Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model

Ricardo A. Daziano

School of Civil and Environmental Engineering, Cornell University, 305 Hollister Hall, Ithaca, NY 14853.

Abstract

Discrete choice modeling is experiencing a reemergence of research interest in the inclusion of latent variables as explanatory variables of consumer behavior. There are several reasons that motivate the integration of latent attributes, including better-informed modeling of random consumer heterogeneity and treatment of endogeneity. However, current work still is at an early stage and multiple simplifying assumptions are usually imposed. For instance, most previous applications assume all of the following: independence of taste shocks and of latent attributes, exclusion restrictions, linearity of the effect of the latent attributes on the utility function, continuous manifest variables, and an a priori bound for the number of latent constructs. We derive and apply a structural choice model with a multinomial probit kernel and discrete effect indicators to analyze continuous latent segments of travel behavior, including inference on the energy paradox. Our estimator allows for interaction and simultaneity among the latent attributes, residual correlation, nonlinear effects on the utility function, flexible substitution patterns, and temporal correlation within responses of the same individual. Statistical properties of the Bayes estimator that we propose are exact and are not affected by the number of latent attributes.

Keywords: Bayesian microeconometrics; discrete choice models; structural equation modeling; energy paradox
JEL classification: C35, C53, D12

Email addresses: daziano@cornell.edu (Ricardo A. Daziano), Phone: (1) 607 255-2018 - Fax: (1) 607 255-9004 (Ricardo A. Daziano)

1. Introduction

Discrete choice models are a powerful tool for analyzing consumers' decisions among mutually exclusive alternatives. However, standard discrete choice models consider only observable hedonic attributes of the alternatives, failing to incorporate other relevant choice components. These additional components may be attitudinal constructs (such as pro-environmental preferences), as well as multi-dimensional attributes of quality (such as performance) that cannot be measured using a single item. Neglecting these components (omission of a relevant variable) or using proxy variables (measurement error) induce endogeneity. Hence, the incorporation of these underlying components is desirable to achieve consistent, efficient preference estimators. In addition, structural choice models that incorporate underlying attitudes through latent variables (see the seminal paper by Ben-Akiva et al., 2001) offer an attractive improvement in modeling choice behavior, because the discrete choice model is only a part of the underlying behavioral process through which the modeler can better represent quality and attitudinal responses. In this paper we propose to use a multinomial probit kernel with latent attributes and discrete (categorical ordered) effect indicators to enrich the representation of random consumer heterogeneity in transportation choices (cf. Burda and Harding, 2013).

Although the number of empirical applications of choice models with latent attributes is increasing at an exponential rate (Vij and Walker, 2014, Palma et al., 2013, Ben-Akiva et al., 2013, Jensen et al., 2013, Hess and Beharry-Borg, 2012, Hildebrandt et al., 2012, Rosenberger et al., 2012, Rungie et al., 2011, just to give a few recent examples), the standard frequentist estimator (a maximum simulated likelihood estimator, see Bolduc and Daziano, 2010) has several problems that have limited applied research. For instance, relatively flat areas of the simulated loglikelihood create problems of weak identification, local maxima may be multiple, and standard numerical approximations of both the gradient and the Hessian do not ensure convergence. In addition, computation cost of simulation-aided inference is high for medium-sized problems: finding the maximum simulated likelihood estimates can take days even when there are no convergence issues. In fact, maximizing the likelihood function exhibits the curse of dimensionality with respect to the number of latent variables.[1] In a very recently published article, Bhat and Dubey (2014) propose to use the maximum approximate composite marginal likelihood (Bhat, 2011) as an analytical approximation of the loglikelihood that is well behaved (even with numerical approximations of the Hessian), avoiding thus the non-convergence problems and dimensionality issues of the standard frequentist estimator. The method of Bhat and Dubey (2014) not only is able to handle a probit kernel and a combination of continuous and discrete indicators but also converges in minutes for problems with 500-2,000 observations, whereas restricted specifications take 15 hours or more with the standard frequentist estimator. The authors note, however, that larger sample sizes are

[1]Each latent variable adds one dimension to the integral of the joint choice probability.

required to best recover the effects of the latent variables on choice.

The purpose of this paper is to explore another estimator that avoids the curse of dimensionality, and addresses other issues such as having exact (small-sample) properties. The main contribution is thus the derivation of a general, simultaneous Bayes estimator for a multinomial probit model with a panel structure and latent attributes that are endogenous and manifested through effect indicators that are discrete, continuous, or both. Our estimator allows for interaction and simultaneity among the latent attributes, residual correlation, nonlinear effects on the utility function, flexible substitution patterns, and temporal correlation within responses of the same individual. We effectively propose to model choice as a covariance structure model with an augmented space of discrete and continuous dependent variables, and identification blocks that are exploited to derive the full conditional distributions for Gibbs sampling the posterior of interest.[2] There are several benefits in the estimator proposed. As discussed in the paper, estimation time is in the order of minutes (1-3 minutes for 500 observations and 10,000 repetitions of the sampler, 5-15 minutes for 2,500 observations); the estimator is integral, gradient, and Hessian free; and inference on transformation of the parameters of interest is eased, through the possibility of post-processing Monte Carlo Markov chains to find posterior distributions and standard errors of welfare measures (willingness to pay, consumer surplus), underlying discount rates, and predicted probabilities and shares.

After analyzing the general behavior of the estimator using a Monte Carlo study, we give an empirical application with important insights that are relevant for better understanding travel behavior. By constructing a model of vehicle purchase and commuting behavior, we generalize previous findings (Bolduc et al., 2008, Bolduc and Daziano, 2010, Daziano and Bolduc, 2013b) about urban transportation choices. In particular, we present the structural discrete choice model as an alternative approach for deriving a continuous, latent market segmentation of consumers. We also provide inference on the **energy paradox** or energy efficiency gap in vehicle fuel efficiency, which aims at explaining the observed slow consumer shift to energy efficient technologies with high-return rates (Jaffe and Stavins, 1994). In particular, we derive implicit discount rates (Hausman, 1979, Train, 1985) that allow for heterogeneity based on a latent variable that identifies cost-conscious consumers.

The rest of the paper is organized as follows. In section 2 we specify both the structural and measurement equations of a generalized structural discrete choice model with a multinomial probit kernel and latent attributes that are manifested by effect indicators that can be either continuous or discrete. We also discuss identification of the parameters

[2] Unlike the estimator analyzed in Daziano and Bolduc (2013b), no Metropolis-Hastings simulation is required for the estimator derived in this paper. Other extensions include simultaneity and interactions, which are both challenging in the Bayesian context.

of the model, and we derive a Gibbs sampler based on reduced form of the model. In fact, we discuss that when introducing interactions, the Gibbs sampler is based on a pseudo-reduced form that requires special attention to take into account stochasticity in the parameters of the full conditional distributions. Section 2 ends with a Monte Carlo study that analyzes behavior of the estimator for a varying number of alternatives (5-10), alternative-specific latent variables, and sample sizes (500; 1,500; 2,500). In section 3 we present the discrete-choice experiment about transportation choices – vehicle-purchase and commuting-mode choices – in Canadian urban centers. Even though we have used a subset of the same dataset in previous work, in this paper we overcome a series of simplifying assumptions that were originally used, and are actually still present in most current work on latent attributes in discrete choice. Furthermore, the commuting mode choice experiment is an addition, as our previous work has focused on specific models of vehicle choice. Section 4 summarizes posterior estimates of the joint model, including a forecasting exercise, and inference on implicit discount rates when comparing upfront costs versus future energy savings. Section 5 concludes by summarizing the main findings of this study.

2. Microeconometric choice model with endogenous attributes

The statistical model representing random utility maximization behavior treats utility as a latent endogenous variable. The problem of latent endogenous variables has led to specific econometric models of qualitative dependent variables, including discrete choice. Standard discrete choice can be seen as a special case of structural equation modeling (SEM) – a class of statistical models common in psychometrics.[3] SEM views the relationship between latent variables (such as utility) and manifest variables or effect indicators (such as choice indicators) as a system of simultaneous equations. Two main sub-models can be distinguished in 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 observable manifest variables.

2.1. Structural choice model

In a standard discrete choice setting all attributes are observable and exogenous. However, we consider that the attributes are partitioned into a set of observable and exogenous attributes and a set of attributes that are not only latent but also determined within the model (see Rungie et al., 2012). The structural model of a multinomial probit with endogenous latent explanatory variables is given by the system of equations we describe in this subsection. The model that we analyze fits the promising avenue of research of

[3]SEM is based on covariance structure analysis.

expanding the explanatory factors included in discrete choice models to psychological constructs such as attitudes (Rungie et al., 2011). McFadden (1986) and Ben-Akiva and Boccara (1987) set the theoretical fundamentals for later development of the comprehensive structural choice modeling framework that we adopt here. These fundamentals were revisited by Walker (2001) and Ben-Akiva et al. (2001) in a seminal work that has motivated the reemergence of what has been called hybrid choice modeling (HCM) by some authors in transportation research (Ben-Akiva et al., 2002).[4]

The system combines a discrete choice kernel with a standard SEM for the latent attributes. There are several relevant expansions in the discrete choice kernel that we analyze in this paper (cf. Bhat and Dubey, 2014). First, the discrete choice kernel is a multinomial probit model with a full covariance matrix that allows for flexible substitution patterns that are determined by the data. Second, we assume a panel structure that accounts for repeated observations that are typical of stated preference data (cf. Elrod and Keane, 1995). In addition, for the latent attributes we adopt a generalization of a Multiple Indicator Multiple Cause (MIMIC) model (Jöreskog and Goldberger, 1975), which is a sub-model of the more general linear structural relations (LISREL) or JKW system of Jöreskog (1973), Keesling (1972), Wiley (1973). The generalization comes from considering effect indicators that can be discrete, continuous, or both. Finally, we allow for interactions not only among the latent factors (simultaneity in the determination of the endogenous latent factors), but also among the latent factors and the observable attributes. Interactions are relevant for a more general representation of the discrete demand system, where latent factors are used to construct a mechanism of continuous market segmentation. This way of representing latent, continuous consumer heterogeneity distributions may be especially appealing for marketing and empirical industrial organization. We will illustrate this consumer-heterogeneity mechanism in the empirical analysis of the next section.

Consider the following simultaneous system of latent variables (we adopt the notation of Bolduc et al. (2005) and Bolduc and Daziano (2010), which is summarized in Table B.1):

Structural equations

$$\underset{(L\times 1)}{\mathbf{z}_n^*} = \underset{(L\times L)}{\mathbf{\Pi}} \mathbf{z}_n^* + \underset{(L\times M)}{\mathbf{B}} \underset{(M\times 1)}{\mathbf{w}_n} + \underset{(L\times 1)}{\boldsymbol{\zeta}_n}, \boldsymbol{\zeta}_n \sim \mathcal{N}(0, \mathbf{H}_\Psi^{-1}) \tag{1}$$

$$\underset{(J\times 1)}{\mathbf{U}_{tn}^*} = \underset{(J\times K)}{\mathbf{X}_{tn}} \underset{(K\times 1)}{\boldsymbol{\beta}} + \underset{(J\times Q)}{\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \mathbf{z}_n^*)} \underset{(Q\times 1)}{\boldsymbol{\varrho}} + \underset{(J\times L)}{\mathbf{\Gamma}} \underset{(L\times 1)}{\mathbf{z}_n^*} + \underset{(J\times 1)}{\boldsymbol{\nu}_{tn}}, \boldsymbol{\nu}_{tn} \sim \mathcal{N}(0, \mathbf{H}_\Sigma^{-1}) \tag{2}$$

$$\underset{(R\times 1)}{\mathbf{I}_n^*} = \underset{(R\times 1)}{\boldsymbol{\alpha}} + \underset{(R\times L)}{\mathbf{\Lambda}} \underset{(L\times 1)}{\mathbf{z}_n^*} + \underset{(R\times 1)}{\boldsymbol{\varepsilon}_n}, \boldsymbol{\varepsilon}_n \sim \mathcal{N}(0, \mathbf{H}_\Theta^{-1}) \tag{3}$$

[4]The model is also known as the Integrated Choice and Latent Variable (ICLV) model.

Measurement equations

$$\underset{(1\times1)}{I_{rn}} = \begin{cases} 1 & \text{if } \mu_{0r} < I^*_{rn} \le \mu_{1r} \\ 2 & \text{if } \mu_{1r} < I^*_{rn} \le \mu_{2r} \\ \vdots & \\ M_r & \text{if } \mu_{M_r-1} < I^*_{rn} \le \mu_{M_r}, \end{cases} \tag{4}$$

$$\underset{(1\times1)}{y_{tn}} = i \in C_n \text{ iff } U_{itn} - U_{jtn} \ge 0, \forall j \in C_n, j \ne i, \forall n \in N. \tag{5}$$

where $\mathbf{z}^*_n$ is an endogenous random vector of individual-specific latent variables that enters the utility function as a latent explanatory variable; the matrix $\mathbf{\Pi}$ allows for the eventual presence of simultaneity or interactions among the latent variables[5] – we assume that $(\mathbf{1}_L - \mathbf{\Pi})$ is invertible, where $\mathbf{1}_L$ represents the identity matrix of size L; $\mathbf{w}_n$ is a vector of *"causal indicators"* or explanatory variables affecting the latent variables; $\mathbf{B}$ is a matrix of unknown regression coefficients used to describe the global effect of $(\mathbf{1}_L - \mathbf{\Pi})^{-1}\mathbf{B}\mathbf{w}_n$ on the latent variables; and $\mathbf{H}^{-1}_{\Psi}$ is a full covariance matrix which describes the relationship among the latent variables through the error term.[6] To simplify notation, we define the following reduced-form parameters $\tilde{\mathbf{B}} = (\mathbf{1}_L - \mathbf{\Pi})^{-1}\mathbf{B}$, $\tilde{\boldsymbol{\zeta}}_n = (\mathbf{1}_L - \mathbf{\Pi})^{-1}\boldsymbol{\zeta}_n$, and $\tilde{\mathbf{H}}^{-1}_{\tilde{\Psi}} = [(\mathbf{1}_L - \mathbf{\Pi})^{-1}]\mathbf{H}^{-1}_{\Psi}[(\mathbf{1}_L - \mathbf{\Pi})^{-1}]'$.

The choice model in equation (2) is written in vector form where we assume that there is a total of J_n available alternatives in the set C_n, as well as T choice situations. Hence, $\mathbf{U}_{tn}$ is a vector of indirect utility functions for individual n and choice situation t; $\mathbf{X}_{tn}$ is a design matrix with $\mathbf{x}'_{tin}$ designating its ith row; and $\boldsymbol{\beta}$ is a vector of unknown taste parameters. $\mathbf{Y}^*_{tn}(\mathbf{X}_{tn}, \mathbf{z}^*_n)$ is a matrix of Q interactions between the observable attributes $\mathbf{X}_{tn}$ and the latent $\mathbf{z}^*_n$ as well as interactions within the latent variables; $\boldsymbol{\varrho}$ is a vector of unknown parameters associated with these interactions. $\mathbf{\Gamma}$ is a matrix of unknown parameters associated with the latent variables, with $\boldsymbol{\gamma}'_i$ designating the ith row of matrix $\mathbf{\Gamma}$.[7] The choice model is completed with equation (5) which contains the choice indicators $y_{tn}, \forall t, n$ that manifest the utility maximization process of consumers. Because of the normality assumptions regarding the distribution of the random term $\boldsymbol{\nu}_{nt}$ the choice kernel of the system is a panel probit model.

Equations (3) and (4) represent a system of independent ordered probit models for measurement of the latent variables $\mathbf{z}^*_n$. Equation (3) is the structural equation of an underlying continuous vector of indicators. Thus, $\mathbf{I}^*_n$ is a (latent) continuous vector of manifestations of the latent variables $\mathbf{z}^*_n$; $\boldsymbol{\alpha}$ is an intercept vector and $\mathbf{\Lambda}$ is a matrix

[5] $\mathbf{\Pi}$ contains zeros in the diagonal.

[6] Current applications of discrete choice models with latent attributes impose a diagonal matrix. We generalize the model to allow for correlated latent variables.

[7] Whereas $\mathbf{\Gamma}$ represents the standard linear effect of the latent variables on the utility function that is common in hybrid choice models, we also allow for nonlinearities through the interactions $\mathbf{Y}^*_{tn}(\mathbf{X}_{tn}, \mathbf{z}^*_n)$.

of unknown factor loadings. $\boldsymbol{\varepsilon}_n$ is a vector of measurement error terms with covariance matrix $\mathbf{H}_\Theta^{-1}$. We assume that there are R measurement elements, i.e. a total of R *"effect indicators"* (usually just labeled as "indicators"). For deriving the estimator we assume, first, that each observable indicator $I_{rn}, r \in 1, .., R$ is a categorical variable that can take M_r multinomial, ordinal values. For the r-th manifest variable, instead of observing the underlying continuous measurement I^*_{rn}, the sample contains the discrete categories of response (for example, answers in a Likert scale.) I_{rn} is therefore a censored version of $I^*_{rn} = \alpha_r + \boldsymbol{\lambda}'_r \mathbf{z}^*_n + \varepsilon_{rn} > 0$.[8] $\boldsymbol{\mu}_r = (\mu_{0r}, \ldots, \mu_{M_r})'$ is a vector of threshold parameters that determine the censorship mechanism. Although we formulate the model with all effect indicators being ordered, it is straightforward to represent a situation with a mixture of binary, ordered, and continuous effect indicators. For instance, if indicator r is dichotomous, then the measurement equation (4) becomes $I_{rn} = 1$ if $I^*_{rn} > 0$ and $I_{rn} = 0$ if $I_{rn} \leq 0$. If indicator r is continuous, then measurement equation (4) for that indicator is not necessary, as I^*_{rn} is directly manifested.

Note that the model of interest is a simultaneous equation system with latent variables representing both preferences and endogenous underlying attributes, with ordinal effect indicators for the underlying attributes. In fact, for the particular case analyzed in this paper what we obtain is a simultaneous system of probit models. However, the derivation of a joint estimator for the parameters of the system is challenging. We will denote by $\boldsymbol{\delta}$ the whole set of unknown parameters of the hybrid choice model. Given our assumptions, the likelihood of observing both $\mathbf{y}_n = (y_{1n}, ..., y_{Tn})'$ and $\mathbf{I}_n = (I_{1n}, ..., I_{Rn})'$ may thus be written as:

$$\ell(\mathbf{y}, \mathbf{I}; \boldsymbol{\delta}) = \prod_{n=1}^{N} \int_{z^*_n} \prod_{t=1}^{T} P_{tn}(i_{tn} \left| \mathbf{z}^*_n, \mathbf{X}_n, \boldsymbol{\theta}, \mathbf{H}_\Sigma^{-1} \right.) \prod_{r=1}^{R} f(I_{rn} | \mathbf{z}^*_n, \boldsymbol{\Lambda}, \boldsymbol{\mu}_r, \mathbf{H}_\Theta^{-1}) g(\mathbf{z}^*_n \left| \mathbf{w}_n, \tilde{\mathbf{B}}, \mathbf{H}_{\tilde{\Psi}}^{-1} \right.) dz^*_n, \tag{6}$$

where $P_{tn}(i_{tn} \left| \mathbf{z}^*_n, \mathbf{X}_n, \boldsymbol{\theta}, \mathbf{H}_\Sigma^{-1} \right.)$ is the probability of the chosen alternative in choice situation t, which is given by the choice probability of a multinomial probit, with $\boldsymbol{\theta}$ being a vector that summarizes the parameters of the utility function; where

$$f(I_{rn} = m) = \Phi\left(\frac{\mu_{m r} - \boldsymbol{\lambda}'_r \mathbf{z}^*_n}{[\mathbf{H}_\Theta^{-1}]_{rr}}\right) - \Phi\left(\frac{\mu_{m-1 r} - \boldsymbol{\lambda}'_r \mathbf{z}^*_n}{[\mathbf{H}_\Theta^{-1}]_{rr}}\right) \tag{7}$$

with Φ being the cumulative distribution function (cdf) of a standard normal distribution, and $[\mathbf{H}_\Theta^{-1}]_{rr}$ being the r-th element of the diagonal of $\mathbf{H}_\Theta^{-1}$;[9] and where $g(\mathbf{z}^*_n | \mathbf{w}_n, \tilde{\mathbf{B}}, \mathbf{H}_{\tilde{\Psi}}^{-1}) \sim \mathcal{N}((\mathbf{1}_L - \boldsymbol{\Pi})^{-1} \mathbf{B} \mathbf{w}_n, [(\mathbf{1}_L - \boldsymbol{\Pi})^{-1}] \mathbf{H}_\Psi^{-1} [(\mathbf{1}_L - \boldsymbol{\Pi})^{-1}]')$.

From the likelihood function it is clear that the latent attributes $\mathbf{z}^*_n$, which we assume

[8] I^*_{rn} is the r-th element of $\mathbf{I}^*_n$. $\boldsymbol{\lambda}'_r$ is the r-th row of matrix $\boldsymbol{\Lambda}$.

[9] This is the contribution to the likelihood function of one observation of an ordered probit.

to be individual-specific variables, are the source of intra-respondent correlation. So, in our model, choices coming from a same individual are correlated.

As stated above, the system can be rewritten to accommodate dichotomous and continuous manifest variables. On the one hand, if I_{rn} is dichotomous, then the measurement equation becomes

$$\underset{(1\times 1)}{I_{rn}} = \mathbb{I}_{[I^*_{rn}>0]}, \forall r, n, \tag{8}$$

and the density of the dichotomous effect indicator is the following

$$f(I_{rn}) = \Phi\left(\frac{\alpha_r + \boldsymbol{\lambda}'_r \mathbf{z}^*_n}{[\mathbf{H}^{-1}_{\Theta}]_{rr}}\right)^{I_{rn}} \left(1 - \Phi\left(\frac{\alpha_r + \boldsymbol{\lambda}'_r \mathbf{z}^*_n}{[\mathbf{H}^{-1}_{\Theta}]_{rr}}\right)\right)^{(1-I_{rn})}. \tag{9}$$

On the other hand, an observable continuous effect indicator $\mathbf{I}^*_n = \mathbf{I}_n$ converts equation (3) into a measurement equation. In this case,

$$f(I_{rn}) = \frac{1}{[\mathbf{H}^{-1}_{\Theta}]_{rr}} \phi\left(\frac{I_{rn} - \alpha_r - \boldsymbol{\lambda}'_r \mathbf{z}^*_n}{[\mathbf{H}^{-1}_{\Theta}]_{rr}}\right), \tag{10}$$

where ϕ is the probability density function (pdf) of a standard normal distribution.

To derive a frequentist maximum likelihood estimator of $\boldsymbol{\delta}$ we would need to find an analytical solution to the problem $\hat{\boldsymbol{\delta}} = \arg\max \ell(\boldsymbol{\delta}; \mathbf{y}, \mathbf{I}|\mathbf{X}, \mathbf{w})$. However, the joint choice probability does not have a closed form and simulation would be required. Note that the solution just for the multinomial probit kernel $P_{tn}(i_{tn} \left| \mathbf{z}^*_n, \mathbf{X}_n, \boldsymbol{\theta}, \mathbf{H}^{-1}_{\Sigma}\right.)$, which is a $J_n - 1$-dimensional integral without a closed form, is computationally very expensive (Bolduc, 1993, Geweke et al., 1994). Deriving a maximum simulated likelihood solution for equation (6) is even more complex, and requires averaging discrete choice probabilities calculated using the GHK simulator (Geweke et al., 1994, Hajivassiliou and McFadden, 1998, Keane, 1994) at every step of the maximization process. Because of the complexity of the standard maximum simulated likelihood estimator (cf. Bhat and Dubey, 2014, Bhat, 2011), we propose a Bayes estimator of $\boldsymbol{\delta}$ (Hastings, 1970, Geweke, 1989, Albert and Chib, 1993).

2.2. Pseudo-reduced form model

Consider the following partition of $\boldsymbol{\delta}$:[10] the taste parameters of the utility function $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\varrho}', \boldsymbol{\gamma}')'$, the parameters associated with the covariance structure of $\mathbf{H}^{-1}_{\Sigma}$, the

[10]The definition of the parameters as vectors is presented in equations A.1, A.2, and A.3 in Appendix A.

parameters of the structural equation $\tilde{\mathbf{b}}$ and the elements in $\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}$, and the measurement equation parameters $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}$ or $\boldsymbol{\mu}$.[11] Bayes estimation requires making draws from the following joint posterior distribution distribution:

$$P(\boldsymbol{\theta}, \tilde{\mathbf{b}}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{H}_{\Sigma}^{-1}, \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} | \mathbf{y}, \mathbf{I}). \tag{11}$$

Even though the latent variables are unobservable by definition, Bayesian estimation allows one to *augment* the observed data by simulating the random latent variables through Markov chain Monte Carlo (MCMC) methods. With the parameter set augmented by the latent $\mathbf{U}_n^*$, $\mathbf{z}_n^*$, and $\mathbf{I}_n^*$ the posterior of interest becomes:

$$P(\mathbf{U}^*, \mathbf{z}^*, \mathbf{I}^*, \boldsymbol{\theta}, \tilde{\mathbf{b}}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{H}_{\Sigma}^{-1}, \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} | \mathbf{y}, \mathbf{I}). \tag{12}$$

As we show below, for estimation of the posterior of equation (12) it is easier to exploit the natural conditional independence structure of the unknown parameters. The system of reduced form equations is essential to derive the conditional structure that is needed for approximating the posterior of interest. In effect, the system of structural and measurement equations (1)-(5) can be written as:

$$\begin{bmatrix} \mathbf{z}_n^* \\ \mathbf{I}_n^* \\ \mathbf{U}_{tn}^* \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{B}}\mathbf{w}_n \\ \boldsymbol{\alpha} + \boldsymbol{\Lambda}\tilde{\mathbf{B}}\mathbf{w}_n \\ \boldsymbol{\Gamma}\tilde{\mathbf{B}}\mathbf{w}_n + \mathbf{X}_{tn}\boldsymbol{\beta} + \mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n)\boldsymbol{\varrho} \end{bmatrix} + \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Lambda} & \mathbf{1} & \mathbf{0} \\ \boldsymbol{\Gamma} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\zeta}}_n \\ \boldsymbol{\varepsilon}_n \\ \boldsymbol{\nu}_{tn} \end{bmatrix} \tag{13}$$

Taking advantage of the fact that each error term is assumed to be normally distributed, and considering the identification restrictions discussed in subsection 2.4, one can show that the reduced form of the system has the following multivariate distribution ($\mathbf{1}_R$ is the identity matrix of size R):

$$\begin{bmatrix} \mathbf{z}_n^* \\ \mathbf{I}_n^* \\ \mathbf{U}_{tn}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{z}_n^*} \\ \boldsymbol{\mu}_{\mathbf{I}_n^*} \\ \boldsymbol{\mu}_{\mathbf{U}_{tn}^*} \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} & \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' & \boldsymbol{\sigma}'_{\mathbf{U}_{tn}^*, \mathbf{z}_n^*} \\ \boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} & \boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathbf{1}_R & \boldsymbol{\sigma}'_{\mathbf{U}_{tn}^*, \mathbf{I}_n^*} \\ \boldsymbol{\sigma}_{\mathbf{U}_{tn}^*, \mathbf{z}_n^*} & \boldsymbol{\sigma}_{\mathbf{U}_{tn}^*, \mathbf{I}_n^*} & \boldsymbol{\sigma}_{\mathbf{U}_{tn}^*} \end{bmatrix} \right), \tag{14}$$

[11] As discussed in the subsection about identification, both $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ cannot be jointly identified. Thus, $\alpha_r = 0$ for multinomial ordered effect indicators.

where the parameters are

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{z}_n^*} &= \tilde{\mathbf{B}}\mathbf{w}_n \\
\boldsymbol{\mu}_{\mathbf{I}_n^*} &= \boldsymbol{\alpha} + \boldsymbol{\Lambda}\tilde{\mathbf{B}}\mathbf{w}_n \\
\boldsymbol{\mu}_{\mathbf{U}_{tn}^*} &= \boldsymbol{\Gamma}\tilde{\mathbf{B}}\mathbf{w}_n + \mathbf{X}_{tn}\boldsymbol{\beta} + \mathbb{E}\left[\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n)|\mathbf{X}_{tn}, \mathbf{w}_n\right]\boldsymbol{\varrho} \\
\boldsymbol{\sigma}_{\mathbf{U}_{tn}^*,\mathbf{z}_n^*} &= \boldsymbol{\Gamma}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} + \mathrm{Cov}\left[\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n)\boldsymbol{\varrho}, \tilde{\boldsymbol{\zeta}}_n|\mathbf{X}_{tn}, \mathbf{w}_n\right] \\
\boldsymbol{\sigma}_{\mathbf{U}_{tn}^*,\mathbf{I}_n^*} &= \boldsymbol{\Gamma}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathrm{Cov}\left[\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n)\boldsymbol{\varrho}, \boldsymbol{\Lambda}\tilde{\boldsymbol{\zeta}}_n|\mathbf{X}_{tn}, \mathbf{w}_n\right] \\
\boldsymbol{\sigma}_{\mathbf{U}_{tn}^*} &= \boldsymbol{\Gamma}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Gamma}' + \mathbf{H}_{\Sigma}^{-1} + \boldsymbol{\varrho}\mathrm{Var}\left[\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n)|\mathbf{X}_{tn}, \mathbf{w}_n\right]\boldsymbol{\varrho}'.
\end{aligned}$$

Note that the derivation of the parameters of the multivariate distribution above faces the challenge of equation (13) being a pseudo-reduced form, due to the fact that the latent attributes $\mathbf{z}^*$ are embedded in the matrix $\mathbf{Y}^*$. Consequently, it is not possible to find a full reduced form. Replacing $\mathbf{z}^*$ with $\mathbf{B}\mathbf{w}_n + \tilde{\boldsymbol{\zeta}}_n$ partly solves the problem, as it introduces stochasticity that cannot be directly added to the error term of equation (13). Our solution is the derivation and use of the expectations and covariance terms that appear in equation (14), and that are relevant for the derivation of the correct conditional distributions that enter the Gibbs sampler.

It is also possible to show that

$$\begin{aligned}
\pi(\mathbf{z}_n^*|\mathbf{I}_n^*) &\sim \mathcal{N}\left(\mathbb{E}(\mathbf{z}_n^*|\mathbf{I}_n^*), \mathrm{Var}(\mathbf{z}_n^*|\mathbf{I}_n^*)\right) &(15)\\
\pi(\mathbf{U}_{tn}^*|\mathbf{I}_n^*) &\sim \mathcal{N}\left(\mathbb{E}(\mathbf{U}_{tn}^*|\mathbf{I}_n^*), \mathrm{Var}(\mathbf{U}_{tn}^*|\mathbf{I}_n^*)\right), &(16)
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}(\mathbf{z}_n^*|\mathbf{I}_n^*) &= \tilde{\mathbf{B}}\mathbf{w}_n + \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}'\left(\boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathbf{1}_R\right)^{-1}\left(\mathbf{I}_n^* - (\boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{B}\mathbf{w}_n)\right) &(17)\\
\mathbb{E}(\mathbf{U}_{tn}^*|\mathbf{I}_n^*) &= \boldsymbol{\mu}_{\mathbf{U}_{tn}^*} + \boldsymbol{\sigma}'_{\mathbf{U}_{tn}^*,\mathbf{I}_n^*}\left(\boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathbf{1}_R\right)^{-1}\left(\mathbf{I}_n^* - (\boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{B}\mathbf{w}_n)\right), &(18)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(\mathbf{z}_n^*|\mathbf{I}_n^*) &= \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} - \tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}'\left(\boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathbf{1}_R\right)^{-1}\boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1} &(19)\\
\mathrm{Var}(\mathbf{U}_{tn}^*|\mathbf{I}_n^*) &= \boldsymbol{\sigma}_{\mathbf{U}_{tn}^*} - \boldsymbol{\sigma}'_{\mathbf{U}_{tn}^*,\mathbf{I}_n^*}\left(\boldsymbol{\Lambda}\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}\boldsymbol{\Lambda}' + \mathbf{1}_R\right)^{-1}\boldsymbol{\sigma}_{\mathbf{U}_{tn}^*,\mathbf{I}_n^*}. &(20)
\end{aligned}$$

The conditional distributions of equations (15) and (16) – and their parameters as found in equations (17-20) – are essential for deriving a Bayes estimator of the parameters of the model. In Appendix A we show the derived closed-form expressions for the full conditional distributions that allow us to exploit Gibbs sampling (Geman and Geman, 1984) for deriving the desired Bayes estimator.

2.3. Implementation of Gibbs sampling

Iteration g of the Gibbs sampler for our model of interest is summarized as follows:[12]

1. Start with $\boldsymbol{\delta}^{(g-1)}$, the values found at the previous iteration.
2. Conditional on $\boldsymbol{\Lambda}^{(g-1)}$, $\mathbf{H}_{\tilde{\Psi}_N}^{(g-1)}$, $\mathbf{b}^{(g-1)}$, and $\mathbf{I}_n^{*(g-1)}$, and for every individual n, sample a new value $\mathbf{z}_n^{*(g)}$ for the latent attributes from the distribution $\pi(\mathbf{z}_n^*|\mathbf{I}_n^*) \sim \mathcal{N}\left(\mathbb{E}(\mathbf{z}_n^*|\mathbf{I}_n^*), \text{Var}(\mathbf{z}_n^*|\mathbf{I}_n^*)\right)$, where $\mathbb{E}(\mathbf{z}_n^*|\mathbf{I}_n^*)$ is defined in equation (17) and $\text{Var}(\mathbf{z}_n^*|\mathbf{I}_n^*)$ is defined in equation (19).
3. Given $\mathbf{z}_n^{*(g)}$ and $\mathbf{H}_{\tilde{\Psi}_N}^{(g-1)}$, update the values of the parameters $\mathbf{B}$ by sampling $\mathbf{b}^{(g)} \sim \mathcal{N}\left((\check{\mathbf{V}}_{\mathbf{b}}^{-1} + \mathbf{W}'\mathbf{H}_{\tilde{\Psi}_N}\mathbf{W})^{-1}(\check{\mathbf{V}}_{\mathbf{b}}^{-1} + \mathbf{W}'\mathbf{H}_{\tilde{\Psi}_N}\mathbf{Z}^*), (\check{\mathbf{V}}_{\mathbf{b}}^{-1} + \mathbf{W}'\mathbf{H}_{\tilde{\Psi}_N}\mathbf{W})^{-1}\right)$.
4. Given $\mathbf{z}_n^{*(g)}$ and $\mathbf{I}_n^{*(g-1)}$, update the values of the parameters $\boldsymbol{\Lambda}$ by sampling $\boldsymbol{\lambda}^{(g)} \sim \mathcal{N}\left((\check{\mathbf{V}}_{\boldsymbol{\lambda}}^{-1} + \mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}(\check{\mathbf{V}}_{\boldsymbol{\lambda}}^{-1} + \mathbf{Z}^{*\prime}\mathbf{I}^*), (\check{\mathbf{V}}_{\boldsymbol{\lambda}}^{-1} + \mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\right)$.
5. Update the covariance matrix of the structural equation by sampling $\mathbf{H}_{\tilde{\Psi}_N}^{(g)} \sim W(\check{\nu}_{\tilde{\Psi}} + N, \check{\mathbf{H}}_{\tilde{\Psi}}^{-1} + \sum_{n=1}^{N} \tilde{\boldsymbol{\zeta}}_n\tilde{\boldsymbol{\zeta}}_n')$.
6. For every continuous effect indicator r and for all n, $I_{rn}^* = I_{rn}$.
7. For every binary effect indicator r and for all n, update the unobserved continuous variable I_{rn}^* by sampling from the following truncated normal distributions: $I_{rn}^{*(g)} \sim \mathcal{TN}_{(0,\infty)}(\alpha_{rn} + \boldsymbol{\lambda}_r'\mathbf{z}_n^*, 1)$ if $I_{rn} = 1$, or $I_{rn}^{*(g)} \sim \mathcal{TN}_{(-\infty,0]}(\alpha_{rn} + \boldsymbol{\lambda}_r'\mathbf{z}_n^*, 1)$ if $I_{rn} = 0$.
8. For every multinomial ordered effect indicator r and for all n, update the unobserved continuous variable I_{rn}^* given I_{rn} by sampling from the following truncated normal distribution: $[I_{rn}^{*(g)}|I_{rn} = m] \sim \mathcal{N}(\boldsymbol{\lambda}_r'\mathbf{z}_n^*, 1)\mathbb{I}_{[\mu_{m-1}<I_{rn}^*\leq\mu_m]}$.
9. For every multinomial ordered effect indicator r and for all n, update the threshold parameters $\boldsymbol{\mu}_r$. If $\mu_{-m_r} = (\mu_{0_r}, \ldots, \mu_{m-1_r}, \mu_{m+1_r}, \ldots, \mu_{M_r r})$, then conditional on $I_{rn}^{*(g)}$, I_{rn},$\boldsymbol{\lambda}^{(g)}$, and $\mu_{-m_r}^{(g)}$ sample $\mu_{m_r} \sim \mathcal{U}\left(\bar{\mu}_{m-1_r}, \bar{\mu}_{m+1_r}\right)$, where

$$\begin{aligned} \bar{\mu}_{m-1_r} &= \max\left\{\max\{I_{rn}^* : I_{rn} = m\}, \mu_{m-1_r}\right\} \\ \bar{\mu}_{m+1_r} &= \min\left\{\min\{I_{rn}^* : I_{rn} = m+1\}, \mu_{m+1_r}\right\}. \end{aligned}$$

10. Conditional on $\mathbf{z}_n^{*(g)}$, $\boldsymbol{\theta}^{(g-1)}$, and $\mathbf{H}_{\Sigma}^{-1(g-1)}$, and given the choice indicators y_{tn}, update the augmented utility function in differences for every individual n and period t by sampling $\Delta_1\mathbf{U}_{tn}^{(g)} \sim \mathcal{TN}_{\Re|y_n}\left(\mathbf{X}_\Delta\boldsymbol{\theta}, \mathbf{H}_{\Sigma_\Delta}^{-1}\right)$, where $\Re$ is the truncation region defined by $\max\Delta_1 U_{itn} \leq 0$ if $y_{tn} = 1$, or by $\Delta_1 U_{itn} > \max\{0, \Delta_1 U_{-itn}\}$ if $y_{tn} > 1$.
11. Given $\Delta_1\mathbf{U}_{tn}^{(g)}$, $\mathbf{H}_{\Sigma}^{-1(g-1)}$, and $\mathbf{z}_n^{*(g)}$, update the parameters $\boldsymbol{\theta}$ by sampling $\boldsymbol{\theta}^{(g)} \sim \mathcal{N}\left((\check{\mathbf{V}}_{\boldsymbol{\theta}}^{-1} + \mathbf{X}_\Delta'\mathbf{H}_{\Sigma_\Delta}\mathbf{X}_\Delta)^{-1}(\check{\mathbf{V}}_{\boldsymbol{\theta}}^{-1} + \mathbf{X}_\Delta{}'\mathbf{H}_{\Sigma_\Delta}\Delta_1\mathbf{U}), (\check{\mathbf{V}}_{\boldsymbol{\theta}}^{-1} + \mathbf{X}_\Delta'\mathbf{H}_{\Sigma_\Delta}\mathbf{X}_\Delta)^{-1}\right)$
12. Update the covariance matrix of the utility function in differences by sampling $\mathbf{H}_{\Sigma}^{(g)} \sim W(\check{\nu}_{\mathbf{H}_{\Sigma}^{-1}} + N, \check{\mathbf{H}}_{\Sigma}^{-1} + \sum_{n=1}^{N} \Delta_1\boldsymbol{\nu}_n\boldsymbol{\nu}_n'\Delta_1' \mid_{\sigma^2_{\Delta,11}=1})$.

[12] Details are provided in Appendix A.

13. Make $g = g + 1$, and go back to step 1.

Steps 10-12 of the estimator outlined above expand on the Gibbs sampler derived by McCulloch et al. (2000) (Appendix A), which is based on ideas first used by Albert and Chib (1993). We note that we also implemented the sampler derived by Imai and van Dyk (2005) for the multinomial probit kernel. Whereas the estimator of Imai and van Dyk (2005) offers a better convergence rate, for illustrative purposes in this paper we preferred to discuss implementation of the Gibbs sampler following the work of McCulloch et al. (2000). A similar situation happens with steps 8 and 9 of the ordered probit models. For illustrative purposes we exploit the estimator of Albert and Chib (1993); however, our Gibbs sampler can be easily adapted to incorporate the more efficient estimator of Jeliazkov et al. (2008), which reverses the order of the conditionals: the threshold parameters are drawn marginally, and then the augmented latent indicators are generated, conditional on the threshold parameters (Chen and Dey, 2000). In sum, further computation gains are possible when using Imai and van Dyk (2005) combined with Jeliazkov et al. (2008).

2.4. A Note on Identification

Whereas parameter identification is well understood for both standard discrete choice models (Ben-Akiva and Lerman, 1985, Train, 2009) and standard latent variable models (Stapleton, 1978), general necessary and sufficient conditions are required for joint identification of the parameters of the structural choice model of interest. A sufficient but not necessary technique for identification is a two-step approach, where separate conditional identification rules for the discrete choice kernel and the MIMIC model are applied (Walker and Ben-Akiva, 2002). Using covariance analysis reduction, Daziano and Bolduc (2013a) show that the joint identification conditions coincide with those established by the two-step analysis. In effect, the reduced form parameters that appear in equation (14) are all identified after normalizing scale of the latent variables $\mathbf{z}_n^*$, $\mathbf{I}_n^*$, and $\mathbf{U}_{tn}^*$. Normalization is in general achieved by normalizing either a structural parameter or an element of the covariance matrix of the latent variable. For instance, normalization of scale of $\mathbf{U}_{tn}^*$ is ensured by fixing the first element of the covariance matrix of the choice model in differences (Dansie, 1985, Bunch, 1991, Bolduc, 1992). For continuous effect indicators it not necessary to assume a latent factor ($\mathbf{I}_n^* = \mathbf{I}_n$). As a result, if the measurement equations of continuous effect indicators are not correlated (i.e. $\mathbf{H}_\Theta^{-1}$ is assumed diagonal), then the covariance matrix is identified. However, discrete effect indicators require the whole diagonal to be normalized, i.e. $\mathbf{H}_\Theta^{-1} = \mathbf{1}_R$, where $\mathbf{1}_R$ is the identity matrix of size R. For multinomial ordered effect indicators, $\boldsymbol{\alpha}$ cannot be identified and must be normalized to zero. For $\mathbf{z}_n^*$, we can set to one any nonzero coefficient in each column of the matrix $\boldsymbol{\Lambda}$ (see Stapleton, 1978).

2.5. Monte Carlo study

In this subsection a Monte Carlo study is carried out to test performance of the Bayes estimator derived above. The simulation plan expands on the study of Daziano and Chiew (2012), where the Bayes estimator of a static multinomial probit model with observable variables only was analyzed, and on Daziano and Bolduc (2013a), where both the Bayes estimator and the maximum likelihood estimator of a hybrid choice model with a logit kernel were compared. We considered three sample sizes, with 500; 1,500; and 2,500 observations. The number of alternatives was varied from 5 to 10. One alternative-specific latent variable, each manifested by three indicators, was considered for each case (i.e. for a model with 7 alternatives there are 7 latent explanatory variables).

The data generating process was constructed as follows. All exogenous variables were generated first using a random number generator for a population of a pre-specified size.[13] A set of fixed values for the true parameter vector $\boldsymbol{\delta}_0$ was then considered. Appropriate elements of $\boldsymbol{\delta}_0$ ($\tilde{\boldsymbol{b}}_0$ and $\mathbf{H}^{-1}_{\tilde{\Psi},0}$) and the population causal factors $\mathbf{w}$ were used to generate the endogenous latent variables $\mathbf{z}$. The deterministic part of the latent utility function was then constructed using the marginal utilities $\boldsymbol{\beta}_0$ as well as the observable attributes and the generated latent variables.

The random utility was completed by adding multivariate normally distributed taste shocks $\boldsymbol{\nu}_{tn}$. As in Daziano and Chiew (2012), we test four covariance structures ($\mathbf{H}^{-1}_{\Sigma}$) for the taste shocks, namely an independent and identically distributed (IID) covariance matrix, an independent but heteroskedastic covariance matrix, a correlated but homoskedastic covariance matrix (a nested structure), and a full covariance matrix. For the IID structure, $\mathbf{H}^{-1}_{\Sigma} = \mathbf{1}_J$ (i.e. the identity matrix of size equal to the total number of alternatives, J). For the heteroskedastic structure,

$$\mathbf{H}^{-1}_{\Sigma} = \begin{bmatrix} 1/J & 0 & \cdots & 0 \\ 0 & 2/J & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

For the simple nested structure, alternatives 1 and 2 have a correlation of 0.5, and alternatives 3 and 4 have a correlation of 0.75. For the full-covariance structure, elements in $\mathbf{H}^{-1}_{\Sigma}$ were generated as follows. For the diagonal elements, deterministic variances equal to $\{1, 2, \ldots, J\}$ were considered. For each off-diagonal element, a correlation coefficient was randomly generated by drawing from a uniform distribution with parameters $[-0.95, 0.95]$ (see Daziano and Chiew, 2012).

[13] Three alternative-specific observable attributes were considered. These attributes were generated by drawing from a uniform distribution with parameters $[0, 1]$.

The choice indicator was built assuming a deterministic discrete utility maximization process, where the chosen alternative for a given individual is the one the maximizes the simulated utility. The effect indicators of the latent variables were generated using the true parameters $\boldsymbol{\lambda}_0$ of the measurement equation (4). For estimation purposes, only the observable attributes $\mathbf{X}$, observable regressors $\mathbf{w}$, choice indicators $\mathbf{y}$, and observable indicators $\mathbf{I}$ were kept.

Given a specific sample size, posterior estimates for the Bayes estimator were calculated for each of 100 samples using a process of repetitive subsampling without replacement from the simulated population. For each parameter in $\boldsymbol{\delta}$ and for each case in the simulation plan (for example, "sample size of 150 individuals and full covariance"), the bias, root mean squared error (RMSE), t-statistic (ratio between the mean of the point estimates and their standard error), and empirical coverage. The empirical coverage was obtained as the proportion of the time that the 95% Highest Density Posterior (HDP) credible intervals contained the true parameter of interest. We also saved the estimation time.

Convergence of the Markov chain was attained after a relatively low number of repetitions of the sampler – the posterior draws exhibited good mixing even with less than 5,000 repetitions. The posterior estimates were produced using 10,000 repetitions of the Gibbs sampler and diffuse priors (precision of 0.1) for a fair comparison with MSLE (Bayes estimators and maximum likelihood estimators coincide asymptotically in this case). Table B.3 reports the performance analytics for selected parameters (a marginal utility for an observable attribute, β_1, and two marginal utilities for latent attributes, γ_1 and γ_2). The Bayes estimator performs really well, independently of the sample size and covariance structure. The magnitudes of the bias are small, but no clear patterns are detected across sample sizes. However, in general, the bias is smallest for the IID structure and largest for the full covariance. Adopting a frequentist approach for hypothesis testing, the t-statistics indicate that the true parameters are always recovered. In fact, true parameters are contained within the bounds of the HDP intervals about 94% of the time, indicating that the empirical coverage almost coincide with the desired probability of 95%. Finally, RMSE decreases as the sample size increases. On average, RMSE decreases 35% when going from 500 to 1,500 observations, and 15% when going from 1,500 to 2,500 observations (the smallest reductions are observed for the full covariance case). Estimation is fast (Table B.2): 1-3 minutes for 500 observations and 10,000 repetitions of the sampler, 5-15 minutes for 2,500 observations.[14]

We note that the frequentist solution using the maximum simulated likelihood estimator – with numerical evaluation of the gradient – did not converge. In general, the search for the maximum stopped at around 200-250 iterations (after about 1-3 days), at which

[14]On a Mac Pro, with 3.0GHz 8-core and 64GB (4×16GB) of 1866MHz DDR3 ECC.

point flat regions of the likelihood were encountered. This situation did not change after testing several starting points. In these flat regions, the optimization algorithm fails to invert the Hessian, and no step can be defined. In fact, we even encountered convergence problems when using a two-step, limited information maximum simulated likelihood estimator. In particular, for the full covariance structure success rates were as follows: 100% for all sample sizes when the number of alternatives was 5; for 6 alternatives, convergence was achieved in 80% of the cases with 500 observations, and in 100% of the subsamples for the sample sizes of 1,500 and 2,500. With 7 alternatives, convergence was achieved in 100% of the subsamples only for 2,500 observations, decreasing to 20% for 1,000 observations and non convergence was observed for all subsamples of 500 observations. With 8 alternatives, convergence was achieved 20% of the time, only for 2,500 observations. For the rest of the cases, a singular Hessian prevented from achieving convergence. We tried a sample size of 10,000 observations for a model with full covariance. Convergence was achieved after about 5 days. For the cases that did converge, the true parameters were recovered, although RMSE was about 10-20% higher when compared to the Bayes estimates. The major difference was observed in terms of the empirical coverage. Whereas in the Bayesian case the HDP intervals effectively reproduce the desired probability, the frequentist confidence intervals did a poor job. For the smallest sample size, the average coverage of the credible intervals was only 78%, reaching 84% for the larger sample sizes. For some parameters the empirical coverage was as low as 27%, and for other parameters the width of the confidence intervals were too large producing a 100% coverage (which is not desirable).

3. The data: transportation choices in Canada

We use data from a discrete-choice-experiment survey conducted in 2002 in Canada by the Energy and Material Research Group (EMRG, Simon Fraser University). The sample consists of 866 commuters randomly drawn from households living in Canadian urban centers with populations over 250,000.[15] The average household income of the sample is $62,000 CAD. 75% of the respondents attained undergraduate degrees or completed graduate school. 59% of the respondents are women, and 59% of the sampled individuals are 41 years or older. Survey participants were first contacted in a quick telephone interview that started with appropriate filters. Results of the telephone interview were used to customize a detailed questionnaire that was then mailed to the respondents. The final questionnaire had the following five different parts: (1) Transportation options, requirements and habits; (2) Personal vehicle choice (discrete choice experiment 1); (3) Transportation mode preferences (discrete choice experiment 2); (4) Views on transportation issues; and (5) Sociodemographics. Full details regarding the survey, including

[15] 866 completed surveys out of 1150 target individuals (75% response rate).

the design of the questionnaire, the process of conducting the survey and analysis of the collected data can be found in Horne (2003).

The survey contained two discrete choice experiments. The first discrete choice experiment was on purchase intentions of alternative fuel vehicles (Horne et al., 2005), and the second one was on mode choice. Even though we have used the vehicle choice data in previous research, either with a frequentist estimator (Bolduc et al., 2008, Bolduc and Daziano, 2010) or a Bayes estimator (Daziano and Bolduc, 2013b), in this paper we address several challenges that were not solved before. As mentioned in the introduction, we now consider discrete indicators, simultaneity in the determination of the latent variables, and these simultaneous latent variables also interact with other attributes and can be correlated. We also incorporate now a larger number of latent attributes, and it is precisely this larger number one of the reasons for the convergence failure of the standard frequentist estimator that we tested before. In terms of empirical contributions, we add inference on the energy paradox. Finally, the mode choice component of the survey has not been used before in the context of modeling latent attributes.

3.1. Ultra-low-emission-vehicle discrete choice experiment

The alternatives of the choice experiment on vehicle purchases are given by four energy sources: (1) Gasoline-operated internal combustion engine vehicle (SGV); (2) Alternative-fuel vehicle (AFV); (3) Gasoline-electric hybrid vehicle (HEV); and (4) Hydrogen fuel cell vehicle (HFC). The experimental attributes are the following:

1. Purchase price (PP): capital cost of the new vehicle in 10K 2002 CAD$ [02CAD$/10,000].
2. Fuel cost (FC): monthly operating costs in 100 2002 CAD$ [02CAD$/100-month].
3. Fuel availability (FA): proportion of stations selling the proper fuel [ratio].
4. Express lane access ($Express$): indicator of whether the vehicle would be granted access to express lanes of the high-occupancy-vehicle (HOV) type.
5. Power (POW): horsepower of the engine of the new vehicle compared to the current household's vehicle [ratio].

Table B.4 presents the experimental attribute levels. Attributes were customized using the stated attribute levels answered by the individuals for their current vehicles as benchmark. The attribute levels were combined following an orthogonal design with randomized blocks of four choice situations each.

3.2. Commuting discrete choice experiment

There are five alternatives in the travel mode choice experiment: (1) Car (driving alone); (2) Carpool; (3) Transit; (4) Park & ride; and (5) Walk or cycle (active transportation). The experimental attributes are the following:

1. Travel cost (TC): average travel cost per trip in 2002 CAD$ (calculated from a monthly expense) [02 CAD$ / month].
2. Travel or driving time (TT): one-way commuting in-vehicle travel time in minutes [min].
3. Pickup & drop-off time (PDT): time spent in pickup and drop-off for carpoolers per trip in minutes [min].
4. Access time (WWT): total walking and waiting time to public transit per trip in minutes [min].
5. Transfers ($TRANS$): whether transfers are needed in public transit or not [indicator].
6. Bike path ($PATH$): availability of a bike path [indicator].

Table B.5 presents the experimental attribute levels. For both travel cost and travel time, the attribute levels were customized according to the actual travel cost (N_{Cost}) and time (N_{Time}) reported by the respondents when describing their commutes. In the case of the active transportation alternative (walk or cycle), travel time was calculated as a function of the stated commuting distance N_{Dist} and a low and high speed for each mode. The attribute levels were combined following an orthogonal design with randomized blocks of four choice situations each.

3.3. Attitudes toward transportation

The survey also included a set of attitudinal and perceptual questions. Whereas attitudinal questions are common in market research, standard choice experiments usually do not consider measurement of attitudes or perceptions. In the survey, individuals were asked to state their degree of support for different transportation policies, as well as their evaluation of the seriousness of different transportation problems.

The transportation policies, evaluated in a discrete Likert scale from one (strongly opposed) to five (strongly supportive), were the following: 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 using bike lanes and speed controls; 4) reducing vehicle emissions with regular testing and manufacturer emission 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 and residential and high-density development; and 8) reducing transportation needs by encouraging compressed work weeks and working from home.

The transportation problems, evaluated in a Likert scale from one (not a problem) to five (major problem), were the following: 1) traffic congestion you experience while driving, 2) traffic noise you hear at home, work, or school, 3) vehicle emissions, which impact local air quality, 4) accidents caused by aggressive or absent-minded drivers, 5) vehicle emissions which contribute to global warming, and 6) unsafe communities due to speeding traffic.

Another perceptual question asked the respondent to compare driving alone with carpooling, transit, walking, and cycling in terms of 1) safety while driving, 2) comfort, 3) impact on the environment, and 4) flexibility. The Likert scale for the rating was from one (much worse) to five (much better).

Respondents rated the importance of the following factors in their mode decisions: 1) cost; 2) travel time; 3) comfort; 4) flexibility; 5) safety; 6) privacy; 7) environmental impact; 8) reliability; and 9) mode availability. The Likert scale for the responses was defined from one (not at all important) to five (very important). Finally, using the same scale, individuals rated the importance of the following attributes in their vehicle purchase decisions: 1) purchase price, 2) vehicle type, 3) fuel economy, 4) horsepower, 5) safety, 6) seating capacity, 7) reliability, and 8) appearance and styling.

After an iterative procedure of dimension reduction common in psychometrics, five latent factors were identified. These factors represent the following underlying consumer segments:

1. Pro-transit consumers: individuals who favor improvements in public transportation
2. Pro-environment consumers: individuals who favor policies protecting the environment
3. Pro-safety consumers: individuals who consider safety as a relevant aspect of their travel decisions
4. Cost-conscious consumers: individuals who are more sensitive to higher prices
5. Pro-performance consumers: individuals who value power at a reasonable cost

These five dimensions were used for defining the MIMIC component of the discrete choice system with latent attributes. As mentioned above, in the MIMIC component we accounted for effect indicators that are ordinal. The latent variables are manifested through a subset of the effect indicators shown in Appendix.

4. Results

The joint model is composed of four sub-structures that interact among each other. First is the vehicle choice model, based on the discrete choice experiment described in subsection 3.2. Second is the mode choice model, which takes into consideration the discrete choice experiment of subsection 3.1. The third element of the joint model is the structural equations of the five underlying segments of consumers that were identified. The last component is the measurement equations that manifest the latent variables. For the first and second sub-structures we considered different models, with differing interactions with the latent constructs. We note that in general we used diffuse priors, with precision equal to 0.1. (A tight prior is used in the next section to discuss how robust the estimates are with respect to the prior choice.)

4.1. Posterior estimates

In both the vehicle and mode choice models, each parameter is estimated according to a probit model. The error terms of the probit kernel for the vehicle choice model were considered independent from those for the mode choice model. However, when introducing latent variables, these random unobservables introduce correlation between vehicle purchase and commuting mode decisions. The posterior estimates – mean, standard deviation, and selected quantiles including the mode – for the vehicle choice model are displayed in Table B.6. Those for the mode choice model are displayed in Table B.7.[16] The estimates reported in both tables were generated with diffuse priors (precision of 0.1). To check robustness to the prior assumptions, table B.8 reports estimates with a tight prior (precision of 100 and zero mean; other means were also tested). Most probably due to the sample size, no clear differences or patterns are detected.

The base vehicle choice model is a standard multinomial probit, without any latent attribute. We note that the base model not only includes the observable attributes, but also interactions with the sociodemographics to represent random taste variations.[17] The parameters for the two cost components, purchase price and fuel cost, both have negative signs (deterministic taste variations with income were not statistically significant). This negative marginal utility suggests that, all else held constant, a vehicle with a higher cost would be less preferred by the consumer. The opposite happens for fuel availability, access to an express lane, and power, which appear as desirable attributes with positive marginal utilities. All the parameters are significantly different from zero at the 5% credible level. This is supported by the 2.5% and 97.5% posterior quantiles which can be used as an approximation of the respective lower and upper bound of the 95% credible interval.

[16]For ease of interpretation, in the tables we omit the alternative specific constants, as well as the nuisance parameters (i.e. the elements of the covariance matrix of the error term in differences) and interactions with sociodemographics for the base model.

[17]Although exclusions restrictions are not necessary, for the models with latent variables the effect of the sociodemographics is only coming from the structural equation of the latent variables. This hypothesis matches most current work on hybrid choice modeling.

Vehicle choice model 1 adds the additive effect of the latent pro-environment and pro-safety variables. Each has alternative-specific parameters, with the internal combustion vehicle set as base. The interpretation is that the underlying concept helps to explain the unobservable random heterogeneity that was absorbed by the error term in the base model. For the alternative fuel vehicle we fail to reject the null hypothesis that pro-environmental attitudes have an effect on the likelihood of choosing this particular energy source. This result may be explained by concerns about sustainability of the production of biofuel using corn, which was questioned in the media around the time that the data was collected.[18] However, for both the hydrogen fuel cell and for the hybrid vehicles, a consumer with higher pro-environmental attitudes is more likely to choose the respective energy-efficient technology. Consistent with previous findings using the same data – in a model with just environmental concerns as a latent variable and a multinomial logit kernel – the impact of pro-environmental attitudes is higher for vehicles propelled by hydrogen. This particular result is consistent with fuel cell vehicles producing no harmful tailpipe emissions. In contrast, hybrid electric cars can be described as having very efficient internal combustion engines that produce less, but not zero, emissions. The effect of pro-safety attitudes are significantly different from zero for both hydrogen fuel cell and hybrid vehicles. Consumers who are more concerned about safety features are less likely to choose hydrogen or hybrid technologies. In particular, fuel cell vehicles are perceived as being much less safe than the other cars. This parameter thus measures the consumer fears regarding the low-ignition point of hydrogen. Regarding the safety concerns of the hybrid electric technology, the data was collected only two years after the introduction of the first hybrid models into the North American market. High voltage discharges in the case of a crash may explain some of the consumer concerns.

Vehicle choice model 2 adds to the base model two interactions. The first one is between power and the latent pro-performance variable. This interaction measures the continuous variation in the marginal utility of power explained by differences in pro-performance attitudes. The posterior mean of the interaction is positive, suggesting that consumers that care more about overall performance of the vehicle value more horsepower. Although this is the expected result, the 95% credible interval contains zero. The second interaction is between fuel cost and the latent cost-consciousness of the consumer. Cost-conscious consumers appear as being less satisfied with increases in fuel cost, which is the expected result. For this interaction, the null hypothesis of a zero parameter is rejected. We note that we also tried the interaction of the latent cost-consciousness with purchase price, but the posterior of this interaction was centered at zero with a very small posterior variance. Vehicle choice model 3 combines models 1 and 2. The same general conclusions about the parameter estimates appear.

[18] In fact, AFVs were negatively perceived in general. Everything else being equal, the choice probability of AFVs was lower than that of any other vehicle.

The base mode choice model is a standard multinomial probit without latent attributes, where the alternative-specific attributes and significant interactions with sociodemographics are considered. As expected, the signs of travel cost, travel or driving time, pickup & drop-off time, access time, and transfers are all negative. Note that each additional minute of both pickup and access time bothers the traveler around 3.17-3.33 times more than an additional minute of in-vehicle travel time (considering main effects only). This result is consistent with previous findings in the literature. The presence of a bicycle path has a positive effect on the probability of commuting via active transportation. The 2.5% and 97.5% posterior quantiles, which can be used as an approximation of the 95% credible interval, indicate that for most of the parameters it is possible to reject the null hypothesis of the single parameter being equal to zero. The exceptions are the number of transfers and the presence of a bike path.

Mode choice model 1 introduces the effect of the underlying pro-transit attitudes as an attribute with alternative-specific parameters. Note that the active-transportation mode (walk or cycle) was set as base. The results show that higher attitudes toward transit favor the probability of choosing not only transit but also of being a carpooler. This may be explained by the more efficient use of private cars when carpooling. The effect of pro-transit on being a solo driver or being a user of park & ride is negative. The latter result implies that someone with higher pro-transit attitudes will be more likely to choose a trip entirely made using public transportation than to choose a trip where only part of the ride is using transit. However, only the effect of the latent pro-transit variable on being a solo driver are statistically different from zero. In mode choice model 2, pro-transit attitudes are only included in the utility of transit. The parameter turns out to be positive and significantly different from zero. Finally, mode choice model 3 extends the previous model by introducing an interaction between travel cost and the latent cost-consciousness. The negative sign of the interaction indicates that the more cost-conscious the consumer is, the more sensitive she is to changes in travel cost.

Table B.9 summarizes the structural equation of the latent attributes. Consumer-specific characteristics were used to explain the variations in the underlying dimensions that are hypothesized as explaining the effect indicators. Pro-transit attitudes are lower for households that own a higher number of vehicles. Solo drivers exhibit on average less pro-transit attitudes, although the respective parameter is not significantly different from zero. Commuters who mostly use public transportation tend to have a more positive view of transit, supporting policies that improve the level of service of mass transit. Females also tend to have higher pro-transit attitudes; the older the individual gets, the higher the pro-transit support; and individuals with medium levels of income also exhibit higher pro-transit behavior. In the case of pro-environmental attitudes, we included the latent pro-transit variable as one of the causal indicators. The parameter is positive, indicating that individuals who favor investments in and priorities for transit also tend to have more favorable views regarding protection of the environment. In terms of the econometric modeling, the possibility of obtaining this parameter is due to the incorporation of simultaneity in equation (1).Through the latent pro-transit, underlying the determinants of pro-environmental attitudes are the pro-transit segments. However, being a female has an even more positive effect on environmental concerns. Also, having completed university studies appears as a factor that increases pro-environmental behavior. In the case of consumers that care about safety, carpoolers appear a one of the determinants, as well as bicyclists (with a surprisingly negative impact). Females are more likely to think about safety issues; and the older an individuals gets, the higher safety appears as a priority in the decisions. Cost-conscious consumers are females and those that have low or medium income levels. Consumers with higher income appear as less sensitive to cost, which is the expected result (in contrast, no income effects could be detected using the base models). Finally, the segment of pro-performance consumers is greatly explained by the same segment of those being cost-conscious. At the same time, an effect of age becomes apparent, with older consumers caring more about overall performance.

Table B.10 reports the loading factors of the measurement equation for each of the latent attributes. The effect indicator with the highest loading factor was normalized. Pro-transit behavior is measured by the support of policies that provide express lanes for public transportation modes, improve mass transit, and discourage automobile use. Pro-environmental preferences are measured by concerns about emissions that contribute to global warming as well as to deterioration of local air quality. Although to a lesser degree, pro-environmental preferences are also measured by opposition to building roads and to expanding the current infrastructure devoted to private vehicles. Pro-safety attitudes are measured by the perception of communities becoming less safe due to speeding traffic, by concerns about drivers causing accidents, and by the importance of safety features when deciding which vehicle to purchase. Cost-conscious consumers are identified as rating purchase price and fuel economy to be highly important attributes that they consider when buying a new car, as well as by the importance of the trip cost when making mode decisions. Finally, pro-performance attitudes are manifested by the importance of the reliability, fuel economy, and horsepower of a potential new vehicle for purchase.

The structural and measurement equations form a MIMIC model that was estimated jointly with the models for vehicle and travel mode choices. For the MIMIC model, standard SEM measures of goodness of fit validate the proposed structure.[19] In addition, we note that in contrast with previous research, we accounted for the ordinal nature of the effect indicators.

[19]Comparative Fit Index (CIF): 0.900; Root Mean Square Error of Approximation (RMSEA): 0.049.

4.2. Forecasting travel behavior

The marginal utilities as well as the parameters of the structural equations of the latent variables[20] describe user behavior in terms of the probability of choosing a specific transportation mode. For instance, the marginal utilities weigh the attributes and latent explanatory variables, allowing us to model the trade-offs faced by the travelers 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 baseline (base scenario), we simulate the impact on the choice probabilities (and thus on the market shares of car, carpool, transit, park & ride, and walk or cycle) of the following four hypothetical market conditions:

1. Scenario 1: Increase in travel cost of car and carpool of 25%.
2. Scenario 2: Increase in the travel cost of car and carpool of 50%.
3. Scenario 3: Increase in gasoline cost of 50%
4. Scenario 4: Increase in power of hybrids of 15%

The first 3 scenarios consider situations where traveling by car becomes less attractive. Scenarios 1 and 2 look at the mode choice impact when driving becomes more expensive (increase in fuel costs, increase in parking costs, congestion pricing, additional taxes). Scenario 3 measures the impact on vehicle choice of a direct increase in the cost of gasoline.[21] Scenario 4 represents an improvement in the technology of hybrid vehicles.

For each scenario, market shares were derived by sample enumeration, i.e. by averaging individual choice probabilities. To obtain the individual choice probabilities, for each individual in the sample and for every alternative we calculated the corresponding predictive posterior probability. Predictive posterior probabilities can be derived by Monte Carlo approximation of

$$P_{tin}(\mathbf{X}_{tn}^{(1)}, y_n) = \int_{\boldsymbol{\Delta}} \int_{\mathbf{z}^*} P_{tn}(i_{tn}|\mathbf{z}_n^*, \mathbf{X}_n^{(1)}, \boldsymbol{\theta}, \mathbf{H}_{\Sigma}^{-1}) f(I_{rn}|\mathbf{z}_n^*, \boldsymbol{\Lambda}, \boldsymbol{\mu}_r, \mathbf{H}_{\Theta}^{-1}) d\mathbf{z}_n^* p(\boldsymbol{\delta}|\mathbf{y}) d\boldsymbol{\delta}, \quad (21)$$

where $\mathbf{X}_{tn}^{(1)}$ is the attribute matrix of the conditions set by the new scenario used for forecasting, and where $p(\boldsymbol{\delta}|\mathbf{y})$ is the posterior distribution of the joint parameter vector $\boldsymbol{\delta} \in \boldsymbol{\Delta}$. Note that Monte Carlo approximation of equation (21) does not require drawing new samples for the posterior of $\boldsymbol{\delta}$, but the same chain generated by the Gibbs sampler for estimation of the model can be used.

[20] The measurement equations provide identification of the latent variables.

[21] Although estimated jointly, the vehicle and mode choice experiments were independent.

Table B.11 presents the results of the forecasting exercise. For each scenario, including the base scenario, both the posterior mean and standard deviation of the market shares are shown. In addition, the percentage change with respect to the corresponding base model estimates are calculated. For mode choice, scenarios 1 and 2 show that the probabilities of choosing the modes with the increased costs decrease. However, note that for the base model – without latent attributes – carpoolers seem to be more elastic than solo drivers. For model 3 – with latent attributes – the percentage change in the market shares of both car (driving alone) and carpool is almost identical. In this scenario we did not increase the cost of the park & ride alternative (i.e. the increased cost of driving may reflect a toll that is avoided when riding transit). Whereas model 3 predicts a 54.3% increase in the market share of park & ride for scenario 2, the model without latent variables predicts an increase of 68.4%. The differences between the impacts of model 3 and the base model respond to the fact that model 3 introduces random consumer heterogeneity in the marginal utility of cost through the latent cost-consciousness variable. In the case of scenario 3, model 3 allows for unobserved heterogeneity in the marginal utility of fuel cost. When we model the effect of an increase in gasoline cost of 50%, model 3 predicts a much higher decrease in the number of consumers buying ICVs (-32.9%) than the base model does (-24.1%). This result is explained by a higher sensitivity to cost coming for cost-conscious consumers. In addition, model 3 predicts a lower decrease in the market share of hybrids (-13.2%) than the base model (-19.4%). This can be explained by the interaction with the latent pro-environment variable. Finally, for scenario 4 the base model predicts similar competition across vehicles after an increase in the horsepower of the hybrid car, but model 3 predicts a stronger competition between HEVs and ICVs, which is a more realistic result.

Although the predictions shown above include interactions with the latent attributes, an additional exercise is to perform forecasting by actually varying the values of the latent variables themselves. We note that since the latent attributes are endogenous and because the measurement scale of the latent constructs is unknown, there is no sense in imposing a shock directly on the underlying concept (as in "increase of pro-environmental behavior of 25%").[22] In structural discrete choice models, a meaningful scenario comes from a shock in the structural equation of the latent attribute. For instance, we can forecast the impact of a latent attribute having a maximum value by considering the predictive posterior for the segment of consumers that exhibit the highest values in the underlying attribute. For example, women were determined to have higher pro-environment attitudes. In the forecasting exercise, we can predict what would happen if men are represented as having the same attitudes as women. Therefore, we define two additional scenarios:

5. Scenario 5: maximum pro-environment consumers
6. Scenario 6: maximum cost-conscious consumers

[22]Scenarios like this one have been analyzed by a few authors

Scenario 5 represents a situation were a shock in the structural equation of the latent pro-environment variable ensures that all consumers behave as the segment with higher pro-environmental attitudes. Scenario 6 introduces the same type of shock but in the structural equation that explains cost-consciousness.

Table B.12 presents the market share forecasts for scenarios 5 and 6. Only the results of model 3 are presented, and the percentage change is calculated with respect to the market shares of the base scenario. For scenario 5, i.e. individuals becoming more conscious about the environment, all commuting modes reduce their market share with the exception of transit. In the case of vehicle choice, HFCs are the only vehicles that increase their penetration in the market. For scenario 6, i.e. consumers becoming more cost conscious, transit and non-motorized modes increase their shares. For the vehicle choice case, cost-conscious consumers opt for HEVs.

4.3. Inference on the energy paradox

Consumers that are sensitive to fuel cost demand more efficient vehicles (Greene, 2010). However, the timing of the vehicle-purchase expense differs from that of fuel expenditures (intertemporal behavior). Furthermore, previous research has noted a rather systematic underestimation of consumers to account for future savings in operating costs (McManus, 2007, Fan and Rubin, 2010, Helfand and Wolverton, 2011, Allcott, 2011, Allcott and Wozny, 2012). In fact, Sallee (2012) argues that some consumers may even be inattentive to fuel costs. The slow consumer shift to energy efficient technologies with high-return rates has been called 'energy paradox' (Jaffe and Stavins, 1994) or 'energy efficiency gap' (Hirst and Brown, 1990). Understanding the undervaluation of cost-efficient energy-efficiency gains is key for better informing policies aiming at promoting consumer adoption of sustainable technologies, including ultra-low-emission vehicles. In particular – as noted by Parry et al. (2010) and Bento et al. (2012) – robust estimation of the energy paradox is critical for the evaluation of the impact of imposing tighter efficiency standards, such as the US Corporate Average Fuel Economy (CAFE) standards, versus other policies such as emission pricing or changes in gasoline taxes. If consumers are misperceiving energy cost savings, energy efficiency standards are superior to Pigouvian taxes (see Bento et al., 2012).

The difference between the marginal utility of purchase price and that of fuel cost can be used as a tool for verifying the presence of a lower elasticity to future energy costs than to out-of-pocket expenses at the time of purchase. The idea is to use the ratio of the consumer valuations of purchase price and operating costs, as derived from estimates of the discrete choice model, to calculate a measure of the energy paradox. In the vehicle choice model analyzed in this paper, fuel costs are monthly expenses. To compare changes in monthly fuel costs with the single payment at the time of purchase, a rational consumer will discount the future costs using her own time preferences. Following Hausman (1979) and Train (1985), the implicit discount rates used by consumers can be inferred from a revealed-preference mechanism that makes use of the present value of future energy costs. Assuming that the lifespan of the vehicle is large enough,[23] the implicit discount rate r can be simply estimated using the following ratio as approximation

$$r = \frac{\beta_{PP}}{\beta_{FC}} = \frac{1}{WTP_{\Delta FC}}, \tag{22}$$

where $WTP_{\Delta FC}$ is the upfront willingness to pay for a marginal improvement in fuel costs. The ratio in equation (22) represents the marginal rate of substitution between the (out-of-pocket) capital cost and discounted lifetime operating costs. For the models that include the interaction of fuel cost and the latent variable for creating segments of cost-conscious consumers, the implicit discount rate can be approximated by

$$r_n = \frac{\beta_{PP}}{\beta_{FC} + \beta_{\text{cost-consciousness}}\text{cost-consciousness}_n}, \tag{23}$$

which is an individual-specific discount rate. Any shock in the causal indicators explaining the latent cost-consciousness of the consumer will change her time preferences. For example, changes in income – measured in discrete intervals in this research – will have an impact on the energy efficiency gap for the consumer. Note that estimation of implicit discount rates results in making inference on parameter ratios. In general, ratios of the parameters of an econometric model are locally almost unidentified (Dufour, 1997). As a result, the problem of interval estimation needs robust identification methods. Because we used Bayes estimators, postprocessing the ratio for the derivation of credible intervals (Daziano and Achtnicht, 2014) addresses the problem of potential weak identification (Edwards and Allenby, 2003).

In Table B.13 we present the posterior mean, standard deviation, and median, as well as bounds for the 95% HDP intervals of the annual discount rate derived from post processing the MCMC parameter samples of both the base model without latent variables and the joint model with latent variables and interactions (model 3). In the case of model 3, because the implicit discount rate is individual-specific, we report the estimates for a representative individual with different treatments in terms of income levels. The mean discount rate for the base model without latent variables is about 27%. When the latent variables are added to the model, it is possible to account for differences in income levels due to the interaction with the latent cost-consciousness of the consumers. Note that for the model with the latent attributes, the implicit discount rate is individual-specific. Table B.12 contains the results for a randomly selected individual. For this consumer, the mean discount rate is in the range 16-18%, depending on the income level. Our results

[23]Because fuel costs were presented to respondents of the survey as a monthly expenditure, and because the expected lifetime for light duty vehicles is 14 years (Bento et al., 2013), the expected lifespan is 168 months.

from the model with latent attributes are in line with previous findings. In the context of a survey to homeowners about appliance choice decisions, Newell and Siikamäki (2013) used the experiment of Coller and Williams (1999) for elicitation of time preferences, and found a mean discount rate of 19%, with a median of 11%, and a standard deviation of 23%. Hausman (1979) obtained implicit discount rates for energy costs of air conditioners in the range of 5.1-89%, with a mean of 26.4%. For vehicle purchases, Allcott and Wozny (2012) find an implicit discount rate of roughly 15%, and Dreyfus and Viscusi (1995) provide a range of 11-17%, whereas Busse et al. (2013) find temporal preferences that are in line with market interest rates. Regarding these market interest rates, the average interest rate for used vehicle loans has been estimated at 6.9% (using information from the Surveys of Consumer Finances for 2001, 2004, and 2007, see Allcott and Wozny, 2012).[24] Additionally, Allcott and Wozny (2012) use the average real return of the S&P 500 from 1945 to 2008 to determine an estimate of 5.8% for the interest rate of the opportunity cost of vehicles paid in cash. We finally note that allowing for heterogeneity in the determination of energy-paradox measures avoids sorting issues (Hausman and Joskow, 1982, Bento et al., 2012).

5. Conclusions

In this paper we have shown that structural discrete choice models with endogenous latent explanatory variables provide a powerful tool for modeling random consumer heterogeneity. Instead of assuming a parametric heterogeneity distribution that is independent from data, structural discrete choice models use effect and causal indicators to construct latent attributes that can be used as means of introducing continuous, unobserved heterogeneity. As a result, the model is enriched because the data used to update the parameters is augmented to include attitudinal responses, while avoiding endogeneity problems. Our paper contains several contributions regarding estimation of discrete choice models with latent attributes. Previous research on simultaneous estimation of the parameters of hybrid choice models has focused on cross-sectional (conditional or mixed) logit-based kernels with continuous indicators for manifesting the latent attributes. We have built our model by assuming a multinomial probit with a full covariance matrix for full flexibility in the competition among alternatives. Our model also accounts for a panel structure through intra-respondent correlation due to individual-specific latent variables. The inclusion of latent variables thus provides a very interesting approach to the problem of repeated observations in stated preference studies.

As discussed in the paper, the standard frequentist estimator (using maximum simulated likelihood) is computationally expensive due to a poorly behaved simulated likelihood

[24]The same authors point out that the real average interest rate reported by dealerships to JD Power is 8.9%.

function. In fact, in our Monte Carlo study the maximum simulated likelihood estimator exhibited serious convergence problems due to flat regions of the likelihood. Whereas Bhat and Dubey (2014) have very recently proposed an analytical approximation of the loglikelihood that is well behaved, we have adopted a Bayes estimator instead. The main finding is that the Bayesian approach is perfectly suited to a complex hybrid choice model that considers temporal effects, correlation, heteroskedasticity, simultaneity in the determination of the latent constructs, and interactions between the latent and the observable attributes. In fact, the generalized discrete choice model with endogenous latent attributes analyzed in this paper becomes a simultaneous system of (multinomial, binary, and ordered) probit models. With the appropriate reduced form, the Bayes estimator for this system of probit equations becomes a series of normal linear regression models, avoiding the curse of dimensionality. In the standard frequentist estimator – in contrast – adding more latent variables increases the dimensionality of the integral that needs to be simulated for solving the estimator. Our estimator has properties that are valid for small samples. For instance, no clear patterns in the bias, t-statistic, and coverage were detected in the Monte Carlo study; however, RMSE did decrease with sample size. In addition, working with a Bayes estimator allows the researcher to make inference on nonlinear transformations of the parameters of interest in a straightforward fashion via postprocessing (required for the derivation of welfare measures), including the construction of credible sets that are interpreted as a region that contains the true parameters with a given probability. Our Bayes estimator converges fast; taking 1-3 minutes for 500 observations, and 5-15 minutes for 2,500 observations, even with 10 alternatives, full covariance, and 10 latent variables.

Although the estimator that we propose is an extension of the Bayes estimator of static multinomial and ordered probit models, one particular challenge that we solved in this paper is that the Gibbs sampler is actually based on a pseudo-reduced form. Equation (13) contains on its right hand side the latent attributes $\mathbf{z}^*$, which are embedded in the matrix $\mathbf{Y}^*$ that accounts for the interactions that have been generally omitted in previous work. As a result, it is not possible to find a full reduced form. Our solution is to use the expectations and covariance terms that appear in equation (14) and that are relevant for the derivation of the correct conditional distributions that enter the Gibbs sampler (equations 15-20).

We also provide an empirical application by constructing a discrete choice model of vehicle purchase and commuting behavior. This model expands on our previous research by exploiting five underlying attitudes to determine segments of pro-transit, pro-environment, pro-safety, cost-conscious, and pro-performance consumers. Interesting insights are derived from the estimates of the structural discrete choice model that includes the latent attributes. For example, cost-conscious consumers appear as having an continuous sensitivity to changes in travel and fuel costs. This pattern of valuation of changes in fuel costs are reflected in an implicit discount rate of future energy savings – which is a measure of the energy paradox – that slightly increases with income. In addition, consumers that value safety exhibit a lower probability of choosing not only hydrogen cars, but also hybrids. We note that postprocessing was used to derive standard errors of the predicted shares and implicit discount rates.

There are several avenues for further research. First, we could compare the performance of the Bayes estimator proposed here with promising frequentist variations, such as the one by Bhat and Dubey (2014). The multinomial probit kernel, although being a flexible model with unrestricted substitution patterns among the alternatives, is highly parametric. In fact, we assumed normally distributed errors for all the equations in the system as well as for prior distributions of the parameters of interest. For deriving the sampler this is a clear advantage, because normality of the error components allows us to exploit data augmentation for the latent variables in a relatively convenient way. In effect, the reduced form of the system and the conditional distributions that are derived from this reduced form are either normal or truncated normal distributions. In addition, normality allowed us to exploit natural conjugacy. For relaxing the Gaussian assumptions, future research should explore implementation of Bayesian semi- and nonparametrics. Finally, derivation and analysis of Bayes factors for statistical model comparison for the joint model is desirable.

Acknowledgements

This research is based upon work supported by the National Science Foundation Faculty Early Career Development CAREER Award No. CBET-1253475. The author also wants to acknowledge and thank Denis Bolduc for his invaluable input into this study (and a previous version of this paper using a different dataset).

References

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.

H. Allcott. Consumers' perceptions and misperceptions of energy costs. *American Economic Review*, 101(3), 2011.

H. Allcott and N. Wozny. Gasoline prices, fuel economy, and the energy paradox. Working Paper 18583, NBER, 2012.

M. Ben-Akiva and B. Boccara. Integrated framework for travel behavior analysis. (17), 1987.

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 M. Ben-Akiva, H. Meersman, and E. Van de Voorde, editors, *Freight Transport Modelling*, pages 121–141. Emerald Publisher, 2013.

A. Bento, K. Roth, and Y. Zuo. Vehicle lifetime trends and scrappage behavior in the US used car market. Working Paper, Cornell University, 2013.

A.M. Bento, S. Li, and K. Roth. Is there an energy paradox in fuel economy? A note on the role of consumer heterogeneity and sorting bias. *Economics Letters*, 115(1):44 – 48, 2012.

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

C.R. Bhat and S.K. Dubey. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, 67:68–85, 2014.

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

D. Bolduc. Maximum simulated likelihood estimation of MNP models using the GHK probability simulation with analytic derivatives. Working Paper, Département d'Économique, Université Laval, 1993.

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

M. Burda and M. Harding. Panel probit with flexible correlated effects: quantifying technology spillovers in the presence of latent heterogeneity. *Journal of Applied Econometrics*, 28(6):956–981, 2013.

M.R. Busse, C.R. Knittel, and F. Zetterlmeyer. Are consumers myopic? Evidence from new and used car purchases. *American Economic Review*, 103(1):220–257, 2013.

M.H. Chen and D.K. Dey. Bayesian analysis for correlated ordinal data models. In D. Dey, S. Ghosh, and B. Mallick, editors, *Generalized Linear Models: a Bayesian perspective*, pages 133–157. Marcel-Dekker, New York, NY, 2000.

M. Coller and M.B. Williams. Eliciting individual discount rates. *Experimental Economics*, 2(2):107–127, 1999.

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

R.A. Daziano and M. Achtnicht. Accounting for uncertainty in willingness to pay for environmental benefits. *Energy Economics*, 44:166–177, 2014.

R.A. Daziano and D. Bolduc. Covariance, identification, and finite-sample performance of the msl and bayes estimators of a logit model with latent attributes. *Transportation*, 40(3):647–670, 2013a.

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

R.A. Daziano and E. Chiew. Analyzing a probit bayes estimator for flexible covariance structures in discrete choice modeling. *Transportation Research Board*, 2012.

M. Dreyfus and K. Viscusi. Rates of time preference and consumer valuations of automobile safety and

fuel efficiency. *Journal of Law and Economics*, 38(1):79–98, 1995.
J.M. Dufour. Some impossibility theorems in econometrics, with applications to structural and dynamic models. *Econometrica*, 65:1365–1389, 1997.
Y. Edwards and G. Allenby. Multivariate analysis of multiple response data. *Journal of Marketing Research*, 40:321–334, 2003.
T. Elrod and M.P. Keane. A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, 32:1–16, 1995.
Q. Fan and J. Rubin. Two-stage hedonic price model for light-duty vehicles. *Transportation Research Record*, 2157(2):119–128, 2010.
A. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.
S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, (6):721–741, 1984.
J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57: 73–89, 1989.
J. Geweke, M. Keane, and D. Runkle. Alternative computational methods approaches to inference in the multinomial probit model. *Review of Economics and Statistics*, 76:609–632, 1994.
C.J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7:473 – 483, 1992.
D.L. Greene. How consumers value fuel economy: A literature review. Technical report, Environmental Protection Agency, 2010.
V. Hajivassiliou and D. McFadden. The method of simulated scores for the estimation of LDV models. *Econometrica*, 66:863–896, 1998.
W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.
J.A. Hausman. Individual discount rates and the purchase and utilization of energy-using durables. *Bell Journal of Economics*, 10(1):33–54, 1979.
J.A. Hausman and P.L. Joskow. Evaluating the costs and benefits of appliance efficiency standards. *American Economic Review*, 72(2):220–225, 1982.
G. Helfand and A. Wolverton. Evaluating the consumer response to fuel economy: a review of the literature. *International Review of Environmental and Resource Economics*, 5(2):103–146, 2011.
S. Hess and N. Beharry-Borg. Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in Tobago. *Environmental and Resource Economics*, 52(1): 109–131, 2012.
L. Hildebrandt, D. Temme, and M. Paulssen. Choice modeling and SEM. *Quantitative Marketing and Marketing Management*, (63-80), 2012.
E. Hirst and M. Brown. Closing the efficiency gap: barriers to the efficient use of energy. *Resources, Conservation and Recycling*, 3(4):267 – 281, 1990.
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.
K. Imai and D.A. van Dyk. A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124:311–334, 2005.
A.B. Jaffe and R.N. Stavins. The energy paradox and the diffusion of conservation technology. *Resource and Energy Economics*, 16(2):91 – 122, 1994.
I. Jeliazkov, J. Graves, and M. Kutzbach. Fitting and comparison of models for multivariate ordinal outcomes. *Advances in Econometrics: Bayesian Econometrics*, 23:115–156, 2008.
A. Fjendbo Jensen, E. Cherchi, and S. Lindhard Mabit. On the stability of preferences and attitudes before and after experiencing an electric vehicle. *Transportation Research Part D: Transport and Environment*, 25:24–32, 2013.

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.

M. Keane. A computationally practical simulation estimator for panel data. *Econometrica*, 62:95–116, 1994.

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

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

W. McManus. The link between gasoline prices and vehicle sales. *Business Economics*, 42(1):53–60, 2007.

R.G. Newell and J. Siikamäki. Nudging energy efficiency behavior: the role of information labels. Working Paper RFF DP 13-17, Resources for the Future, 2013.

A. Nobile. Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics*, 99:335–345, 2000.

D. Palma, J.D. Ortúzar, G. Casaubon, L.I. Rizzi, and E. Agosin. Measuring consumer preferences using hybrid discrete choice models. Working Paper 137, American Association of Wine Economists, 2013.

I.W.H. Parry, D. Evans, and W.E. Oates. Are energy efficiency standards justified? Working Paper RFF DP 10-59, Resources for the Future, 2010.

R.S. Rosenberger, M.D. Needham, A.T. Morzillo, and C. Moehrke. Attitudes, willingness to pay, and stated values for recreation use fees at an urban proximate forest. *Journal of Forest Economics*, 18 (4):271 – 281, 2012.

C.M. Rungie, L.V. Coote, and J.J. Louviere. Structural choice modelling: Theory and applications to combining choice experiments. *Journal of Choice Modelling*, 4(3):1–29, 2011.

C.M. Rungie, L.V. Coote, and J.J. Louviere. Latent variables in discrete choice experiments. *Journal of Choice Modelling*, 5(3):145 – 156, 2012.

J. Sallee. Rational inattention and energy efficiency. Unpublished manuscript, NBER, 2012.

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

K. Train. Discount rates in consumers' energy-related decisions: a review of the literature. *Energy*, 10: 1243–1253, 1985.

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

A. Vij and J. Walker. Hybrid choice models: the indetification problem. In S. Hess and A. Daly, editors, *Handbook of Choice Modelling*. Edward Elgar Publishing, 2014.

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.

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.

Appendix A. Gibbs sampler

When presenting the system of equations we mentioned that we were keeping the matrix notation for parameters of the latent attributes that is characteristic in structural equation modeling. This notation allows us to derive the reduced form of the model. However, when analyzing the estimator once the conditional distributions of equations (15) and (16) have been found, it is more useful the standard notation of parameters as vectors. We then rewrite the system of structural equations using the following equivalent form:

$$\underset{(J\times 1)}{\mathbf{U}^*_{tn}} = \underset{(J\times K)}{\mathbf{X}_{tn}}\underset{(K\times 1)}{\boldsymbol{\beta}} + \underset{(J\times Q)}{\mathbf{Y}^*_{tn}(\mathbf{X}_{tn},\mathbf{z}^*_n)}\underset{(Q\times 1)}{\boldsymbol{\varrho}} + \underset{(J\times P)}{\mathbf{Z}^{*s}_n}\underset{(P\times 1)}{\boldsymbol{\gamma}} + \underset{(J\times 1)}{\boldsymbol{\nu}_{tn}}, \quad \nu_{tn}\sim\mathcal{N}(0,\mathbf{H}_\Sigma^{-1}) \tag{A.1}$$

$$\underset{(L\times 1)}{\mathbf{z}^*_n} = \underset{(L\times LM)}{\mathbf{W}_n}\underset{(LM\times 1)}{\tilde{\mathbf{b}}} + \underset{(L\times 1)}{\tilde{\boldsymbol{\zeta}}_n}, \quad \tilde{\boldsymbol{\zeta}}_n\sim\mathcal{N}(0,\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}) \tag{A.2}$$

$$\underset{(R\times 1)}{\mathbf{I}^*_n} = \underset{(R\times 1)}{\boldsymbol{\alpha}} + \underset{(R\times LR)}{\mathbf{Z}^{*m}_n}\underset{(LR\times 1)}{\boldsymbol{\lambda}} + \underset{(R\times 1)}{\boldsymbol{\varepsilon}_n}, \quad \boldsymbol{\varepsilon}_n\sim\mathcal{N}(0,\mathbf{1}_R) \tag{A.3}$$

where the new matrix and vector representations of parameters and variables are defined such that $\mathbf{Z}^{*s}_n\boldsymbol{\gamma} = \boldsymbol{\Gamma}\mathbf{z}^*_n$, $\mathbf{W}_n\tilde{\mathbf{b}} = \mathbf{B}\mathbf{w}_n$, and $\mathbf{Z}^{*m}_n\boldsymbol{\lambda} = \boldsymbol{\Lambda}\mathbf{z}^*_n$. (The matrix notation for the parameters is needed to derive the covariance matrix of the pseudo-reduced form.)

Although the Gibbs sampler is performed simultaneously for the whole demand system, we present the conditional distributions separately for both the multinomial probit kernel and the MIMIC sub-models.

Appendix A.1. Conditional distributions of the multinomial probit kernel

We start the Gibbs sampler by determining the posterior simulator for the multinomial probit kernel. Note that the vector of conditional indirect utility functions is an unobservable dependent variable. However, using the notion of data augmentation allows us to treat equation (2) – or its equivalent form given by equation (A.12) – as a standard regression. It is important to mention that the use of data augmentation is straightforward only when working with a multinomial probit model (Albert and Chib, 1993, Bolduc et al., 1997, McCulloch et al., 2000, Nobile, 2000).[25]

To set location of preferences we work with utility differences with respect to an arbitrary base alternative.[26] Let $\Delta_1(\cdot)_{jn} = (\cdot)_{jn} - (\cdot)_{1n}$ be a matrix difference operator. For example,

[25]For other discrete choice kernels, such as a conditional logit or a mixed logit model, Metropolis-Hastings within Gibb sampling is needed (Daziano and Bolduc, 2013b).

[26]Without loss of generality we take the first alternative as base.

$\Delta_1 \mathbf{U}_n$ takes each element of $\mathbf{U}_n$ and subtracts the base element U_{1n} such that

$$\underset{(J-1\times 1)}{\Delta_1 \mathbf{U}_{tn}} = \Delta_1 \begin{bmatrix} U_{1tn} \\ U_{2tn} \\ \vdots \\ U_{Jtn} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} U_{1tn} \\ U_{2tn} \\ \vdots \\ U_{Jtn} \end{bmatrix} = \begin{bmatrix} U_{2tn} - U_{1tn} \\ \vdots \\ U_{Jtn} - U_{1tn} \end{bmatrix}.$$

If we stack equation (A.12), we get a regression expression with an unobservable dependent variable, observed and latent explanatory variables, and interactions between the observed and unobserved attributes:

$$\begin{aligned} \Delta_1 \mathbf{U} &= \Delta_1 \mathbf{X} \boldsymbol{\beta} + \Delta_1 \mathbf{Y}^*(\mathbf{X}, \mathbf{Z}^*) \boldsymbol{\varrho} + \Delta_1 \mathbf{Z}^{*s} \boldsymbol{\gamma} + \Delta_1 \boldsymbol{\nu}, \quad \Delta_1 \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}^{-1}_{\Sigma_{\Delta,N}}), \\ \Delta_1 \mathbf{U} &= \mathbf{X}_\Delta \boldsymbol{\theta} + \Delta_1 \boldsymbol{\nu} \end{aligned} \tag{A.4}$$

where $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \boldsymbol{\varrho}', \boldsymbol{\gamma}')$ is the vector of regression coefficients of the utility function; $\mathbf{X}_\Delta$ is an extended attribute matrix built by appropriately stacking the matrices $\Delta_1 \mathbf{X}$, $\Delta_1 \mathbf{Y}^*(\mathbf{X}, \mathbf{Z}^*)$ and $\Delta_1 \mathbf{Z}^{*s}$; and where

$$\underset{((J-1)TN \times (J-1)TN)}{\mathbf{H}^{-1}_{\Sigma_{\Delta,N}}} = \Delta_1 \begin{bmatrix} \mathbf{H}^{-1}_\Sigma & \mathbf{0}_{J\times J} & \cdots & \mathbf{0}_{J\times J} \\ \mathbf{0}_{J\times J} & \mathbf{H}^{-1}_\Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{J\times J} \\ \mathbf{0}_{J\times J} & \cdots & \mathbf{0}_{J\times J} & \mathbf{H}^{-1}_\Sigma \end{bmatrix} \Delta_1'.$$

In the next subsection we provide a simulator for the latent variable $\mathbf{z}^*$. Therefore, if we take the unconditional observations of the latent attributes, the terms $\mathbf{Z}^{*s}$ and $\mathbf{Y}^*(\mathbf{X}, \mathbf{Z}^*)$ simply enter equation (A.4) as standard observable exogenous attributes. Then, if we simulate observations for the latent utility function then equation (A.4) 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 has a normal distribution. In fact, recall that from the reduced form of the model we can use equation (16) which says that $\pi(\mathbf{U}^*_n | \mathbf{I}^*_n) \sim \mathcal{N}\left(\mathbb{E}(\mathbf{U}^*_n | \mathbf{I}^*_n), \mathrm{Var}(\mathbf{U}^*_n | \mathbf{I}^*_n)\right)$.

However, we need to describe the conditional distribution of the utility function taking into account the choice indicators y_n (McCulloch et al., 2000). Since

$$y_{tn} = \begin{cases} 1 & \text{if } \max \Delta_1 U_{itn} \leq 0 \\ i & \text{if } \Delta_1 U_{itn} > \max\{0, \Delta_1 U_{-itn}\} \end{cases}, \tag{A.5}$$

where U_{-itn} represents the set of all utility functions except U_{itn}, then conditional on

y_{tn}, $\Delta_1 \mathbf{U}_{tn}$ has a truncated multivariate normal distribution:

$$\pi(\Delta_1 \mathbf{U}_{tn}|y_n) \sim \begin{cases} \mathcal{N}\left(\mathbf{X}_\Delta \boldsymbol{\theta}, \mathbf{H}^{-1}_{\Sigma_\Delta}\right) \mathbb{I}_{[\max \Delta_1 U_{itn} \leq 0]} \\ \mathcal{N}\left(\mathbf{X}_\Delta \boldsymbol{\theta}, \mathbf{H}^{-1}_{\Sigma_\Delta}\right) \mathbb{I}_{[\Delta_1 U_{itn} > \max\{0, \Delta_1 U_{-itn}\}]} \end{cases}, \tag{A.6}$$

where $\mathbf{H}^{-1}_{\Sigma_\Delta} = \Delta_1 \mathbf{H}^{-1}_{\Sigma} \Delta_1'$ corresponds to the $(J-1 \times J-1)$ covariance matrix of the utility-difference error term $\Delta_1 \boldsymbol{\nu}_{tn} \sim \mathcal{N}(\mathbf{0}, \Delta_1 \mathbf{H}^{-1}_{\Sigma} \Delta_1')$. We summarize this truncated normal distribution using the following notation:

$$\pi(\Delta_1 \mathbf{U}_{tn}|\mathbf{Z}^{*s}, \boldsymbol{\theta}, \mathbf{H}^{-1}_{\Sigma_\Delta}, y_{tn}) \sim \mathcal{TN}_{\Re|y_{tn}}\left(\mathbf{X}_\Delta \boldsymbol{\theta}, \mathbf{H}^{-1}_{\Sigma_\Delta}\right), \forall t, n, \tag{A.7}$$

where the truncation region $\Re$ is defined by the inequalities in the measurement equation (A.5).

Although data augmentation transforms the estimation problem of the discrete choice kernel into a Bayesian regression, simulations required for $\mathbf{H}^{-1}_{\Sigma}$ must address the normalization of scale. Because we are working with the equation that considers utility differences, using the information contained in the covariance matrix of the model in differences $\mathbf{H}^{-1}_{\Sigma_\Delta}$ it is not possible to identify the $J(J-1)/2$ elements in the original covariance matrix $\mathbf{H}^{-1}_{\Sigma}$. Standard practice is to set the scale of the model by fixing the first diagonal element of $\mathbf{H}^{-1}_{\Sigma_\Delta}$ such that $\sigma^2_{\Delta,11} \equiv \text{var}(\Delta_1 \nu_{1tn}) = \text{var}(\nu_{2tn} - \nu_{1tn}) = 1$.[27] Following Nobile (2000), it is possible to generate Wishart draws given a diagonal element if we assume the Wishart prior $p(\mathbf{H}_\Sigma) \sim W(\check{\nu}_{\mathbf{H}^{-1}_\Sigma}, \check{\mathbf{H}}_\Sigma)\,|_{\sigma^2_{\Delta,11}=1}$ such that

$$\pi(\mathbf{H}_\Sigma) \sim W(\bar{\nu}_{\mathbf{H}^{-1}_\Sigma}, \bar{\mathbf{H}}_\Sigma)\,|_{\sigma^2_{\Delta,11}=1}, \tag{A.8}$$

where

$$\bar{\nu}_{\mathbf{H}^{-1}_\Sigma} = \check{\nu}_{\mathbf{H}^{-1}_\Sigma} + N, \;\; \bar{\mathbf{H}}^{-1}_\Sigma = \check{\mathbf{H}}^{-1}_\Sigma + \sum_{n=1}^{N} \Delta_1 \boldsymbol{\nu}_n \boldsymbol{\nu}_n' \Delta_1'. \tag{A.9}$$

Finally, we take $p(\boldsymbol{\theta}) \sim \mathcal{N}(\check{\boldsymbol{\theta}}, \check{\mathbf{V}}_{\boldsymbol{\theta}})$ as prior belief, and the regression coefficients of the discrete choice kernel can be sampled from the following posterior conditional distribution:

$$\pi(\boldsymbol{\theta}|\mathbf{Z}^*, \Delta_1 \mathbf{U}, \mathbf{H}^{-1}_\Sigma) \sim \mathcal{N}(\bar{\boldsymbol{\theta}}, \bar{\mathbf{V}}_{\boldsymbol{\theta}}), \tag{A.10}$$

where

$$\bar{\mathbf{V}}_{\boldsymbol{\theta}} = (\check{\mathbf{V}}^{-1}_{\boldsymbol{\theta}} + \mathbf{X}_\Delta' \mathbf{H}_{\Sigma_\Delta} \mathbf{X}_\Delta)^{-1}, \;\; \bar{\boldsymbol{\theta}} = \bar{\mathbf{V}}_\theta (\check{\mathbf{V}}^{-1}_{\boldsymbol{\theta}} + \mathbf{X}_\Delta{}' \mathbf{H}_{\Sigma_\Delta} \Delta_1 \mathbf{U}). \tag{A.11}$$

[27] This normalization is equivalent to set the value of the first element of the Cholesky decomposition of the covariance matrix in differences.

Appendix A.2. Conditional distributions of the MIMIC model

Given $\mathbf{I}_n^*$, equation (15) contains the conditional distribution needed for the data augmentation step for the latent $\mathbf{z}_n^*$. Effectively, simulated observations of $\mathbf{z}_n^*$ can be drawn from the normal distribution $\mathcal{N}\left(\mathbb{E}(\mathbf{z}_n^*|\mathbf{I}_n^*), \mathrm{Var}(\mathbf{z}_n^*|\mathbf{I}_n^*)\right)$. Note that $\pi(\mathbf{z}_n^*|\mathbf{I}_n^*)$ accounts for identification of the latent explanatory variables of choice $\mathbf{z}_n^*$ through the effect indicators $\mathbf{I}_n^*$.

Using the simulated observations of $\mathbf{z}_n^*$, both equations 2 and 3 become linear regression models with general covariance matrices. First, we rewrite these equations considering the regression coefficients in vector form and the explanatory variables as a design matrix. Then we stack the N observations together. For the structural equation of $\mathbf{z}^*$ we obtain

$$\underset{(LN\times 1)}{\mathbf{z}^*} = \underset{(LN\times LM)}{\mathbf{W}}\underset{(LM\times 1)}{\tilde{\mathbf{b}}} + \underset{(LN\times 1)}{\tilde{\boldsymbol{\zeta}}}, \quad \tilde{\boldsymbol{\zeta}} \sim \mathcal{N}(0, \mathbf{H}_{\tilde{\Psi}_N}^{-1}), \tag{A.12}$$

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

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

The equivalent expression for equation (3) is

$$\underset{(RN\times 1)}{\mathbf{I}^*} = \underset{(RN\times 1)}{\boldsymbol{\alpha}} + \underset{(RN\times LR)}{\mathbf{Z}^*}\underset{(LR\times 1)}{\boldsymbol{\lambda}} + \underset{(RN\times 1)}{\boldsymbol{\varepsilon}}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{1}_{RN}), \tag{A.13}$$

where $\mathbf{Z}^*$ is a specification matrix formed by appropriately using the elements in $\mathbf{z}_n^*$, $\forall n$; and $\boldsymbol{\lambda}$ is the vector of free factor loadings in $\boldsymbol{\Lambda}$.

Let $\tilde{\boldsymbol{\lambda}} = (\boldsymbol{\alpha}', \boldsymbol{\lambda}')$. If prior beliefs for $\mathbf{b}$ and $\tilde{\boldsymbol{\lambda}}$ are described by $p(\mathbf{b}) \sim \mathcal{N}(\check{\mathbf{b}}, \check{\mathbf{V}}_{\mathbf{b}})$ and $p(\tilde{\boldsymbol{\lambda}}) \sim \mathcal{N}(\check{\boldsymbol{\lambda}}, \check{\mathbf{V}}_{\boldsymbol{\lambda}})$ respectively, then it can verified that, conditional on the other parameters of the model, the posteriors of $\mathbf{b}$ and $\boldsymbol{\lambda}$ are multivariate normal:

$$\pi(\mathbf{b}|\mathbf{Z}^*, \boldsymbol{\theta}, \mathbf{b}, \tilde{\boldsymbol{\lambda}}, \mathbf{y}, \mathbf{I}) \sim \mathcal{N}(\bar{\mathbf{b}}, \bar{\mathbf{V}}_{\mathbf{b}}) \tag{A.14}$$

$$\pi(\tilde{\boldsymbol{\lambda}}|\mathbf{Z}^*, \boldsymbol{\theta}, \mathbf{b}, \mathbf{y}, \mathbf{I}) \sim \mathcal{N}(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{V}}_{\boldsymbol{\lambda}}), \tag{A.15}$$

where

$$\bar{\mathbf{V}}_{\mathbf{b}} = (\check{\mathbf{V}}_{\mathbf{b}}^{-1} + \mathbf{W}'\mathbf{H}_{\tilde{\Psi}_N}\mathbf{W})^{-1}, \;\; \bar{\mathbf{b}} = \bar{\mathbf{V}}_{\mathbf{b}}(\check{\mathbf{V}}_{\mathbf{b}}^{-1} + \mathbf{W}'\mathbf{H}_{\tilde{\Psi}_N}\mathbf{Z}^*) \tag{A.16}$$

$$\bar{\mathbf{V}}_{\boldsymbol{\lambda}} = (\check{\mathbf{V}}_{\boldsymbol{\lambda}}^{-1} + \mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}, \;\; \bar{\boldsymbol{\lambda}} = \bar{\mathbf{V}}_{\boldsymbol{\lambda}}(\check{\mathbf{V}}_{\boldsymbol{\lambda}}^{-1} + \mathbf{Z}^{*\prime}\mathbf{I}). \tag{A.17}$$

The conditional posterior for the general covariance matrix $\mathbf{H}_{\tilde{\Psi}_N}^{-1}$ 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 iid, then the resulting block-diagonal structure, combined with Wishart prior beliefs $p(\mathbf{H}_{\tilde{\Psi}}) \sim W(\check{\nu}_{\tilde{\Psi}}, \check{\mathbf{H}}_{\tilde{\Psi}})$ allow us to obtain

$$\pi(\mathbf{H}_{\tilde{\Psi}}) \sim W(\bar{\nu}_{\tilde{\Psi}}, \bar{\mathbf{H}}_{\tilde{\Psi}}), \tag{A.18}$$

where

$$\bar{\nu}_{\tilde{\Psi}} = \check{\nu}_{\tilde{\Psi}} + N, \;\; \bar{\mathbf{H}}_{\tilde{\Psi}}^{-1} = \check{\mathbf{H}}_{\tilde{\Psi}}^{-1} + \sum_{n=1}^{N} \tilde{\boldsymbol{\zeta}}_n \tilde{\boldsymbol{\zeta}}_n'. \tag{A.19}$$

The estimator for the parameters of the structural equation of the latent variables $\mathbf{z}^*$ assumes that the effect indicators are observed. This is the case of a continuous effect indicator, i.e. when the latent variables are manifested through continuous variables. However, it is not possible to condition directly on the effect indicators if these are dichotomous.

When the effect indicators are dichotomous, the structural equation becomes a binary probit model. In this case, it is possible to use a simpler version of the sampler proposed for the mutinomial probit kernel, which is the standard form of the sampler as suggested by Albert and Chib (1993). In effect, the underlying continuous effect indicators are generated by data augmentation using the following truncated normal posterior distribution:

$$p(I_{rn}^*|I_{rn}) \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(\alpha_{rn} + \boldsymbol{\lambda}_r'\mathbf{z}_n^*, 1) & \text{if } I_{rn} = 1 \\ \mathcal{TN}_{(-\infty,0]}(\alpha_{rn} + \boldsymbol{\lambda}_r'\mathbf{z}_n^*, 1) & \text{if } I_{rn} = 0 \end{cases}. \tag{A.20}$$

Under fairly mild conditions (Gelfand and Smith, 1990) 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 Bayes point estimates are calculated taking the sample means of the Gibbs sampler draws. The mean of the Gibbs sampler draws – the Bayes estimates – are consistent estimators of the corresponding posterior means (Geyer, 1992). 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 errors are simply the standard deviations of the artificial samples generated by the Gibbs sampler.

Appendix B. Tables

Notation	Definition	Dimensions
	Structural equation of the latent variables – Eq. (1)	
$\mathbf{z}_n^*$	Endogenous random vector of latent variables for individual n	$L \times 1$
$\mathbf{\Pi}$	Simultaneity matrix for the latent variables	$L \times L$
$\mathbf{w}_n$	Vector of exogenous explanatory variables (SEM: "causal indicators")	$M \times 1$
$\mathbf{B}$	Matrix of unknown parameters associated with $\mathbf{w}_n$ (Reduced form: $\tilde{\mathbf{B}}$)	$L \times M$
$\mathbf{b}$	Vectorization of $\mathbf{B}$	$LM \times 1$
$\boldsymbol{\zeta}_n$	Error term of the structural equation of the latent variables (Reduced form: $\tilde{\boldsymbol{\zeta}}_n$)	$L \times 1$
$\mathbf{H}_{\Psi}^{-1}$	Covariance matrix of $\boldsymbol{\zeta}_n$ (Reduced form: $\tilde{\mathbf{H}}_{\tilde{\Psi}}^{-1}$)	$L \times L$
	Structural equation of the indicators – Eq. (3)	
$\mathbf{I}_n^*$	Latent (or observed) vector of indicators measuring $\mathbf{z}_n^*$	$R \times 1$
$\boldsymbol{\alpha}$	Intercept vector	$R \times 1$
$\mathbf{\Lambda}$	Matrix of unknown factor loadings	$R \times L$
$\boldsymbol{\lambda}$	Vectorization of $\mathbf{\Lambda}$	$RL \times 1$
$\boldsymbol{\varepsilon}_n$	Measurement error term	$R \times 1$
$\mathbf{H}_{\Theta}^{-1}$	Covariance matrix of $\boldsymbol{\varepsilon}_n$	$R \times R$
	Measurement equation of the indicators – Eq. (4)	
$\mathbf{I}_n$	Vector of observed indicators manifesting both $\mathbf{I}_n^*$ and $\mathbf{z}_n^*$ (SEM: "effect indicators")	$R \times 1$
$\boldsymbol{\mu}_r$	Vector of threshold parameters when $I_{rn} \in \mathbf{I}_n$ in an ordered variable	$M_r - 1 \times 1$
	Choice model – Eqs (2) and (5)	
$\mathbf{U}_{tn}^*$	Vector of indirect utility functions for choice situation t	$J \times 1$
$\mathbf{X}_{tn}$	Design matrix of exogenous attributes (with row elements $\mathbf{x}'_{tin}$)	$J \times K$
$\boldsymbol{\beta}$	Vector of unknown preference parameters (marginal utilities) associated with $\mathbf{X}_{tn}$	$K \times 1$
$\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \mathbf{z}_n^*)$	Interaction matrix among exogenous and endogenous, latent attributes	$J \times Q$
$\boldsymbol{\varrho}$	Vector of unknown parameters associated with the interactions $\mathbf{Y}_{tn}^*(\mathbf{X}_{tn}, \mathbf{z}_n^*)$	$Q \times 1$
$\mathbf{\Gamma}$	Vector of unknown parameters associated with the endogenous, latent attributes $\mathbf{z}_n^*$	$J \times L$
$\boldsymbol{\nu}_{nt}$	Taste shock (random term of the choice kernel	$J \times 1$
$\mathbf{H}_{\Sigma}^{-1}$	Covariance matrix of $\boldsymbol{\nu}_{nt}$	$J \times J$
y_{tn}	Choice indicator	1×1

Table B.1: Notation being used for the system of equations describing the general model

	N=500				N=1500				N=2500			
Nalt	IID	Heterosk.	Nested	Full	IID	Heterosk.	Nested	Full	IID	Heterosk.	Nested	Full
5	0.94	3.61	4.67	0.95	3.17	4.67	1.03	3.76	7.20	0.97	3.48	6.87
6	1.08	3.95	5.32	1.15	3.80	5.39	2.37	4.21	8.35	1.05	4.25	8.41
7	1.31	4.76	9.74	1.38	4.38	6.18	2.74	4.97	9.70	1.67	5.02	9.78
8	1.68	5.54	10.99	1.61	4.95	7.35	3.11	5.78	11.26	1.93	5.83	11.29
9	1.83	6.18	12.16	1.80	4.64	8.28	3.33	6.44	12.26	2.16	4.63	12.29
10	1.98	6.68	13.14	1.99	5.99	9.15	3.58	7.09	13.30	2.39	7.00	13.14

Table B.2: Estimation time – 10,000 repetitions of the Gibbs sampler [min]

IID		N=500				N=1500				N=2500			
Nalt	Param.	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage
5	β_1	0.0155	0.1746	0.0902	93	0.0140	0.3031	0.0481	92	0.0134	0.3529	0.0401	90
	γ_1	-0.0043	-0.0508	0.0851	90	0.0248	0.5123	0.0543	94	0.0176	0.4582	0.0422	96
	γ_2	0.0264	0.2834	0.0970	96	0.0305	0.6399	0.0566	90	0.0251	0.7343	0.0424	90
6	β_1	0.0227	0.3485	0.0689	91	0.0167	0.4180	0.0434	92	0.0164	0.4696	0.0385	93
	γ_1	0.0177	0.2517	0.0723	92	0.0124	0.2925	0.0441	96	0.0182	0.4906	0.0413	96
	γ_2	0.0142	0.2036	0.0713	94	0.0181	0.4264	0.0461	96	0.0285	0.8636	0.0436	95
7	β_1	0.0168	0.2327	0.0741	96	0.0054	0.1405	0.0386	94	0.0114	0.3496	0.0346	91
	γ_1	0.0125	0.1613	0.0786	95	0.0123	0.3044	0.0424	95	0.0212	0.5972	0.0414	96
	γ_2	0.0242	0.3406	0.0750	94	0.0250	0.6678	0.0450	96	0.0213	0.6534	0.0390	91
8	β_1	0.0165	0.2713	0.0631	91	0.0145	0.4213	0.0374	91	0.0132	0.4036	0.0353	95
	γ_1	0.0197	0.3568	0.0586	96	0.0204	0.4891	0.0465	91	0.0166	0.5592	0.0340	96
	γ_2	0.0203	0.3033	0.0700	96	0.0249	0.6347	0.0464	94	0.0306	1.0450	0.0423	95
9	β_1	0.0186	0.2858	0.0676	94	0.0120	0.2923	0.0426	96	0.0114	0.3625	0.0333	93
	γ_1	0.0221	0.3860	0.0612	96	0.0193	0.5058	0.0427	93	0.0184	0.5885	0.0363	90
	γ_2	0.0166	0.2542	0.0673	92	0.0248	0.6475	0.0456	93	0.0301	1.0010	0.0426	92
10	β_1	0.0124	0.2081	0.0609	96	0.0171	0.4765	0.0397	94	0.0096	0.3594	0.0285	92
	γ_1	0.0258	0.4244	0.0659	95	0.0161	0.4279	0.0409	96	0.0157	0.5111	0.0346	91
	γ_2	0.0104	0.1548	0.0682	92	0.0250	0.7505	0.0417	91	0.0305	1.0766	0.0416	91
Heterosk.		N=500				N=1500				N=2500			
Nalt	Param.	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage
5	β_1	-0.0035	-0.0440	0.0805	96	0.0042	0.0937	0.0447	91	0.0019	0.0467	0.0402	96
	γ_1	-0.0132	-0.1501	0.0890	91	0.0070	0.1381	0.0513	92	0.0022	0.0569	0.0393	91
	γ_2	0.0053	0.0602	0.0886	95	0.0125	0.2415	0.0531	95	0.0011	0.0317	0.0357	92
6	β_1	0.0052	0.0727	0.0712	93	0.0043	0.1050	0.0412	91	0.0043	0.1113	0.0385	96
	γ_1	0.0001	0.0008	0.0745	95	0.0009	0.0183	0.0488	90	0.0033	0.0842	0.0398	92
	γ_2	-0.0165	-0.2207	0.0768	95	-0.0014	-0.0269	0.0518	92	0.0058	0.1637	0.0360	92
7	β_1	0.0039	0.0477	0.0818	90	-0.0074	-0.1743	0.0430	94	0.0003	0.0093	0.0336	96
	γ_1	0.0060	0.0704	0.0856	93	-0.0056	-0.123	0.0455	91	0.0047	0.1241	0.0378	93
	γ_2	0.0013	0.0184	0.0679	91	0.0032	0.0743	0.0429	96	-0.0017	-0.0490	0.0356	94
8	β_1	0.0025	0.0369	0.0690	91	0.0051	0.1117	0.0464	92	-0.0035	-0.1032	0.0336	90
	γ_1	0.0084	0.1212	0.0696	92	0.0085	0.1917	0.0451	95	-0.0006	-0.0199	0.0298	92
	γ_2	0.0053	0.0685	0.0780	96	-0.0006	-0.0149	0.0389	96	0.0074	0.2346	0.0324	96
9	β_1	0.0042	0.0664	0.0631	91	0.0015	0.0351	0.0430	94	0.0005	0.0168	0.0287	95
	γ_1	0.0051	0.0749	0.0682	94	0.0044	0.1069	0.0416	93	0.0019	0.0588	0.0326	93
	γ_2	0.0062	0.0839	0.0737	93	0.0059	0.1429	0.0416	93	0.0059	0.1773	0.0337	94
10	β_1	-0.0012	-0.0212	0.0586	95	0.0015	0.0413	0.0370	90	-0.0040	-0.1321	0.0306	95
	γ_1	0.0064	0.0946	0.0683	90	0.0060	0.1409	0.0427	92	0.0037	0.1094	0.0336	93
	γ_2	-0.0028	-0.0417	0.0661	93	0.0030	0.0767	0.0388	93	0.0084	0.2747	0.0319	93
Nested		N=500				N=1500				N=2500			
Nalt	Param.	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage
5	β_1	0.0216	0.3129	0.0724	94	0.0200	0.5051	0.0444	90	0.0178	0.5625	0.0364	93
	γ_1	0.0071	0.0970	0.0736	92	0.0290	0.7287	0.0493	91	0.0216	0.6119	0.0413	95
	γ_2	0.0346	0.4275	0.0880	94	0.0371	0.9405	0.0541	94	0.0325	1.1996	0.0424	90
6	β_1	0.0262	0.3909	0.0721	91	0.0178	0.4456	0.0436	92	0.0216	0.6448	0.0398	96
	γ_1	0.0243	0.3907	0.0669	94	0.0179	0.4189	0.0463	96	0.0245	0.7352	0.0414	90
	γ_2	0.0217	0.3154	0.0720	91	0.0261	0.6685	0.0470	91	0.0348	1.1088	0.0468	93
7	β_1	0.0170	0.2427	0.0721	91	0.0096	0.2434	0.0405	95	0.0157	0.5437	0.0329	93
	γ_1	0.0185	0.2402	0.0793	91	0.0192	0.4682	0.0453	92	0.0276	0.8495	0.0426	92
	γ_2	0.0304	0.4332	0.0765	96	0.0326	0.8483	0.0504	92	0.0285	0.8865	0.0429	93
8	β_1	0.0223	0.3257	0.0719	96	0.0187	0.5083	0.0412	92	0.0176	0.5347	0.0373	96
	γ_1	0.0226	0.4177	0.0586	94	0.0274	0.6265	0.0516	95	0.0186	0.6164	0.0354	91
	γ_2	0.0258	0.3441	0.0792	96	0.0318	0.8107	0.0505	95	0.0362	1.2467	0.0464	91
9	β_1	0.0216	0.3676	0.0627	96	0.0174	0.3922	0.0477	94	0.0158	0.5202	0.0342	91
	γ_1	0.0256	0.4202	0.0660	94	0.0255	0.6560	0.0465	94	0.0234	0.7703	0.0383	92
	γ_2	0.0267	0.4281	0.0679	95	0.0320	0.8668	0.0489	92	0.0359	1.1934	0.0468	90
10	β_1	0.0169	0.2979	0.0591	91	0.0243	0.6886	0.0428	90	0.0107	0.3836	0.0300	90
	γ_1	0.0265	0.4196	0.0686	95	0.0209	0.5659	0.0424	92	0.0201	0.6190	0.0382	96
	γ_2	0.0233	0.3608	0.0687	93	0.0312	0.8834	0.0472	91	0.0384	1.4442	0.0467	96
Full		N=500				N=1500				N=2500			
Nalt	Param.	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage	Bias	t-stat	RMSE	Coverage
5	β_1	0.0119	0.1233	0.0973	93	0.0176	0.3183	0.0579	90	0.0244	0.5502	0.0505	94
	γ_1	0.0169	0.1586	0.1079	90	0.0323	0.5975	0.0629	94	0.0270	0.5904	0.0530	90
	γ_2	0.0451	0.4635	0.1073	94	0.0448	0.8035	0.0716	95	0.0439	0.9544	0.0635	95
6	β_1	0.0353	0.4402	0.0877	90	0.0206	0.4148	0.0537	93	0.0274	0.6495	0.0503	95
	γ_1	0.0302	0.3648	0.0881	94	0.0291	0.5392	0.0613	95	0.0300	0.7751	0.0490	96
	γ_2	0.0362	0.4305	0.0916	90	0.0411	0.8107	0.0652	90	0.0396	1.0123	0.0556	91
7	β_1	0.0209	0.2388	0.0901	95	0.0201	0.4675	0.0475	94	0.0252	0.6618	0.0457	90
	γ_1	0.0258	0.3309	0.0823	93	0.0257	0.5654	0.0522	95	0.0310	0.7331	0.0524	92
	γ_2	0.0412	0.5065	0.0912	96	0.0370	0.8673	0.0565	93	0.0382	0.9284	0.0562	93
8	β_1	0.0396	0.6593	0.0719	96	0.0339	0.9395	0.0495	94	0.0274	0.9753	0.0392	93
	γ_1	0.0382	0.5869	0.0754	96	0.0309	0.7027	0.0538	93	0.0338	1.1271	0.0452	93
	γ_2	0.0481	0.8073	0.0766	93	0.0467	0.9978	0.0661	95	0.0533	1.7403	0.0615	94
9	β_1	0.0301	0.3673	0.0874	94	0.0259	0.6677	0.0467	94	0.0308	1.0923	0.0418	95
	γ_1	0.0380	0.5621	0.0775	96	0.0276	0.7527	0.0459	93	0.0328	1.1360	0.0438	95
	γ_2	0.0464	0.6888	0.0818	92	0.0468	1.0628	0.0643	90	0.0507	1.5540	0.0603	90
10	β_1	0.0363	0.5297	0.0776	94	0.0241	0.6056	0.0465	90	0.0217	0.7083	0.0375	95
	γ_1	0.0363	0.5885	0.0716	90	0.0419	1.1593	0.0553	92	0.0366	1.2153	0.0473	92
	γ_2	0.0562	0.9575	0.0812	93	0.0557	1.5476	0.0664	93	0.0531	1.7630	0.0610	92

Table B.3: Summary of Monte Carlo Study for selected parameters, Bayes estimator

Attribute	SGV	AFV	HEV	HFC
PP	100% PP 105% PP 110% PP 115% PP	105% PP 110% PP	105% PP 120% PP	110% PP 120% PP
FC	100% FC 110% FC 120% PP 130% PP	110% FC 120% PP	75% ICV	110% FC 120% FC
FA	1	0.25 0.75	1	0.25 0.75
Express	No	No Yes	= AFV	No Yes
POW	100% POW	100% POW 90% POW	100% POW 90% POW	100% POW 90% POW

Table B.4: Discrete choice experiment: attribute levels for the vehicle choice experiment. Source: Horne (2003).

Attribute	car	carpool	transit	park & ride	walk or cycle
Travel cost TC	$100\% N_{Cost}$ $110\% N_{Cost}$	$50\% N_{Cost}$ $75\% N_{Cost}$	\$60 /month \$100 /month	$25\% N_{Cost} + TC_{transit}$ $50\% N_{Cost} + TC_{transit}$	\$0 / month
Travel time TT	$90\% N_{Time}$ $100\% N_{Time}$ $110\% N_{Time}$ $120\% N_{Time}$	$90\% N_{Time}$ $100\% N_{Time}$	$105\%\ N_{Time}$ $115\%\ N_{Time}$	$95\% N_{Time}$ $105\% N_{Time}$	N_{Dist} / (6 or 15 km/hr) N_{Dist} / (8 or 20 km/hr))
Pickup & drop-off PDT		5 minutes 10 minutes			
Access time WWT			5 minutes 15 minutes	5 minutes 10 minutes	
Transfers $TRANS$			None One	None One	
Bike path $PATH$					Yes No

Table B.5: Discrete choice experiment: attribute levels for the travel mode choice experiment. Source: Horne (2003).

Base vehicle choice model: probit without latent attributes*

Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Purchase price (PP)	-2.222	0.714	-3.620	-3.399	-2.235	-1.044	-0.823
Fuel cost (FC)	-1.219	0.437	-2.076	-1.940	-1.229	-0.497	-0.361
Fuel availability (FA)	0.740	0.187	0.373	0.431	0.742	1.049	1.107
Express lane access ($Express$)	0.093	0.040	0.014	0.027	0.094	0.159	0.171
Power (POW)	0.976	0.367	0.256	0.370	0.982	1.582	1.696

Vehicle choice model 1: probit with two additive latent attributes

Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Purchase price (PP)	-0.334	0.118	-0.607	-0.560	-0.330	-0.180	-0.160
Fuel cost (FC)	-0.319	0.103	-0.546	-0.507	-0.307	-0.170	-0.150
Fuel availability (FA)	0.560	0.161	0.282	0.314	0.547	0.850	0.910
Express lane access ($Express$)	0.071	0.032	0.017	0.024	0.068	0.130	0.140
Power (POW)	1.093	0.375	0.487	0.555	1.054	1.760	1.900
Pro-environment on HFC	0.635	0.086	0.476	0.499	0.632	0.780	0.810
Pro-environment on AFV	0.071	0.133	-0.228	-0.160	0.082	0.260	0.300
Pro-environment on HEV	0.359	0.080	0.218	0.238	0.354	0.500	0.530
Pro-safety on HFC	-0.315	0.087	-0.492	-0.462	-0.313	-0.180	-0.150
Pro-safety on AFV	0.094	0.138	-0.129	-0.097	0.075	0.340	0.410
Pro-safety on HEV	-0.194	0.077	-0.363	-0.330	-0.188	-0.080	-0.061

Vehicle choice model 2: probit with two latent attributes, interacting

Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Purchase price (PP)	-0.337	0.109	-0.578	-0.533	-0.328	-0.176	-0.152
Fuel cost (FC)	-0.368	0.102	-0.590	-0.549	-0.360	-0.216	-0.192
Fuel availability (FA)	0.585	0.131	0.354	0.386	0.575	0.817	0.864
Express lane access ($Express$)	0.067	0.031	0.012	0.020	0.065	0.121	0.134
Power (POW)	0.999	0.373	0.352	0.438	0.968	1.658	1.792
Pro-performance × power	0.358	0.387	-0.406	-0.274	0.356	0.994	1.141
Cost-cons. × fuel cost	-0.228	0.104	-0.449	-0.407	-0.223	-0.066	-0.037

Vehicle choice model 3: probit with four latent attributes, including interactions

Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Purchase price (PP)	-0.358	0.116	-0.616	-0.564	-0.346	-0.190	-0.163
Fuel cost (FC)	-0.396	0.107	-0.628	-0.587	-0.389	-0.230	-0.210
Fuel availability (FA)	0.616	0.146	0.361	0.396	0.605	0.870	0.926
Express lane access ($Express$)	0.074	0.033	0.016	0.024	0.072	0.130	0.145
Power (POW)	0.963	0.365	0.317	0.403	0.941	1.590	1.729
Pro-performance × power	0.531	0.425	-0.262	-0.137	0.517	1.260	1.431
Cost-cons. × fuel cost	-0.283	0.108	-0.510	-0.470	-0.278	-0.120	-0.086
Pro-environment on HFC	0.635	0.081	0.480	0.503	0.633	0.770	0.800
Pro-environment on AFV	-0.037	0.134	-0.313	-0.258	-0.034	0.180	0.223
Pro-environment on HEV	0.353	0.084	0.213	0.230	0.344	0.500	0.533
Pro-safety on HFC	-0.301	0.087	-0.477	-0.447	-0.300	-0.160	-0.133
Pro-safety on AFV	0.189	0.142	-0.062	-0.019	0.177	0.440	0.499
Pro-safety on HEV	-0.195	0.084	-0.382	-0.347	-0.187	-0.070	-0.047

*The base model contains statistically significant interactions between the observed attributes and the sociodemographic variables.

Table B.6: Vehicle choice model – diffuse prior, precision = 0.1

Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Base mode choice model: probit without latent attributes*							
Travel cost (TC)	-0.016	0.003	-0.021	-0.021	-0.016	-0.012	-0.011
Travel or driving time (TT)	-0.027	0.003	-0.033	-0.032	-0.027	-0.022	-0.021
Pickup & drop-off time (PDT)	-0.074	0.011	-0.095	-0.092	-0.075	-0.057	-0.053
Access time (WWT)	-0.057	0.015	-0.087	-0.082	-0.057	-0.033	-0.028
Transfers ($TRANS$)	0.294	0.101	0.095	0.126	0.296	0.461	0.492
Bike path ($PATH$)	-0.022	0.054	-0.127	-0.110	-0.022	0.067	0.083
Mode choice model 1: probit with one alternative-specific latent attribute							
Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Travel cost (TC)	-0.007	0.001	-0.009	-0.009	-0.007	-0.005	-0.005
Travel or driving time (TT)	-0.011	0.002	-0.014	-0.014	-0.011	-0.009	-0.008
Pickup & drop-off time (PDT)	-0.039	0.008	-0.056	-0.053	-0.038	-0.027	-0.026
Access time (WWT)	-0.041	0.008	-0.057	-0.054	-0.040	-0.029	-0.028
Transfers ($TRANS$)	-0.035	0.020	-0.077	-0.070	-0.035	-0.003	0.003
Bike path ($PATH$)	0.016	0.024	-0.029	-0.022	0.016	0.056	0.063
Pro-transit on solo driver	-0.270	0.054	-0.371	-0.354	-0.265	-0.178	-0.161
Pro-transit on carpooler	0.012	0.031	-0.044	-0.036	0.010	0.065	0.077
Pro-transit on transit	0.061	0.040	-0.012	-0.001	0.059	0.129	0.145
Pro-transit on park & ride	-0.069	0.040	-0.148	-0.135	-0.069	-0.005	0.007
Mode choice model 2: probit with one additive latent attribute (on transit)							
Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Travel cost (TC)	-0.007	0.001	-0.010	-0.009	-0.007	-0.005	-0.005
Travel or driving time (TT)	-0.012	0.002	-0.015	-0.015	-0.012	-0.009	-0.009
Pickup & drop-off time (PDT)	-0.043	0.007	-0.058	-0.055	-0.042	-0.031	-0.029
Access time (WWT)	-0.042	0.008	-0.058	-0.055	-0.041	-0.030	-0.028
Transfers ($TRANS$)	-0.036	0.021	-0.081	-0.073	-0.035	-0.003	0.003
Bike path ($PATH$)	0.014	0.025	-0.035	-0.027	0.014	0.053	0.062
Pro-transit on transit	0.062	0.031	0.006	0.014	0.060	0.115	0.127
Mode choice model 3: probit with two latent attributes, including interactions							
Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Travel cost (TC)	-0.007	0.001	-0.010	-0.010	-0.007	-0.005	-0.005
Travel or driving time (TT)	-0.012	0.002	-0.015	-0.015	-0.012	-0.010	-0.009
Pickup & drop-off time (PDT)	-0.043	0.007	-0.058	-0.055	-0.043	-0.032	-0.030
Access time (WWT)	-0.042	0.009	-0.060	-0.058	-0.042	-0.029	-0.027
Transfers ($TRANS$)	-0.036	0.021	-0.082	-0.073	-0.035	-0.004	0.002
Bike path ($PATH$)	0.016	0.026	-0.035	-0.026	0.015	0.058	0.067
Pro-transit on transit	0.064	0.032	0.009	0.017	0.062	0.122	0.136
Travel cost × cost cons.	-0.002	0.001	-0.005	-0.004	-0.002	-0.001	-0.0001

*The base model contains statistically significant interactions between the observed attributes and the sociodemographic variables.

Table B.7: Travel mode choice model – diffuse prior, precision = 0.1

Vehicle choice model 3: probit with four latent attributes, including interactions – tight prior							
Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Purchase price (PP)	-0.337	0.107	-0.572	-0.525	-0.326	-0.177	-0.153
Fuel cost (FC)	-0.363	0.101	-0.577	-0.539	-0.356	-0.209	-0.185
Fuel availability (FA)	0.711	0.147	0.421	0.466	0.713	0.945	0.990
Express lane access ($Express$)	0.071	0.031	0.015	0.023	0.069	0.126	0.138
Power (POW)	1.038	0.373	0.366	0.465	1.018	1.687	1.822
Pro-performance × power	0.660	0.425	-0.156	-0.021	0.651	1.366	1.516
Cost-cons. × fuel cost	-0.303	0.107	-0.525	-0.486	-0.298	-0.134	-0.101
Pro-environment on HFC	0.616	0.073	0.477	0.498	0.615	0.739	0.762
Pro-environment on AFV	-0.183	0.161	-0.546	-0.471	-0.173	0.064	0.108
Pro-environment on HEV	0.295	0.069	0.174	0.192	0.290	0.416	0.443
Pro-safety on HFC	-0.291	0.082	-0.459	-0.430	-0.290	-0.159	-0.134
Pro-safety on AFV	0.310	0.187	0.002	0.047	0.291	0.658	0.766
Pro-safety on HEV	-0.167	0.071	-0.320	-0.288	-0.162	-0.060	-0.041
Mode choice model 3: probit with two latent attributes, including interactions – tight prior							
Attribute	Estimate	s.d.	2.5%	5%	50%	95%	97.5%
Travel cost (TC)	-0.008	0.001	-0.010	-0.010	-0.008	-0.006	-0.005
Travel or driving time (TT)	-0.013	0.001	-0.015	-0.015	-0.013	-0.010	-0.010
Pickup & drop-off time (PDT)	-0.042	0.007	-0.059	-0.056	-0.042	-0.032	-0.031
Access time (WWT)	-0.042	0.007	-0.059	-0.056	-0.042	-0.032	-0.031
Transfers ($TRANS$)	-0.037	0.022	-0.082	-0.073	-0.036	-0.003	0.003
Bike path ($PATH$)	0.014	0.027	-0.040	-0.031	0.014	0.059	0.069
Pro-transit on transit	0.063	0.032	0.006	0.015	0.061	0.119	0.133
Travel cost × cost cons.	-0.002	0.001	-0.005	-0.005	-0.002	0.000	0.000

Table B.8: Prior Sensitivity – model 3 estimates, precision = 100

Pro-transit				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Number of vehicles	-0.205	0.073	-0.348	-0.062
Solo driver	-0.125	0.112	-0.345	0.095
Transit user	0.485	0.179	0.134	0.836
Gender	0.200	0.092	0.020	0.380
Income 40K-60K	0.155	0.135	-0.110	0.420
Age 26-55	0.361	0.164	0.040	0.682
Age 56+	0.685	0.195	0.303	1.067
Pro-environment				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Pro-transit	0.544	0.087	0.373	0.715
Gender	0.124	0.100	-0.072	0.320
University	0.159	0.105	-0.047	0.365
Pro-safety				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Carpooler	0.196	0.183	-0.163	0.555
Bicyclist	-0.744	0.362	-1.454	-0.034
Gender	0.457	0.089	0.283	0.631
Age 26-55	0.292	0.165	-0.031	0.615
Age 56+	0.629	0.193	0.251	-1.007
Cost-consciousness				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Gender	0.296	0.096	0.108	0.484
University	0.075	0.088	-0.097	0.247
Income 60K-80K	-0.242	0.149	-0.534	0.050
Income 80K+	-0.533	0.165	-0.856	-0.210
Pro-performance				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Cost-consciousness	0.456	0.186	0.091	0.821
Bicyclist	-0.678	0.377	-1.417	0.061
Age 26-55	0.414	0.216	-0.009	0.837
Age 56+	0.608	0.256	0.106	1.110
Income 80K+	0.355	0.213	-0.062	0.772

Table B.9: Structural equation

Pro-transit				
Effect indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Express lanes for carpooling and transit	1.000			
Making transit more attractive	0.865	0.120	0.639	1.100
Discouraging automobile use	0.619	0.098	0.427	0.811
Pro-environment				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Emissions contributing to global warming	1.000			
Vehicle emissions impacting local air quality	0.937	0.057	0.825	1.049
Building new roads	-0.218	0.056	-0.328	-0.108
Pro-safety				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Unsafe communities due to speeding traffic	1.000			
Accidents caused by drivers	0.972	0.078	0.819	1.125
Importance of safety (veh purchase)	0.779	0.078	0.626	0.932
Cost-consciousness				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Importance of purchase price (veh purchase)	1.000			
Importance of fuel economy (veh purchase)	0.821	0.233	0.364	1.278
Importance of cost (mode choice)	0.257	0.115	0.032	0.482
Pro-performance				
Causal indicator	Estimate	s.d.	Lower bound 95% CI	Upper bound 95% CI
Importance of reliability (veh purchase)	1.000			
Importance of fuel economy (veh purchase)	0.528	0.109	0.314	0.742
Importance of horsepower (veh purchase)	0.420	0.089	0.246	0.594

Table B.10: Measurement equation

	Base model			Model 3		
				Base scenario		
Alternative	share	s.d.	% change	share	s.d.	% change
Car	0.392	0.011		0.393	0.011	
Carpool	0.271	0.010		0.268	0.010	
Transit	0.204	0.009		0.202	0.009	
Park & ride	0.038	0.004		0.038	0.004	
Walk or cycle	0.097	0.006		0.099	0.006	
HFC	0.362	0.011		0.361	0.011	
AFV	0.038	0.004		0.038	0.004	
HEV	0.488	0.011		0.487	0.012	
ICV	0.114	0.007		0.115	0.007	
	Scenario 1: increase of travel cost of car and carpool of 25%					
Alternative	share	s.d.	% change	share	s.d.	% change
Car	0.366	0.011	-5.1%	0.371	0.012	-5.7%
Carpool	0.255	0.009	-7.8%	0.253	0.010	-5.6%
Transit	0.225	0.010	10.4%	0.218	0.010	8.1%
Park & ride	0.049	0.005	31.6%	0.047	0.005	25.6%
Walk or cycle	0.107	0.006	9.6%	0.111	0.007	11.6%
	Scenario 2: increase of travel cost of car and carpool of 50%					
Alternative	share	s.d.	% change	share	s.d.	% change
Car	0.340	0.010	-10.2%	0.348	0.013	-11.5%
Carpool	0.238	0.011	-15.6%	0.237	0.011	-11.5%
Transit	0.245	0.010	19.8%	0.234	0.011	16.0%
Park & ride	0.061	0.007	68.4%	0.058	0.006	54.3%
Walk or cycle	0.118	0.007	18.9%	0.122	0.007	23.5%
	Scenario 3: increase in gasoline cost of 50%					
Alternative	share	s.d.	% change	share	s.d.	% change
HFC	0.470	0.021	29.8%	0.442	0.027	22.6%
AFV	0.052	0.008	37.5%	0.058	0.008	54.8%
HEV	0.394	0.024	-19.4%	0.422	0.024	-13.2%
ICV	0.087	0.010	-24.1%	0.077	0.011	-32.9%
	Scenario 4: increase in power of hybrids of 15%					
Alternative	share	s.d.	% change	share	s.d.	% change
HFC	0.300	0.013	-17.5%	0.293	0.018	-18.7%
AFV	0.030	0.004	-16.2%	0.034	0.004	-10.6%
HEV	0.580	0.021	18.2%	0.591	0.024	21.4%
ICV	0.091	0.010	-18.3%	0.082	0.010	-28.3%

Table B.11: Market share forecasts

	Model 3		
	Scenario 5: maximum pro-environment consumers		
Alternative	share	s.d.	% change
Car	0.383	0.013	-2.6%
Carpool	0.255	0.010	-4.9%
Transit	0.236	0.016	16.9%
Park & ride	0.035	0.004	-6.2%
Walk or cycle	0.091	0.006	-8.3%
HFC	0.380	0.011	5.3%
AFV	0.036	0.004	-5.1%
HEV	0.480	0.012	-1.4%
ICV	0.104	0.007	-9.1%
	Scenario 6: maximum cost-conscious consumers		
Alternative	share	s.d.	% change
Car	0.390	0.011	-0.9%
Carpool	0.267	0.010	-0.3%
Transit	0.202	0.009	0.1%
Park & ride	0.037	0.004	-2.8%
Walk or cycle	0.104	0.007	5.3%
HFC	0.353	0.011	-2.1%
AFV	0.037	0.004	-1.2%
HEV	0.500	0.013	2.8%
ICV	0.110	0.007	-4.6%

Table B.12: Market share forecasts, after a shock in the structural equation of the latent attributes

Vehicle choice base model					
Annualized rate	Mean	Median	s.d.	LB 95% CI	UB 95% CI
r	27.23%	21.09%	22.37%	6.54%	69.30%
Vehicle choice model 3 - randomly selected individual					
Annualized rate	Mean	Median	s.d.	LB 95% CI	UP 95% CI
r_n (Income <60K)	16.13%	15.51%	4.22%	7.85%	26.13%
r_n (Income 60K-80K)	17.08%	16.05%	4.33%	8.56%	27.47%
r_n (Income 80K+)	18.24%	17.67%	4.50%	9.16%	29.00%

Table B.13: Annual implicit discount rate for energy savings