn}}\underset{(K\times 1)}{\beta} + \underset{(J\times Q)}{W_{tn}^*(X_{tn}, z_n^*)}\underset{(Q\times 1)}{\varrho} + \underset{(J\times L)}{\Gamma}\underset{(L\times 1)}{z_n^*} + \underset{(J\times 1)}{\nu_{tn}}, \nu_{tn} \sim MVN(0, H_\Sigma^{-1}) \tag{6.2}$$

$$\underset{(R\times 1)}{I_n^*} = \underset{(R\times 1)}{\alpha} + \underset{(R\times L)}{\Lambda}\underset{(L\times 1)}{z_n^*} + \underset{(R\times 1)}{\varepsilon_n}, \varepsilon_n \sim MVN(0, H_\Theta^{-1}) \tag{6.3}$$

Measurement equations

$$\underset{(1\times 1)}{I_{rn}} = \mathbf{1}_{[I_{rn}^*>0]}, \forall r, n \tag{6.4}$$

$$\underset{(1\times 1)}{y_{tn}} = i \in C_n \text{ iff } U_{itn} - U_{jtn} \geq 0, \forall j \in C_n, j \neq i, \forall n \in N. \tag{6.5}$$

where z_n^* is an endogenous random vector of latent variables that enters the utility function as a latent explanatory variable; the matrix Π allows for 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 vector of explanatory variables affecting the latent variables; B is a matrix of K unknown regression coefficients used to describe the global effect of $(I_L - \Pi)^{-1}Bw_n$ on the latent variables; and H_{Ψ}^{-1} is a covariance matrix which describes the relationship among the latent variables through the error term. To simplify notation, we define $\tilde{B} = (I_L - \Pi)^{-1}B$, $\tilde{\zeta}_n = (I_L - \Pi)^{-1}\zeta_n$, and $H_{\tilde{\Psi}}^{-1} = [(I_L - \Pi)^{-1}]H_{\Psi}^{-1}[(I_L - \Pi)^{-1}]'$.

The choice model in equation (**??**) is written in vector form where we assume that there is a total of J_n available alternatives in the set C_n. Hence, U_n is a vector of indirect utility functions; X_n is a matrix with X_{in} designating its ith row; and β is a vector of unknown parameters. $W_n^*(X_n, z_n^*)$ is a matrix of Q interactions between the observable X_n and the latent z_n^* as well as interactions within the latent variables; ϱ is a vector of unknown parameters associated with these interactions. Γ is a matrix of unknown parameters associated with the latent variables present in the utility function, with Γ_i designating the ith row of matrix Γ. The analytical form of the discrete choice kernel depends on the assumptions regarding the distribution of the random term ν_n.

In the set of measurement equations, I_n corresponds to a vector of manifest variables that serve as indicator responses for the latent variables z_n^*; α is an intercept vector and Λ is a matrix of G unknown factor loadings. The term ε_n is a vector of error terms with covariance matrix H_{Θ}^{-1}. Finally, we stack the choice indicators y_{in}'s into a vector called y_n.

The structural model translates into a parametric hybrid choice model dominated by the Lebesgue measure

$$(\mathcal{Y}, \mathcal{P} = P_\delta = \ell(y, I; \delta), \delta \in \Delta \subseteq \mathbb{R}^p, p \geq 1) \tag{6.6}$$

where $\mathcal{Y} = C_n \times \{0,1\}^R$ is the sample space of a hybrid choice sampling model composed by both choice y among the alternatives in the set C_n and the group of R dichotomous indicators stacked in I; $\mathcal{P}$ is a parameterized family of probability density functions P_δ on $\mathcal{Y}$, $\ell(y;\theta)$ is the likelihood function as in equation **??**, δ is a vector of p parameters, and Δ is the parameter space.

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:

$$\ell(y, I; \delta) = \prod_{n=1}^{N} \prod_{i \in C_n} \left(\int_{z_n^*} P_n(i \,|z_n^*, X_n, \theta) f(I_n | z_n^*, \Lambda, \Theta) g(z_n^* \,| w_n, B, \Pi, \Psi) dz_n^* \right)^{y_{in}}, \tag{6.7}$$

where

$$f(I_n) = \prod_{r=1}^{R} \Phi\left(\frac{\alpha_r + \Lambda_r z_n^*}{\theta_r}\right)^{I_{rn}} \left(1 - \Phi\left(\frac{\alpha_r + \Lambda_r z_n^*}{\theta_r}\right)\right)^{(1-I_{rn})},$$

with Φ being the cumulative distribution function (cdf) of a standard normal distribution; and where $g(z_n^*|w_n, B, \Pi, \Psi)$ corresponds to the multivariate normal distribution $MVN((\mathrm{I}_L - \Pi)^{-1}Bw_n, [(\mathrm{I}_L - \Pi)^{-1}]\Psi[(\mathrm{I}_L - \Pi)^{-1}]')$.

6.2.2 Reduced form

The system of structural equations **??**, **??**, and **??** can be written in the reduced form **??**.

$$\begin{bmatrix} z_n^* \\ I_n^* \\ U* \end{bmatrix} = \begin{bmatrix} 0 & \tilde{B} & 0 & 0 \\ \alpha & \Lambda\tilde{B} & 0 & 0 \\ 0 & \Gamma\tilde{B} & \beta & \varrho \end{bmatrix} \mathbf{X}_n\,(1, w_n, X_n, W_n^*(X_n, z_n^*)) + \begin{bmatrix} 1 & 0 & 0 \\ \Lambda & 1 & 0 \\ \Gamma & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\zeta}_n \\ \varepsilon_n \\ \nu_n \end{bmatrix} \tag{6.8}$$

Taking advantage of the fact that each error term is assumed to be normal distributed, the reduced form of an HCM with a probit kernel follows the following multivariate distribution:

$$\begin{bmatrix} z_n^* \\ I_n^* \\ U_n^* \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mu_{z_n^*} \\ \mu_{I_n^*} \\ \mu_{U_n^*} \end{bmatrix}, \begin{bmatrix} \tilde{\Psi} & \tilde{\Psi}\Lambda' & \tilde{\Psi}\Gamma' \\ \Lambda\tilde{\Psi} & \Lambda\tilde{\Psi}\Lambda' + \Theta & \Lambda\tilde{\Psi}\Gamma' \\ \Gamma\tilde{\Psi} & \Gamma\tilde{\Psi}\Lambda' & \Gamma\tilde{\Psi}\Gamma' + \Sigma \end{bmatrix}\right), \tag{6.9}$$

where

$$\begin{aligned} \mu_{z_n^*} &= \tilde{B}w_n \\ \mu_{I_n^*} &= \alpha + \Lambda\tilde{B}w_n \\ \mu_{U_n^*} &= X_n\beta + \Gamma\tilde{B}w_n + W_n^*(X_n, z_n^*)\varrho. \end{aligned}$$

It is possible to show that

$$\pi(z_n^*|I_n^*) \sim \mathrm{MVN}\left(\mathbb{E}(z_n^*|I_n^*), \mathbb{V}(z_n^*|I_n^*)\right) \tag{6.10}$$

$$\pi(U_n^*|I_n^*) \sim \mathrm{MVN}\left(\mathbb{E}(U_n^*|I_n^*), \mathbb{V}(U_n^*|I_n^*)\right), \tag{6.11}$$

where

$$\mathbb{E}(z_n^*|I_n^*) = \tilde{B}w_n + \Psi\Lambda'\left(\Lambda\Psi\Lambda' + \Theta\right)^{-1}\left(I_n^* - (\alpha + \Lambda Bw_n)\right) \tag{6.12}$$

$$\begin{aligned} \mathbb{E}(U_n^*|I_n^*) &= X_n\beta + \Gamma\tilde{B}w_n + W_n^*(X_n, z_n^*)\varrho + \Gamma\Psi\Lambda'\left(\Lambda\Psi\Lambda' + \Theta\right)^{-1}\left(I_n^* - (\alpha + \Lambda Bw_n)\right) \\ &= X_n\beta + \Gamma\tilde{B}w_n + W_n^*(X_n, z_n^*)\varrho + \Gamma\mathbb{E}(z_n^*|I_n^*), \end{aligned} \tag{6.13}$$

and

$$\mathbb{V}(z_n^*|I_n^*) = \Psi - \Psi\Lambda'\left(\Lambda\Psi\Lambda' + \Theta\right)^{-1}\Lambda\Psi \tag{6.14}$$

$$\mathbb{V}(U_n^*|I_n^*) = \Gamma\Psi\Gamma' + \Sigma - \Gamma\Psi\Lambda'\left(\Lambda\Psi\Lambda' + \Theta\right)^{-1}\Lambda\Psi\Gamma'. \tag{6.15}$$

6.2.3 MCMC estimator

Consider the following partition of the parameter space Δ: the taste parameters of the utility function $\theta = (\beta, \varrho, \Gamma)'$, the parameters associated with the covariance structure implied by the precision matrix H_Σ^{-1}, the parameters of the structural equation B, and Λ which contains the measurement equation parameters. This partition allows us to implement the Gibbs sampler developed in Chapter 5. Note that equations **??** and **??** provide the conditional distributions that are necessary for implementing Gibbs sampling with data augmentation.

6.3 The Data

I use data from a survey conducted in 2004 by NTT (Nippon Telegraph and Telephone Corporation) Labs in Japan. The market research survey gathered information about stated behavioral intentions of accessing the Internet and IP Telephony under hypothetical future scenarios. Information on current usage of both the Internet and telephone service was also collected. The sample consists of 3369 Japanese consumers living in different metropolitan areas. Figure **??** shows how the sample is composed by gender, and also how many current VoIP users are represented.

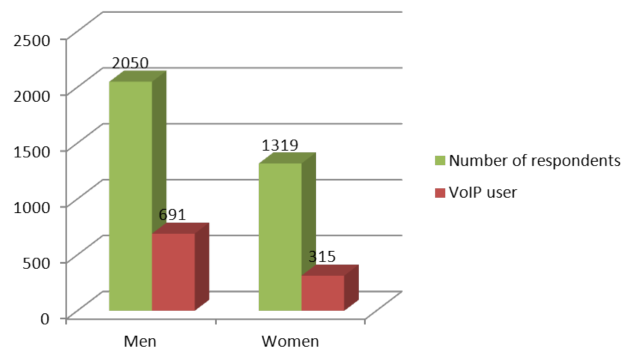

Figure 6.10: Total number of respondents and VoIP users by gender

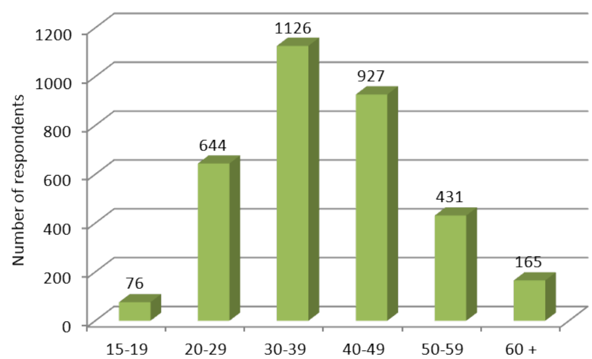

Figure 6.11: Total number of respondents by age range

In terms of Internet access, most of the individuals are subscribed to BB services[8] (Figure **??**). ADSL dominates with 45.83%, followed by CATV with 24.55%, and FTTH with 13%. Access to BB is important since it is a requisite for VoIP; current BB subscribers

[8]The BB penetration rate of the sample (83%) is higher than the actual penetration rate in 2004 (66%).

that are not currently subscribed to VoIP represent an easily approachable potential market. Narrowband access is represented by only 10.00% in the sample. 6.62% of the sample does not have Internet access.

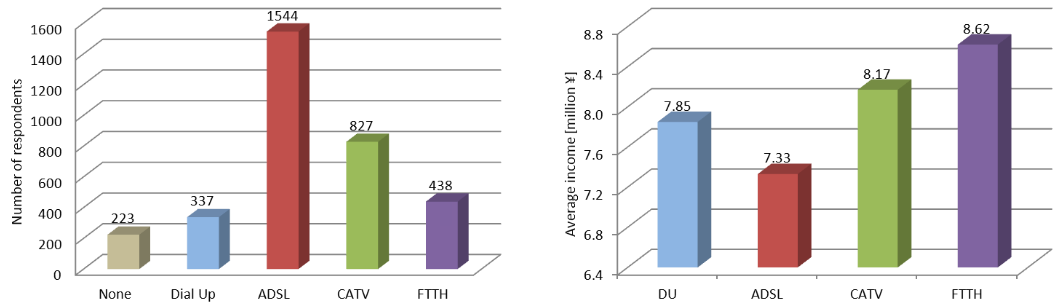

Figure 6.12: Number of respondents and average income by Internet service

Because each of the 3369 individuals that were interviewed responded to 5 different choice situations, there is a total of 16845 pseudo-individuals for estimation. The stated preference experiment considered choice over 3 alternatives: plain old telephone service (POTS), combination of POTS and IP telephony (POTS/IP or POIP), and VoIP (IP). In Table **??** the experimental attributes are shown.

Experimental attribute	Mean	Stdv	Min	Max
POTS monthly usage charge [¥]	1.40	0.33	1.00	1.80
POTS charge for 3 minutes of local conversation [¥]	6.99	0.82	6.00	8.00
POTS charge for 3 minutes of long distance conversation [¥]	31.09	8.10	20.00	40.00
POTS charge for calling a mobile phone [¥]	89.71	20.99	60.00	120.00
POTS/IP charge for 3 minutes of local conversation [¥]	6.01	0.82	5.00	7.00
POTS/IP charge for 3 minutes of long distance conversation [¥]	6.01	0.82	5.00	7.00
POTS/IP charge for calling a mobile phone [¥]	67.51	28.07	30.00	120.00
IP initial cost of service [¥]	0.27	0.21	0.00	0.50
IP monthly usage charge [¥]	0.27	0.20	0.00	0.50
IP charge for 3 minutes of local conversation [¥]	6.01	0.82	5.00	7.00
IP charge for 3 minutes of long distance conversation [¥]	6.01	0.82	5.00	7.00
IP charge for calling a mobile phone [¥]	22.40	23.62	0.00	60.00

Table 6.1: Descriptive statistics of experimental attributes

6.4 Modeling the adoption of VoIP

I will now present the results of the HCM Bayesian estimation process for the telephony choice data. Using the R language, I implemented a specific case of the Gibbs sampler for a hybrid choice model with a probit kernel with dichotomous indicators and individual-specific latent variables. The probit kernel is more general than an MNL, since a general structure of covariance can be tested. When dichotomous effect indicators are introduced, mathematically the problem involves additional latent variables that are not necessary in the continuous case[9]. In addition, the telephony choice model I implement takes into account the repeated observation problem. Effectively, the latent variables represent individual attitudes, so that for each individual the same realization of each latent variable is considered.

After analyzing the data it is possible to recognize 5 underlying dimensions that may affect the decision of adopting VoIP. Specifically, knowledge (of VoIP functionalities, requirements, and charges), VoIP awareness, and early adoption of new technologies are tested to determine their effect on telephony choice (Table **??**).

Variable	Description
z^*_{1n}	Knowledge of IP telephony functionalities
z^*_{2n}	Knowledge of IP telephony requirements
z^*_{3n}	Knowledge of IP telephony charges
z^*_{4n}	Awareness of IP telephony
z^*_{5n}	Early adoption of new technologies

Table 6.2: Latent variables

The latent variables are manifested through a subset of the effect indicators shown in Table **??**[10]. Note that the possible answers for each indicator are binary perceptions.

Before introducing the results of the model, note that Appendix A presents the notation for both the dependent and independent variables. Using this notation[11], the following table provides the results of the choice model. The simultaneously estimated structural and measurement models for the latent variables are shown in Appendix B.

[9]The empirical applications in the previous chapters consider continuous effect indicators.

[10]The actual subset for each latent variable is shown when presenting the results of the estimated model.

[11]The notation of the coefficients follows the general notation introduced to present the general model.

Indicator	Description
I_1	It is possible to make calls from IP phone to fixed phone
I_2	It is possible to make calls from IP phone to cellular phone and PHS
I_3	It is possible to make calls from IP phone to overseas
I_4	It is possible to make calls from overseas to IP phone
I_5	It is possible to make calls from fixed phone to IP phone
I_6	It is possible to make calls from cellular phone or PHS to IP phone
I_7	It is possible to send faxes from IP phone
I_8	Ordinary telephone sets can be used for IP phone service
I_9	A dedicated special device is needed
I_{10}	A PC is not needed to use IP phone service
I_{11}	No special knowledge is needed to use IP phone service
I_{12}	Charge for a local call from IP phone to a fixed phone is lower than from a fixed phone
I_{13}	Charge for a long distance call is less expensive from an IP phone than from a fixed phone
I_{14}	Charge for a call to a cell phone is less expensive from an IP phone than from a fixed phone
I_{15}	Charge for an overseas call is less expensive from an IP phone than from a fixed phone
I_{16}	There is no charge for a call between IP phones from same provider or same provider group
I_{17}	There is no initial subscription charge for IP phone service
I_{18}	Monthly basic charge for IP phone service is less than ¥1000
I_{19}	Have you heard about IP phone service?
I_{20}	I prefer to adopt new products or services ahead of other people
I_{21}	I prefer to take up new products or services after they have become generally accepted
I_{22}	Are you already using IP phone service?

Table 6.3: Telephony choice: Effect Indicators

In the choice model, each paramater is estimated according to a probit model. The estimates are displayed in Table **??**. The effect of the latent variables is added to the probit kernel. These latent variables are individual-specific (to account for repeated observations) with parameters that are alternative-specific (for the alternatives POIP and IP).

Some interesting results can be derived from the estimates. For instance, all attributes associated with cost[12] have a negative marginal utility, which means that the probability of choosing a given alternative will decrease if its cost increases. Other valuations are related with certain features that VoIP sometimes does not provide. As discussed in the introduction to this chapter, the lack of certain features within the VoIP services may be an issue when QoS is not guaranteed[13]. If it is not possible to call mobile phones using

[12]Namely monthly fixed charges, charges for local and long distance calls, mobile charges, and initial cost.

[13]Lack of QoS is a characteristic of the earliest stages of VoIP. However, when the broadband is adequate QoS can be guaranteed and VoIP can be offered featuring number portability, emergency calling, and fax capabilities. However, as in 2008 the Japanese market was still dominated by best-

Parameter	Description	Estimate	s.e.	t-stat
ASC_{PO}	PO Constant	-0.5200	0.1978	-2.63
ASC_{POIP}	POIP Constant	-0.0880	0.1366	-0.64
β_1	Monthly Charge PO	-0.2400	0.0392	-6.12
β_2	Monthly Charge IP	-0.7800	0.1219	-6.40
β_3	Initial Cost IP	-0.2200	0.0858	-2.56
β_4	Local Call	-0.0130	0.0089	-1.46
β_5	Long Distance	-0.0023	0.0013	-1.77
β_6	Call Mobile	-0.0011	0.0003	-3.67
β_7	No Mobile IP	-0.3500	0.0422	-8.29
β_8	Emergency	0.3300	0.0422	7.82
β_9	Special Number	0.1200	0.0235	5.11
β_{10}	Tokyo POIP	-0.0310	0.0381	-0.81
β_{11}	Tokyo IP	-0.0370	0.0452	-0.82
β_{12}	Web POIP	-0.2300	0.0494	-4.66
β_{13}	Web IP	-0.3000	0.0440	-6.82
β_{14}	Male POIP	-0.0670	0.0303	-2.21
β_{15}	Male IP	0.0840	0.0368	2.28
β_{16}	Age 30- POIP	0.0026	0.0047	0.55
β_{17}	Age 30- IP	-0.0140	0.0056	-2.50
β_{18}	Age 30-50 POIP	0.0069	0.0024	2.88
β_{19}	Age 30-50 IP	0.0041	0.0026	1.58
β_{20}	Age 50+ POIP	0.0120	0.0036	3.33
β_{21}	Age 50+ IP	0.0000	0.0041	0.00
β_{22}	Mobile Charge 8- POIP	0.0069	0.0045	1.53
β_{23}	Mobile Charge 8- IP	-0.0064	0.0052	-1.23
β_{24}	Mobile Charge 8+ POIP	0.0051	0.0029	1.76
β_{25}	Mobile Charge 8+ IP	0.0048	0.0033	1.45
β_{26}	PO Charge 3- POIP	-0.0072	0.0143	-0.50
β_{27}	PO Charge 3- IP	-0.1600	0.0194	-8.25
β_{28}	PO Charge 3+ POIP	0.0240	0.0079	3.04
β_{29}	PO Charge 3+ IP	-0.0006	0.0092	-0.07
β_{30}	Voice Often POIP	0.1200	0.0423	2.84
β_{31}	Voice Often IP	0.1600	0.0464	3.45
β_{32}	Switched I-net POIP	0.0031	0.0240	0.13
β_{33}	Switched I-net IP	0.1300	0.0302	4.30
β_{34}	Awareness x Monthly Charge IP	-0.0870	0.0518	-1.680
Γ_{11}	z_1 on POIP	0.0580	0.0266	2.180
Γ_{12}	z_2 on POIP	0.1100	0.0275	4.000
Γ_{13}	z_3 on POIP	-0.0110	0.0190	-0.579
Γ_{14}	z_4 on POIP	0.2400	0.0370	6.486
Γ_{15}	z_5 on POIP	0.0590	0.0203	2.906
Γ_{21}	z_1 on IP	-0.0680	0.0321	-2.118
Γ_{22}	z_2 on IP	0.1300	0.0264	4.924
Γ_{23}	z_3 on IP	0.0610	0.0233	2.618
Γ_{24}	z_4 on IP	0.2600	0.0213	12.207
Γ_{25}	z_5 on IP	0.0650	0.0212	3.066
$\ell(\theta)$	-15136.78			
$\ell(ASC)$	-18096.36			

Table 6.4: Telephony choice model

VoIP, then the overall satisfaction with the VoIP service is reduced. However, when other services are provided, the overall satisfaction improves. For instance, when VoIP does offer the possibility of making emergency calls there is a consequent satisfaction perceived by the user, which can be derived from the positive marginal utility of the corresponding attribute (Emergency). The same conclusion holds when VoIP allows users to call other special numbers. In addition, users dislike the constraint of some VoIP services which do not allow them to make calls to mobile phones; this is reflected by the negative marginal utility of the variable *No mobile IP*.

A complete subset of the parameters represents the effect of user segmentation. For instance, there is the effect of where the sample was taken[14], gender, and age. For example, men prefer VoIP over the combination of POTS and VoIP.

Finally, the HCM specification allows us to determine the effect of the latent constructs[15] that were identified on the choice probabilities. Alternative-specific parameters were adopted for each of the latent variables for the alternatives POIP and IP. It is expected that the latent variables have positive marginal utilities. The results show that this fact is almost always true, with two exceptions. The third latent variable has a negative effect on POIP, but this result is not significantly different from zero. However, the first latent variable has a negative effect on the utility function of IP. Note however that since the the first latent variable represents knowledge of VoIP functionalities, this latent construct may reflect the situation of the constraints of best-effort VoIP. Because of these constraints, VoIP is preferred when combined with POTS rather than as a service by itself[16]. But in general the results show that knowledge[17] as well as a positive attitude toward the adoption of new technologies favor the adoption of VoIP either directly (IP) or in combination with traditional telephony (POIP). Note that the marginal effect on the utility function of both the POIP and IP alternatives (and hence the marginal effect on the choice probabilities of adopting these two alternatives) is considerably higher for the latent variable *awareness of VoIP*. Awareness of VoIP is an underlying measure of general knowledge of VoIP and it summarizes the specific dimensions covered in the first three latent variables. In effect, the more people know about VoIP, the more they are willing to accept this new service.

---

effort VoIP.

[14] Tokyo or Web sample.

[15] Knowledge of VoIP functionalities, Knowledge of VoIP requirements, Knowledge of VoIP charges, Awareness of VoIP, and Early adoption of new technologies.

[16] The effect of the first latent variable on POIP is positive.

[17] The first four latent variables are related to different dimensions of knowledge about VoIP services.

In Table ?? the Bayesian quantiles are given. These quantiles are calculated from the posterior distribution of the parameters. Recall that unlike frequentist estimation, in the Bayesian approach the parameters have an associated distribution.

Since a probit kernel was adopted, the covariance matrix for the model expressed in differences with respect to VoIP is estimated (Table ??). The results indicate the presence of both heteroscedasticity and correlation. Because of this result, it can be argued that a simpler model, such as an MNL, should be avoided.

The choice model that has been discussed in detail is estimated simultaneously with the structural and measurement equations of the latent variables. Both the structural and measurement equations make it possible to have draws for the latent variables that are introduced in the choice model. I emphasize the choice model since the analysis of the marginal effects of the latent variables on the choice probabilities of adopting VoIP is made using the discrete choice kernel. However, some interesting conclusions can be deduced from the structural equations (Table ??). For instance, broadband access has a positive effect on all latent variables. This positive effect can be seen in the effect of the type of I-net access that the user currently has (through the estimates of the indicator variables *DU user*, *ADSL/CATV user*, and *FTTH user*). The faster the I-net connection, the higher the effect on the latent variables. This is particularly true for users with FTTH access, which appears as a significant explanatory variable of early adoption of new technologies.

In Appendix B I show the complementary results of the estimation for the measurement equations which provide identification of the latent variables.

6.5 Forecasting consumer response to VoIP

The marginal utilities as well as the parameters of the structural equations of the latent variables[18] describe user behavior in terms of the probability of adopting VoIP. For instance, the marginal utilities weigh the attributes and latent explanatory variables, allowing us to model the trade-offs faced by the consumers and to forecast the market shares of the different alternatives. However, a true understanding of the meaning of the estimates beyond analyzing sign and magnitude of the marginal utilities comes from applying the model to forecast different scenarios. Taking the experimental design as

[18]The measurement equations provide identification of the latent variables.

Parameter	Description	2.50%	5%	50%	95%	97.50%
ASC_{PO}	PO Constant	-0.9247	-0.8473	-0.5147	-0.1969	-0.1392
ASC_{POIP}	POIP Constant	-0.3480	-0.3102	-0.0904	0.1366	0.1784
β_1	Monthly Charge PO	-0.3206	-0.3062	-0.2323	-0.1768	-0.1673
β_2	Monthly Charge IP	-1.0215	-0.9846	-0.7784	-0.5861	-0.5569
β_3	Initial Cost IP	-0.4028	-0.3663	-0.2149	-0.0874	-0.0626
β_4	Local Call	-0.0308	-0.0281	-0.0128	0.0012	0.0039
β_5	Long Distance	-0.0049	-0.0045	-0.0023	-0.0002	0.0002
β_6	Call Mobile	-0.0017	-0.0016	-0.0011	-0.0005	-0.0004
β_7	No Mobile IP	-0.4234	-0.4101	-0.3493	-0.2709	-0.2568
β_8	Emergency	0.2499	0.2597	0.3308	0.3988	0.4100
β_9	Special Number	0.0717	0.0792	0.1156	0.1572	0.1647
β_{10}	Tokyo POIP	-0.1094	-0.0948	-0.0300	0.0309	0.0454
β_{11}	Tokyo IP	-0.1267	-0.1121	-0.0366	0.0376	0.0500
β_{12}	Web POIP	-0.3363	-0.3184	-0.2254	-0.1557	-0.1451
β_{13}	Web IP	-0.3854	-0.3720	-0.3002	-0.2266	-0.2128
β_{14}	Male POIP	-0.1341	-0.1193	-0.0653	-0.0217	-0.0127
β_{15}	Male IP	0.0110	0.0226	0.0861	0.1436	0.1531
β_{16}	Age 30- POIP	-0.0064	-0.0050	0.0025	0.0105	0.0121
β_{17}	Age 30- IP	-0.0250	-0.0234	-0.0141	-0.0047	-0.0031
β_{18}	Age 30-50 POIP	0.0027	0.0033	0.0067	0.0112	0.0121
β_{19}	Age 30-50 IP	-0.0009	-0.0002	0.0040	0.0083	0.0092
β_{20}	Age 50+ POIP	0.0054	0.0063	0.0113	0.0183	0.0195
β_{21}	Age 50+ IP	-0.0080	-0.0067	-0.0001	0.0068	0.0081
β_{22}	Mobile Charge 8- POIP	-0.0017	-0.0004	0.0067	0.0146	0.0163
β_{23}	Mobile Charge 8- IP	-0.0165	-0.0148	-0.0065	0.0023	0.0036
β_{24}	Mobile Charge 8+ POIP	-0.0002	0.0006	0.0050	0.0102	0.0113
β_{25}	Mobile Charge 8+ IP	-0.0016	-0.0007	0.0048	0.0103	0.0112
β_{26}	PO Charge 3- POIP	-0.0344	-0.0302	-0.0074	0.0163	0.0219
β_{27}	PO Charge 3- IP	-0.2005	-0.1950	-0.1653	-0.1310	-0.1240
β_{28}	PO Charge 3+ POIP	0.0100	0.0121	0.0234	0.0377	0.0410
β_{29}	PO Charge 3+ IP	-0.0188	-0.0157	-0.0005	0.0144	0.0172
β_{30}	Voice Often POIP	0.0388	0.0499	0.1120	0.1905	0.2065
β_{31}	Voice Often IP	0.0655	0.0792	0.1578	0.2329	0.2452
β_{32}	Switched I-net POIP	-0.0465	-0.0375	0.0036	0.0408	0.0472
β_{33}	Switched I-net IP	0.0644	0.0759	0.1269	0.1749	0.1854
β_{34}	Awareness x Monthly Charge IP	-0.1892	-0.1727	-0.0863	-0.0013	0.0135
Γ_{11}	z_1 on POIP	0.0115	0.0173	0.0565	0.1053	0.116
Γ_{12}	z_2 on POIP	0.0656	0.0725	0.1126	0.1639	0.1738
Γ_{13}	z_3 on POIP	-0.0503	-0.0431	-0.0108	0.0185	0.0257
Γ_{14}	z_4 on POIP	0.1788	0.1839	0.2398	0.305	0.3149
Γ_{15}	z_5 on POIP	0.0231	0.0292	0.0581	0.0945	0.1022
Γ_{21}	z_1 on IP	-0.1296	-0.1199	-0.0689	-0.0142	-0.003
Γ_{22}	z_2 on IP	0.0807	0.0898	0.1331	0.1768	0.1835
Γ_{23}	z_3 on IP	0.0147	0.0226	0.0607	0.0988	0.1055
Γ_{24}	z_4 on IP	0.2188	0.2241	0.2608	0.2938	0.2993
Γ_{25}	z_5 on IP	0.0246	0.0306	0.0648	0.0994	0.1062

Table 6.5: Telephony choice model - quantiles

Estimates		s.e.	
1.00	0.35	-	0.20
0.35	0.67	0.20	0.13

Table 6.6: Telephony choice model - covariance matrix

	z_1		z_2		z_3		z_4		z_5	
	est.	s.e.	est.	s.e.	estimate	s.e.	est.	s.e.	est.	s.e.
Intercept	-1.6000	0.506	0.0680	0.337	0.1600	0.367	-4.8000	0.173	0.9900	0.321
Age 30-	0.0530	0.015	0.0065	0.013	0.0030	0.012	0.0920	0.006	-0.0390	0.011
Age 30-50	-0.0290	0.007	0.0150	0.005	0.0081	0.005	-0.0170	0.002	-0.0180	0.004
Age 50+	-0.0110	0.009	0.0140	0.008	0.0080	0.008	-0.0045	0.003	0.0047	0.007
male	0.3100	0.072	-0.2900	0.061	0.0025	0.057	0.5300	0.028	0.3800	0.054
DU user	-0.1400	0.179	0.4100	0.177	0.2000	0.205	0.3300	0.087	0.0450	0.157
ADSL/CATV user	0.3600	0.165	0.6300	0.155	0.0025	0.201	1.2000	0.083	0.0840	0.124
FTTH user	0.4300	0.183	0.3300	0.174	0.2600	0.210	1.1000	0.090	0.6200	0.143
Cell only	-	-	-	-	-	-	-0.6100	0.112	-	-
Years PC [2,10]	0.3900	0.122	-0.2500	0.122	-0.0600	0.123	1.1000	0.055	-0.0900	0.098
Years PC 10+	0.6700	0.142	-0.4000	0.129	-0.4000	0.133	1.5000	0.060	0.1200	0.105

Table 6.7: Telephony structural latent variable model

baseline, I simulate the impact on the choice probabilities (and thus on the market shares of POTS, POIP, and VoIP) of the following eight hypothetical market conditions:

1. Scenario 1: No initial charges for VoIP service
2. Scenario 2: Initial charges drop to 50% of the experimental figures
3. Scenario 3: Experimental monthly VoIP charges are cut by half combined with no initial charges for VoIP
4. Scenario 4: Emergency calls are made available for all users
5. Scenario 5: Awareness of VoIP is increased by 50%
6. Scenario 6: Early adoption of new technologies is increased by 100%
7. Scenario 7: All consumers become early adopters
8. Scenario 8: All consumers become aware of VoIP

The first 3 scenarios consider situations where the costs of accessing (initial charge) or of using (monthly charge) VoIP are reduced. Scenario 4 reflects the situation where calling to emergency numbers from VoIP is no longer an issue because this feature is made available to everyone. The last 4 scenarios are related with latent *awareness* and *early adoption*. Scenarios 5 and 6 both represent a situation where the distribution of the latent variables is adjusted. Scenarios 7 and 8 consider a degenerate distribution of the latent variable where all the mass is concentrated at its maximum. For instance, from the estimation procedure it is possible to describe the distribution of the latent early adoption. Early adopters show an increased probability of adopting VoIP. The highest early adoption is taken and in Scenario 7 we make the assumption that all consumers behave as the best new adopter. In a similar way, in Scenario 8 all consumers are assumed to have the same behavior as the most aware user (for example, as a result of informative campaigns.)

In Table **??** the predicted market shares and percent change are displayed (See also Figure **??**.) Because a probit kernel was assumed, 200 repetitions of the GHK simulator for the choice probabilities were used for forecasting with the model.

		Market shares			Percent change		
		POTS	POIP	VoIP	POTS	POIP	VoIP
Base	SP experimental design	41.4%	38.3%	20.3%			
Scenario 1	No initial charge IP	41.1%	37.9%	20.9%	-0.7%	-1.0%	3.0%
Scenario 2	50% initial charge IP	41.3%	38.1%	20.6%	-0.2%	-0.5%	1.5%
Scenario 3	50% monthly charge IP + no initial charge	40.6%	37.3%	22.1%	-1.9%	-2.6%	8.9%
Scenario 4	Emergency calls always available	37.9%	33.8%	28.3%	-8.5%	-11.7%	39.4%
Scenario 5	150% more awareness	40.4%	37.1%	22.6%	-2.4%	-3.1%	11.3%
Scenario 6	200% more adoption of new technologies	41.0%	37.8%	21.2%	-1.0%	-1.3%	4.4%
Scenario 7	Max adoption	39.6%	36.1%	24.4%	-4.3%	-5.7%	20.2%
Scenario 8	Max awareness	39.3%	36.0%	24.7%	-5.1%	-6.0%	21.7%

Table 6.8: Predicted VoIP penetration

All of the hypothetical scenarios represent situations where VoIP becomes more attractive. As a result of the improved features of VoIP, its market share increases in all situations. However, the impact on the market share of VoIP depends on the scenario being modeled. For instance, the biggest impact on the VoIP market share within the considered scenarios is the capability of calling emergency numbers. When this feature is made available for all users in the sample, the VoIP penetration increases by 39.4%. As discussed previously, VoIP may not be able to perform emergency calls[19] (basically because of how VoIP works it is not trivial to locate the position from where the call is being made.) This result shows the importance that users allocate to this VoIP feature, and hence in order to increase the current market shares and to consolidate the presence

[19] Even though QoS VoIP makes it possible to call emergency numbers, the Japanese market is currently dominated by best-effort VoIP for which emergency calls are not available.

of VoIP in the market it is essential to provide access to emergency calls (which is possible with QoS VoIP). Increasing general awareness of and knowledge about VoIP[20] also considerably increases the penetration of VoIP. Almost the same result is obtained for underlying early adoption. Information campaigns can be targeted as a tool to increase knowledge about how VoIP works.

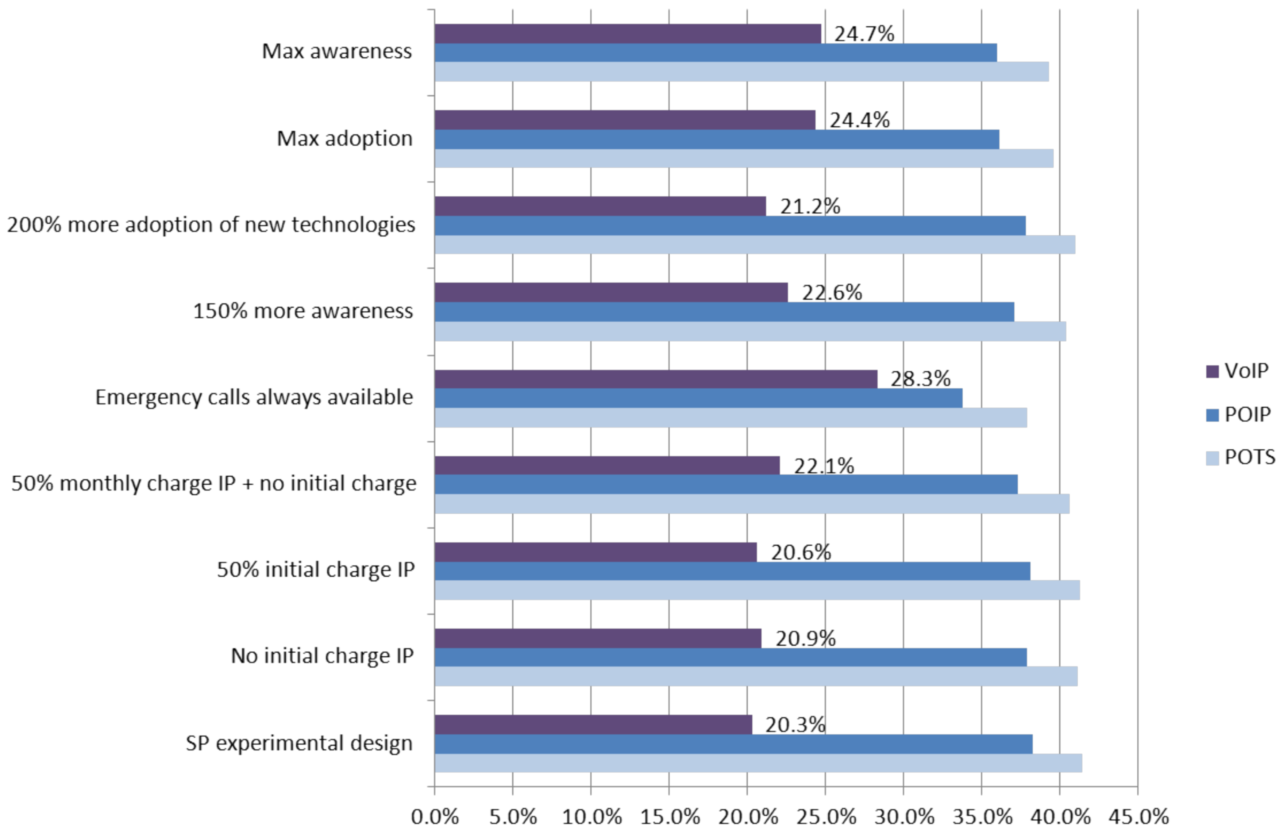

Figure 6.13: Predicted shares for VoIP

6.6 Conclusions

Hybrid choice modeling is a powerful tool for modeling the adoption of new technologies because it allows one to incorporate the effect of qualitative attributes and attitudes on the probabilities of adopting the new technologies. Effectively, using an HCM with a

[20] Recall that Scenarios 7 and 8 consider a situation where all users are assumed to have the maximum value obtained during the estimation process for the latent variables describing early adoption and awareness.

probit kernel and dichotomous effect indicator variables I have shown that knowledge as well as a positive attitude toward the adoption of new technologies favor the adoption of VoIP either directly or in combination with traditional telephony. For instance, the model describes the segments of the population that can be identified as early adopters. Technically, early adopters in the model are individuals with a high value for the attitude *early adoption of new technologies*, which is statistically identified through self reported behavior related to technological adoption habits. According to my results, early adopters of new technologies in Japan are men that have FTTH I-net access, and that have had access to a computer for more than 10 years. This segment of the population has a correspondingly higher probability of adopting VoIP. According to the forecasts of the model, in order to increase the penetration of IP telephony it is essential to assure QoS in order to provide access to desired features such as emergency calls. Information campaigns targeted at improving telephony users' knowledge about VoIP also appear as a suitable tool to consolidate the adoption of VoIP in Japan.

Regarding the estimation of the model, the main finding is that the Bayesian approach is perfectly suited for a complex hybrid choice model that considers the presence of correlation and heteroscedasticity, together with a relatively large number of latent variables and dichotomous effect indicators. Finally, the inclusion of latent variables provides a very interesting approach for taking into account the problem of repeated observations in stated preference studies.

References

W. Beckert. Estimation of heterogenous preferences, with an application to demand for Internet services. *The Review of Economics and Statistics*, 87(3):495–502, 2005.

M. Ben-Akiva, 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): 163–175, 2002.

J. Bodnar, P. Dilworth, and S. Iacono. Cross-section analysis of residential telephone subscription in Canada. *Information Economics and Policy*, 4(3):311–331, 1988.

D. Bolduc and M. Ben-Akiva. Estimation of IP telephony demand using the integrated choice and latent variables approach. Computing in Economics and Finance 2006 303, Society for Computational Economics, 2006.

K.T. Duffy-Deno. Business demand for broadband access capacity. *Journal of Regulatory Economics*, 24:359–372, 2003.

J. Eisner and T. Waldon. The demand for bandwidth: second telephone lines and one-line services. *Information Economics and Policy*, 13:301–309, 2001.

R. Fildes and V. Kumar. Telecommunications demand forecasting – a review. *International Journal of Forecasting*, 18:489–522, 2002.

A. Goolsbee. The value of broadband and the deadweight loss of taxing new technology. Mimeo, University of Chicago, 2000.

S. Greenstein. Building and delivering the virtual world: commercializing services for Internet access. *The Journal of Industrial Economics*, XLVIII(4):391–411, 2000.

R.S. Hartman, F. Zareen, and Z.F. Naqvi. Estimation of household preference for long distance telecommunications carrier. *Journal of Regulatory Economics*, 6:197–220, 1994.

T. Ida and Y. Horiguchi. Consumer benefits of public services over FTTH in Japan: Comparative analysis of provincial and urban areas by using discrete choice experiment. *The Information Society*, 24(1):1–17, 2008.

T. Ida and T. Kuroda. Discrete choice analysis of demand for broadband in Japan. *Journal of Regulatory Economics*, 29:5–22, 2006.

T. Ida and T. Kuroda. Discrete choice model analysis of mobile telephone service demand in Japan. *Empirical Economics*, 36:65–80, 2009.

T. Ida and K. Sakahira. Broadband migration and lock-in effects: Mixed logit model analysis of Japan's high-speed Internet access services. *Telecommunications Policy*, 32:615–625, 2008.

T. Ida, S. Kinoshita, and M. Sato. Conjoint analysis of demand for IP telephony: the case of Japan. *Applied Economics*, 40:1279–1287, 2008.

D.J. Kridel. A consumer surplus approach to predicting extended area service (EAS) development and stimulation rates. *Information Economics and Policy*, 3:379–390, 1988.

D.J. Kridel, P.N. Rappoport, and L.D. Taylor. An economic model of the demand for access to the internet by cable modem. In D.G. Loomis and L.D. Taylor, editors, *Forecasting the internet: understanding the explosive growth of data communications*. Kluwer Academic Publishers, Massachusetts, 2001.

G. Madden and M. Simpson. Recursive models with qualitative endogenous variables. *Annals of Economic and Social Measurement*, 5:525–545, 1997.

G. Madden, S.G. Savage, and G. Coble-Neal. Subscriber churn in the Australian ISP market. *Information Economics and Policy*, 11:195–208, 1999.

L.J. Perl. Economic and demographic determinants of telephone availability. AT&T filing before the Federal Communications Comission docket no. 20003, exhibit 21, 1978a.

L.J. Perl. Economic and demographic detrminants for basic telephone service. Working Paper, National Economic Research Associates, White Plains, New York, 1978b.

L.J. Perl. Residential demand for telephoneservice 1983. Working Paper prepared for the Central Service Organization of the Bell Operating Companies, Inc., National Economic Research Associates, White Plains, New York, 1978c.

J. Raina, R. Fildes, and K. Day. Forecasting Internet telephony. *OR Insight*, 11(4):11–21, 1998.

P.N. Rappoport, D.J. Kridel, L.D. Taylor, and J. Alleman. A residential demand for access to the Internet. In G. Madden, editor, *The International Handbook of Telecommunications Economics Volume II*. Edward Elgar Publishers, Cheltenham, UK, 2002.

P.N. Rappoport, D.J. Kridel, L.D. Taylor, and J. Alleman. The demand for voice over IP: an econometric analysis using survey data on willingness-to-pay. *Telektronikk*, 4.04:70–83, 2004.

S.G. Savage and D. Waldman. Broadband Internet access, awareness, and use: analysis of United States household data. *Telecommunications Policy*, 29:615–633, 2005.

T.J. Tardiff. Effects of presubscription and other attributes on long-distance carrier choice. *Information Economics and Policy*, 7:353–366, 1995.

L.D. Taylor and D.J. Kridel. Residential demand for access to the telephone network. In A. de Fontanay, M.H. Shugard, and D.S. Sibley, editors, *Telecommunications demand modeling*. North Holland Publishing Co., Amsterdam, 1990.

K.E. Train, D.L. McFaddden, and M. Ben-Akiva. The demand for local telephone service: a fully discrete model of residential calling paterns and service choices. *The Rand Journal of Economics*, 18(1):109–123, 1987.

K.E. Train, M. Ben-Akiva, and T. Atherton. Consumption patterns and self-selecting tariffs. *The Review of Economics and Statistics*, 71(1):62–73, 1989.

H.R. Varian. The demand for bandwidth: evidence from the INDEX Project. In J. Alleman and R. Crandall, editors, *Broadband: should we regulate high-speed Internet access?* Brookings Institution, Washington DC, 2002.

Appendix A: Variable description

Variable	Description
U_{PO}	Utility associated with Plain Old Telephone Service POTS
U_{POIP}	Utility associated with mixed use of both POTS and VoIP
U_{IP}	Utility associated with VoIP
EC	Knowledge of VoIP functionalities
EC	Knowledge of VoIP requirements
EC	Knowledge of VoIP charges
EC	Awareness of VoIP
EC	Early adoption of new technologies
I_1	It is possible to make calls from VoIP to fixed phone
I_2	It is possible to make calls from VoIP to cellular phone and PHS
I_3	It is possible to make calls from VoIP to overseas
I_4	It is possible to make calls from overseas to VoIP
I_5	It is possible to make calls from fixed phone to VoIP
I_6	It is possible to make calls from cellular phone or PHS to VoIP
I_7	It is possible to send faxes from VoIP
I_8	Ordinary telephone sets can be used for VoIP service
I_9	A dedicated special device is needed
I_{10}	A PC is not needed to use VoIP
I_{11}	No special knowledge is needed to use VoIP
I_{12}	Charge for a local call from VoIP to a fixed phone is lower than from a fixed phone
I_{13}	Charge for a long distance call is less expensive from VoIP than from a fixed phone
I_{14}	Charge for a call to a cell phone is less expensive from VoIP than from a fixed phone
I_{15}	Charge for an overseas call is less expensive from VoIP than from a fixed phone
I_{16}	There is no charge for a call between IP phones from same provider or same provider group
I_{17}	There is no initial subscription charge for IP phone service
I_{18}	Monthly basic charge for VoIP is less than ¥1000
I_{19}	Have you heard about IP phone (VoIP) service?
I_{20}	I prefer to adopt new products or services ahead of other people
I_{21}	I prefer to take up new products or services after they have become generally accepted
I_{22}	Are you already using VoIP?

Table 6.9: Dependent Variables

Variable	Description
Monthly Charge PO	PO monthly usage charge / log(income) [Thousand ¥]
Monthly Charge IP	VoIP monthly usage charge / log(income) [Thousand ¥]
Initial Cost IP	Initial cost of VoIP service / log(income) [Thousand ¥]
Local Call	Charge for 3 minutes of local conversation [¥]
Long Distance	Charge for 3 minutes of long distance conversation [¥]
Call Mobile	Charge for calling a mobile phone from an IP telephone [¥]
No Mobile IP	IP telephone offered cannot connect to a mobile phone
Emergency	IP telephone offered can make emergency calls
Special Number	IP telephone offered can make special number calls
Tokyo	Respondent is in the Tokyo sample
Web	Respondent is in the Web sample
Male	Gender indicator
Age 30-	Piecewise linear on age (less than 30 years)
Age [30,50]	Piecewise linear on age (30-50 years)
Age 50+	Piecewise linear on age (more than 50 years)
Mobile Charge 8-	Piecewise linear on monthly charge usually spent on cell (less than ¥8)
Mobile Charge 8+	Piecewise linear on monthly charge usually spent on cell (¥8 or more)
PO Charge 3-	Piecewise linear on monthly charge usually spent on fixed phone (less than ¥3)
PO Charge 3+	Piecewise linear on monthly charge usually spent on fixed phone (¥3 or more)
Voice Often	Voice telecommunications used very often
Switched I-net	Respondent switched internet Service Provider 2 times or more
DU user	Respondent currently uses Dial Up
ADSL user	Respondent currently uses ADSL
CATV user	Respondent currently uses CATV I-net
ADLS/CATV user	Respondent currently uses either ADSL or CATV I-net
FTTH user	Respondent currently uses FTTH
No I-net	Respondent does not have Internet access
Cell only	Only a cellular phone is used at home
Years PC 2-	Respondent has been using a PC for less than 2 years
Years PC [2,10]	Respondent has been using a PC between 2 and 10 years
Years PC 10+	Respondent has been using a PC for more than 10 years

Table 6.10: Dependent Variables

Appendix B: Measurement Model

	Intercept		z_1		z_2		z_3		z_4		z_5	
Ind.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
I_1	1.5000	0.352	1.0000	-					-0.0710	0.019		
I_2	0.4800	0.040	1.2000	0.027					0.6400	0.012		
I_3	0.2800	0.073	-0.1100	0.023					0.7500	0.022		
I_4	-0.2900	0.106	0.7900	0.015					0.2700	0.017		
I_5	0.7200	0.036	-0.2400	0.022					0.9400	0.019		
I_6	0.2100	0.062	0.7700	0.015					0.1700	0.018		
I_7	-0.3700	0.051	-0.3500	0.025					0.5900	0.015		
I_8	0.0420	0.142			1.0000	-			0.0260	0.018		
I_9	0.2400	0.054			-0.0140	0.020			0.7500	0.015		
I_{10}	0.1200	0.043			0.8600	0.017			0.0027	0.018		
I_{11}	0.5700	0.035			-0.2100	0.025			0.0550	0.009		
I_{12}	-0.5200	0.114					1.0000	-	0.0017	0.017		
I_{13}	0.1800	0.049					-0.1700	0.020	0.1900	0.010		
I_{14}	-0.8400	0.033					0.5300	0.015	-0.0700	0.020		
I_{15}	-0.4500	0.032					0.9700	0.026	0.3900	0.010		
I_{16}	-0.9000	0.012					-0.2800	0.021	0.1200	0.018		
I_{17}	-1.3000	0.016					0.2500	0.010	-0.0670	0.009		
I_{18}	-0.4300	0.021					-0.0920	0.020	-0.2600	0.019		
I_{19}	0.2500	0.041							1.0000	-		
I_{20}	-0.8900	0.173							0.0830	0.025	1.0000	-
I_{21}	-0.2300	0.046							0.1200	0.021	0.7500	0.020
I_{22}	-1.1000	0.056	0.5400	0.014	0.7100	0.016	0.7000	0.014	-0.0740	0.019	-0.0480	0.018

Table 6.11: Telephony measurement latent variable model

Appendix C: Posterior distributions and MCMC sequences

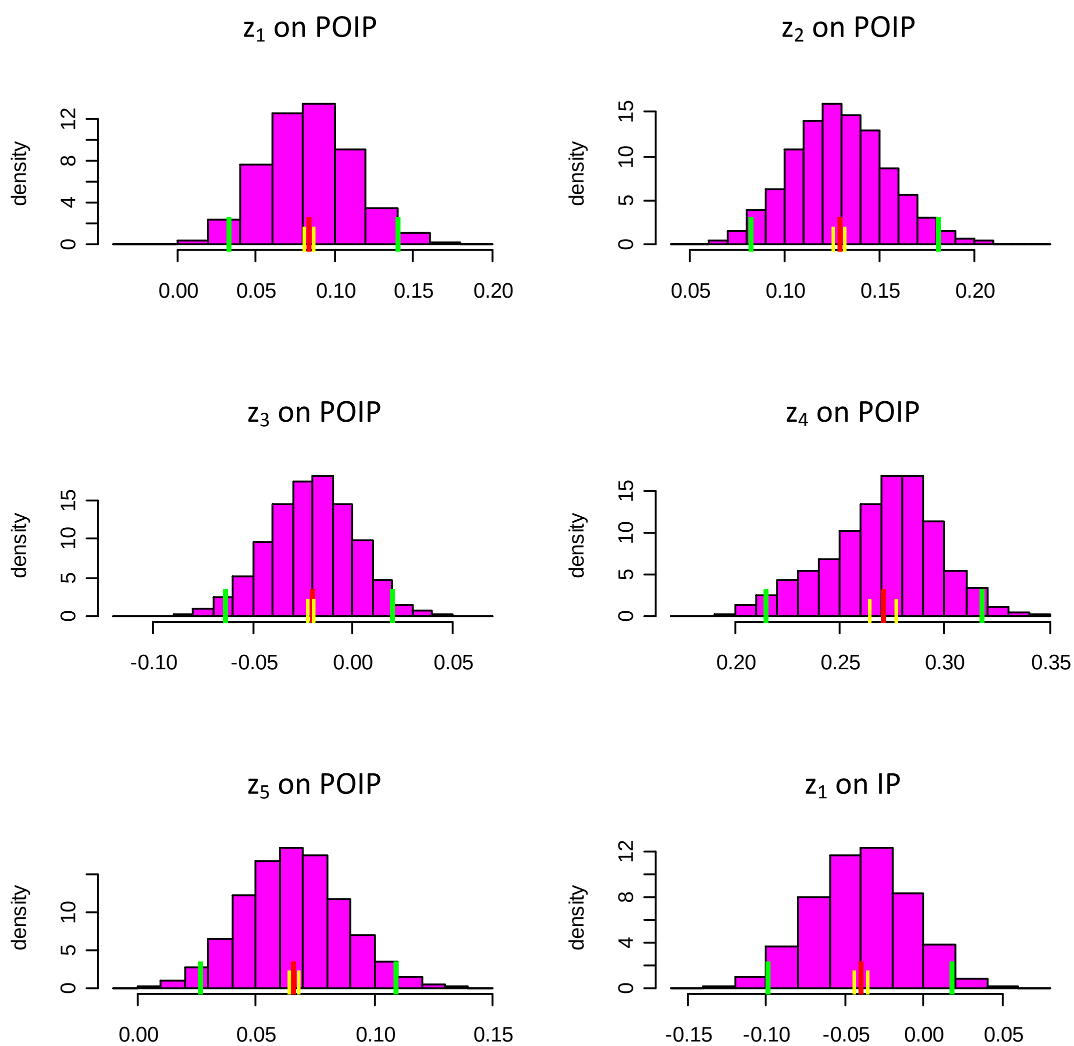

Figure 6.14: Posterior distributions of the latent variable parameters on POIP

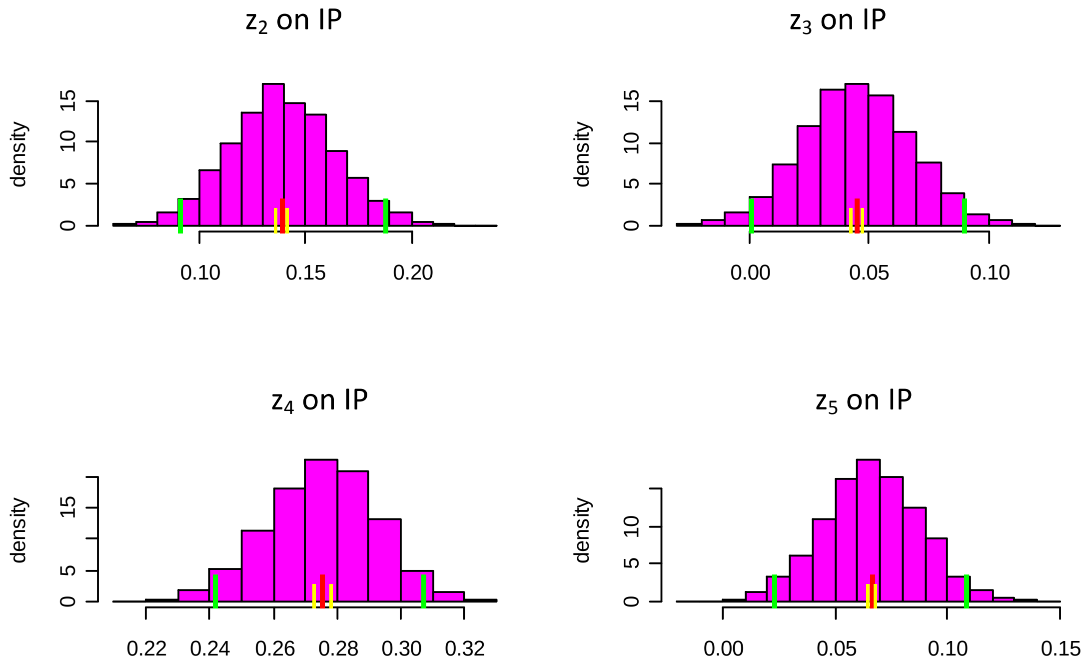

Figure 6.15: Posterior distributions of the latent variable parameters on IP

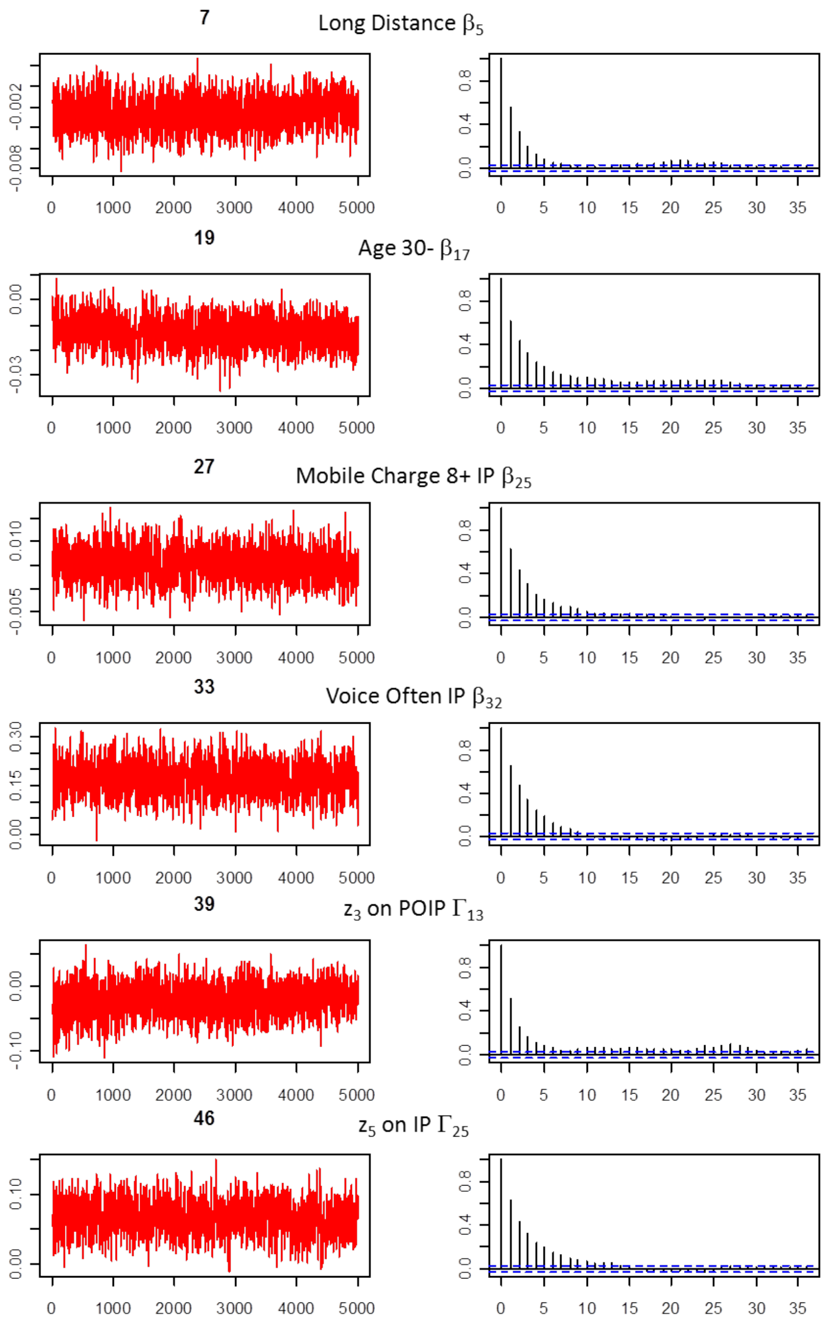

Figure 6.16: MCMC sequence of selected parameters

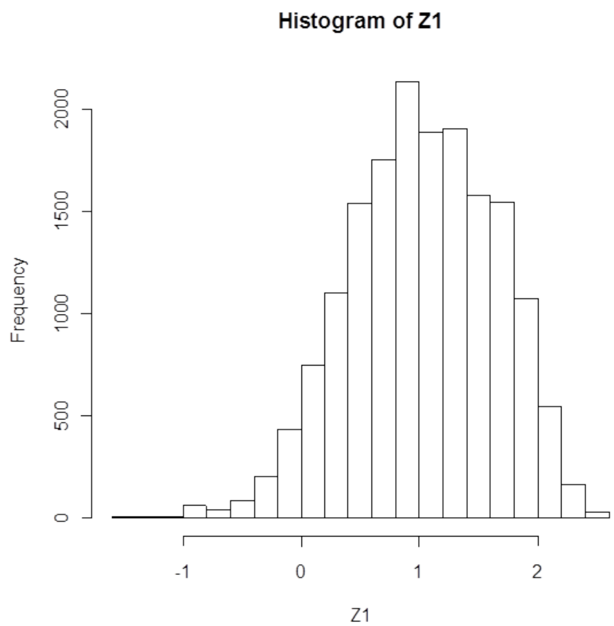

Figure 6.17: Knowledge of IP telephony functionalities

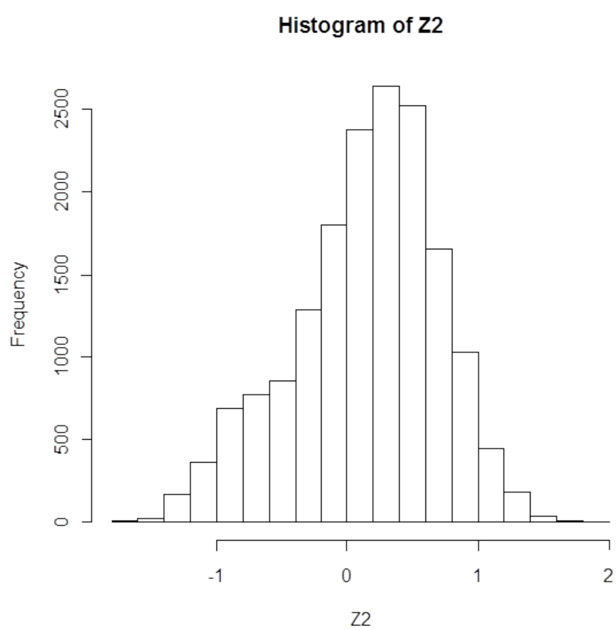

Figure 6.18: Knowledge of IP telephony requirements

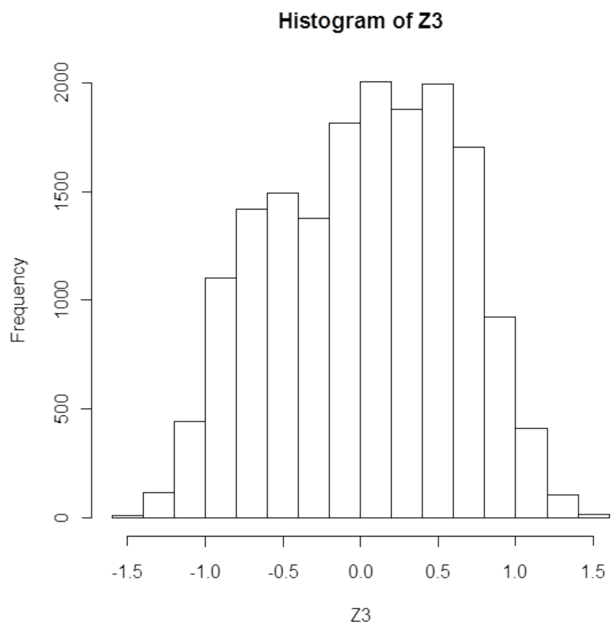

Figure 6.19: Knowledge of IP telephony charges

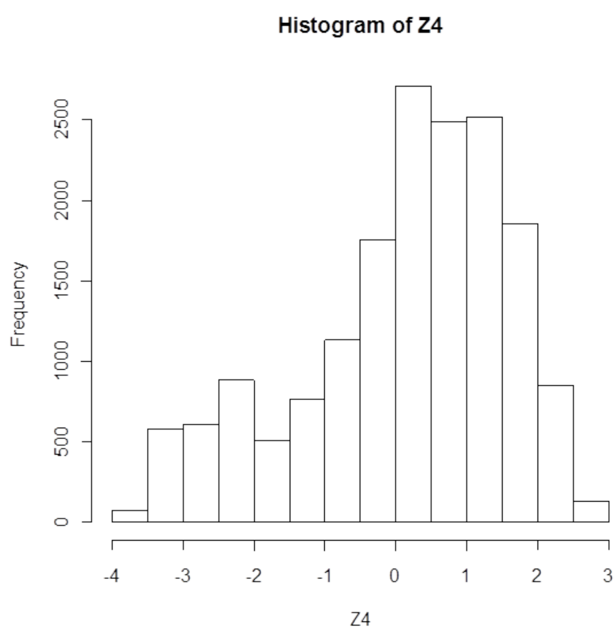

Figure 6.20: Awareness of IP telephony

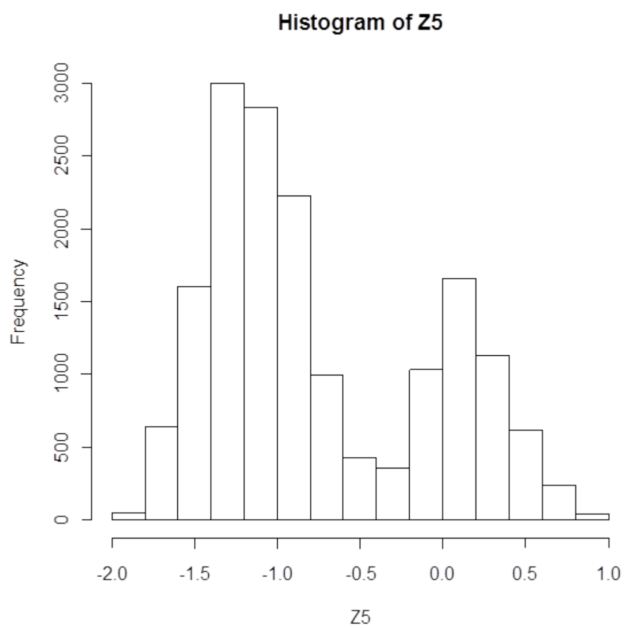

Figure 6.21: Early adoption of new technologies

Chapter 7

Conclusion

Hybrid choice models are a generalization of standard discrete choice models where independent expanded models are considered simultaneously. In particular, the extension that accommodates a discrete choice kernel with latent explanatory variables is of particular interest. The hybrid choice model that represents integrated choice with latent variables is written as a simultaneous system of structural equation models, where the latent variables are mapped using effect and causal indicators.[1]

The results of my dissertation, which has as its main objective to study the application of hybrid choice models both theoretically and empirically, is consistent with the reemerged trend in discrete choice modeling toward incorporating attitudinal factors into the behavioral representation of the decision process. Hybrid choice models offer an attractive improvement in modeling choice behavior, because the choice model is only a part of the underlying behavioral process in which we now incorporate individual attitudes, opinions and perceptions, thus yielding a more realistic econometric model. As discussed in Chapter 2, the introduction of attitudes into a structural model of choice is supported by different theories in social psychology and cognitive science. If we omit the role of attitudes, which is the case in standard economic preference models, we may face problems related to endogeneity.[2] In addition, the hybrid choice modeling framework is also perfectly suited for the introduction of quality as an explanatory variable of choice. Qualitative attributes[3], as opposed to quantitative attributes[4], do not have a natural

[1]For a definition of these concepts, consult Chapter 2.

[2]I.e. an independent variable may be correlated with the error term.

[3]Such as comfort in a travel mode choice context. See the case study example in Chapter 3.

[4]Such as travel time in a travel mode choice context

order or an overt measurement scale, thus qualitative attributes are often introduced as categorical variables on a nominal scale. When the qualitative attribute is measured with error, econometric problems related to endogeneity arise; however, a hybrid choice model is capable of representing the underlying structure of quality.

From the point of view of econometrics, **the most important contribution of my current research is that even though the estimation of hybrid choice models requires the evaluation of intractable complex multidimensional integrals, frequentist full information simulated maximum likelihood[5] and Bayesian MCMC[6] methods can successfully be implemented to practical situations, and both offer unbiased, consistent, and smooth estimators of the true probabilities**.

Although feasible, the estimation of HCMs using frequentist (classical) methods can become extremely complex; however in this dissertation I have verified the practical feasibility of the Gibbs sampler I developed for HCM Bayesian estimation[7], exploiting data augmentation techniques for the latent variables. Whereas Gibbs sampling for a probit kernel is analytically straightforward because it also admits the use of data augmentation, in the case of both a multinomial logit (MNL) kernel and a mixed logit (MMNL) kernel one fails to find a closed form full conditional distribution for the taste parameters of the utility function. However, I have shown that it is possible to exploit Metropolis-Hastings (MH) methods for both the MNL and MMNL cases. In fact, I have shown that even though the probit kernel formulation breaks down the methodological complexity of the model, the data augmentation step for the utility function is very demanding in computational terms, and eventually could be outperformed by a logit-based kernel - even with the additional MH step required by logit models. In addition, classical estimation of HCMs is very demanding in situations with a large number of latent variables - each additional latent variable sums another dimension in the joint choice probability. Thus, according to my analysis Bayesian HCM estimation outperforms simulated maximum likelihood: the inclusion of additional latent variables under the Bayesian approach entails simply working with ordinary regressions (i.e. sampling additional draws from a normal distribution.) Another advantage of the Bayesian approach is that it allows us to forecast using the same sample generated for estimation. In fact, since the Bayesian estimates describe the posterior distribution, I have shown that we can directly calculate

[5] See Chapter 3.

[6] See Chapter 5 for a general description and evaluation of the Bayesian estimator, and Chapters 4 and 5 for empirical applications.

[7] Gibbs sampler for a specific choice context is derived in Chapter 4. In Chapter 5, the Gibbs sampling procedure is generalized.

confidence intervals for willingness to pay measures as well as standard deviations for both the choice probabilities and market shares.

Because of the empirical applications used to test the performance of the frequentist and Bayesian estimators, my work also has interesting applied contributions. In general, the hybrid choice modeling approach is extremely promising for studying behavioral intentions in choice situations where qualitative attributes or consumers' attitudes play major roles.[8] Among these choice situations we can envision consumer response to new products. When new products are developed it is important to forecast consumers' reactions in terms of purchase behavior not only for marketing plans aimed at introducing the new product in the market but also for policymaking.[9]

In particular, in this dissertation I analyzed pro-environmental preferences toward low-emission vehicles.[10] The relevance of this choice situation comes from understanding the effects of climate change and energy security concerns on travel behavior. Using a hybrid choice model to explain purchase intentions by Canadian consumers, I show that environmentally-conscious consumers are aware of the dangers of climate change and oil dependency. Whereas standard demand models have a hard time representing eco-friendly behavior, the hybrid choice model is capable of modeling the consequent change in consuming behavior motivated by the consumer's concerns about the environmental externalities of transportation. Effectively, in the model I propose that environmentally-conscious consumers are willing to pay more for sustainable solutions (low-emission vehicles) despite potential drawbacks (such as a reduced refueling availability). This is the first empirical application of Gibbs sampling to a hybrid choice model with real data.

I also analyzed consumer response to innovation in telecommunications[11]. The telecom industry is especially dynamic, with new products[12] and services being constantly introduced. With standard choice models it is hard to explain the decision to adopt new technologies that are at an early stage of development or that currently have a low overall penetration in the market. The adoption of new products depends on behavioral intentions that are not only a result of observable attributes but also of perceptions, attitudes and knowledge. I specifically study the role of knowledge as well as awareness

[8]In Chapter 2, I discussed how attitudes and behavioral intentions that may eventuate in actual behavior are closely (and causally) related. I also introduced the role of qualitative attributes.

[9]For example, use of a new product may need to be regulated depending on potential externalities. Products that are environmentally friendly may have subsidies seeking a higher penetration rate.

[10]See Chapter 4.

[11]In Chapter 6.

[12]New products that potentially create new needs.

and promptness to adopt new technologies as qualitative attributes in the choice of IP telephony access in Japan. Using stated IP telephony choice by Japanese consumers, I consider and apply a hybrid choice model with a probit kernel and dichotomous effect indicators. The model allows one to measure the effect of qualitative attributes related to knowledge as well as the individual attitude toward the adoption of new technologies on the choice probabilities of adopting IP telephony. The main findings of this study suggest that consumers desire a quality of service of IP telephony that will assure some features that in best effort IP telephony are not guaranteed. In particular, according to the forecasts of the model, in order to increase the penetration of IP telephony it is essential to provide access to emergency calls.

7.1 Future Research

Based on the results of my dissertation I identify three immediate lines of future research, where the hybrid choice modeling framework can be applied:

1. Econometric challenges in choice modeling
2. Toward a more comprehensive attitudinal model of choice behavior
3. Consumer response to innovation

7.1.1 Econometric challenges in choice modeling

There are several econometric challenges that emerge from expanding the standard discrete choice modeling framework. Effectively, the estimation of hybrid choice models is computationally very demanding in situations with numerous latent variables and large sets of potentially interrelated choices. To continue the study of hybrid choice modeling using both Bayesian and frequentist techniques, an analysis of MCMC convergence, and the application of flexible nonparametric methods (Bayesian and classical) are areas that can be exploited. In my dissertation I analyzed hybrid choice models that integrate a discrete choice kernel with latent variables; however, more general estimators need to be derived when other expansions are considered. Parameter identification analysis and techniques for forecasting when latent variables are present are two other issues that need to be addressed.

In the specific case of classical estimation, the latent variables affect the behavior of the simulated likelihood function in such a way that a standard optimization algorithm may require a huge number of iterations to converge. This major weakness translates into a specific research question: how to speed up the optimization process, while ensuring consistency properties and numerical convergence. A solution to explore will be the analysis, implementation and testing of adaptive-sampling trust-region techniques.

7.1.2 Toward a more comprehensive attitudinal model of choice behavior

In approaching a full understanding of the underlying process of decision making, my dissertation is only a preliminary investigation of the role of attitudes in discrete choice behavior. Recent research in social psychology has been centered on analyzing the impact of attitudes on behavioral intentions. Potential expansions of the hybrid choice modeling framework can be analyzed by incorporating the economic grounds of decision making into well-established psychometric models of attitude-behavior, such as the Theory of Planned Behavior. Other topics of future research include cognitive theories explaining the formation of, and stability and change in, attitudes; the impact of information, knowledge and habits on both stated intentions and actual actions; and finally, potential semiotic effects on the dynamics of choice behavior.

7.1.3 Consumer's response to innovation

From the empirical application analyzed in my dissertation, hybrid choice modeling emerges as a powerful tool to model consumer response to innovation. First, it would be interesting to develop a methodology for designing a survey conceived for the hybrid choice modeling framework.[13] This methodology could be applied to the context of consumer response to technological innovation in energy supply, which is a relevant element for energy policies.

Expanding on the empirical and theoretical work of my dissertation, as well as on the technical outcomes of the two previously mentioned research projects, an immediate question to address is the demand for low-emission vehicles. Consumers' preferences for

[13] Although the data used in this dissertation is valid for estimation of a hybrid choice model, the surveys were not designed for these models.

low-emission vehicles must be understood, first, in a context where the new technologies often have a low or even zero market share and hence the role of knowledge, experience and information is critical. Second, demand for low-emission vehicles is a decision-making process guided by environmental preferences. Third, because travel demand is derived from the activity system, demand for green vehicles shows complex interactions within the transportation system and with other systems including the natural environment. A comprehensive research project in the context of demand for cleaner and more efficient vehicles should provide tools for analyzing all these components together by developing comprehensive models of travel behavior based on integrating multidisciplinary complementary methods. These models will have to be consistent with the complexity of the transportation, environmental and urban systems, allowing us to predict travel behavior and to model policy scenarios that are compatible with reducing environmental impacts, as well as with an efficient use of energy resources and existent infrastructure.

οὐκ ἔστ᾽ ἐραστὴς ὅστις οὐκ ἀεὶ φιλεῖ