

Forecasting adoption of ultra-low-emission vehicles using Bayes estimates of a multinomial probit model and the GHK simulator

Ricardo A. Daziano* and Martin Achnicht†

January 2013

Abstract

In this paper we use Bayes estimates of a multinomial probit model with fully flexible substitution patterns to forecast consumer response to ultra-low-emission vehicles. In this empirical application of the probit Gibbs sampler, we use stated-preference data on vehicle choice from a Germany-wide survey of potential light-duty-vehicle buyers using computer-assisted personal interviewing. We show that Bayesian estimation of a multinomial probit model with a full covariance matrix is feasible for this medium-scale problem and provides results that are very similar to maximum simulated likelihood estimates. Using the posterior distribution of the parameters of the vehicle choice model as well as the GHK simulator we derive the choice probabilities of the different alternatives. We first show that the Bayes point estimates of the market shares reproduce the observed values. Then, we define a base scenario of vehicle attributes that aims at representing an average of the current vehicle choice situation in Germany. Consumer response to qualitative changes in the base scenario is subsequently studied. In particular, we analyze the effect of increasing the network of service stations for charging electric vehicles as well as for refueling hydrogen. The result is the posterior distribution of the choice probabilities that represent adoption of the energy-efficient technologies.

JEL classification: C25, D12, Q42.

Keywords: Discrete choice models; Bayesian econometrics; Low emission vehicles; Charging infrastructure

*School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853; Email: daziano@cornell.edu

†Centre for European Economic Research (ZEW), L7,1, D-68161 Mannheim, Germany; Email: achnicht@zew.de

1 Introduction

Consumer shift to ultra-low-emission vehicles has been regarded as a way to promote sustainable personal transportation. Whereas new low-emission technologies – including battery electric vehicles – have clear benefits such as efficiency gains and emission reductions, there are several barriers preventing broad adoption. On the one hand, electric vehicles are much more expensive than standard gas vehicles with a similar build. On the other hand, consumers face reliability issues, namely limited and variable driving range, and lack of refueling stations. Discrete choice models are a powerful tool to understand how consumers evaluate these tradeoffs and decide which vehicle to purchase (Bunch et al., 1993; Brownstone et al., 1996; Brownstone and Train, 1999; Brownstone et al., 2000; Horne et al., 2005; Daziano and Bolduc, 2011; Hensher et al., 2011; Achtnicht et al., 2012). Additionally, since the automotive market presents highly differentiated products with several qualitative attributes that are hard to measure, it is desirable to work with flexible discrete choice models that allow for both consumer and error heterogeneity. For instance, unobservable qualitative attributes that may be shared – completely or partially – among differentiated products, such as light duty vehicles, may create correlation patterns that can be fairly complex (see Train, 2009).

The multinomial probit model (Thurstone, 1927) is a direct strategy for addressing heterogeneity of the error term in random utility maximization. In effect, to avoid the econometric problems of biased and inconsistent parameters related to specification error, there are two possible strategies for dealing with random heterogeneity. First, the modeler can include additional additive error terms that create correlation or heteroskedasticity. This is the modeling strategy of mixed logit models (McFadden and Train, 2000). A second strategy is to introduce more general structures that are derived directly from the covariance matrix. In the case of the multinomial probit model, the general assumption is a direct generalization of the covariance structure through error terms that have a multivariate normal distribution. Applications of the multinomial probit model have included both constrained and unconstrained versions of the covariance matrix of the multivariate normally distributed error term (see Daganzo, 1979; Bolduc and Ben-Akiva, 1991; Munizaga and Daziano, 2005; Ziegler,

2012).

To forecast consumer response to ultra-low-emission vehicles, in this paper we adopt a multinomial probit model with a general covariance structure that offers fully flexible substitution patterns among alternatives.

Fully flexible models do not need to assume a particular covariance structure; instead, the substitution patterns are revealed from the data. Because the choice probabilities of the probit model do not have a convenient closed form, simulation is required to evaluate the multi-fold integral that represents the probit choice probabilities. On the one hand, the GHK recursive probability simulator has been proposed and successfully tested for deriving a frequentist estimator of the parameters of the model (Geweke, 1991; Hajivassiliou and McFadden, 1998; Keane, 1994). However, the multinomial probit loglikelihood function is not globally concave. In fact, the maximum simulated likelihood estimator using the GHK simulator is relatively computationally expensive for high dimensions, may produce a poor approximation of the asymptotic covariance matrix (Bhat, 2011), and may have convergence issues for large-scale applications. Many empirical transportation modelers share the perception that the use of the GHK simulator is somewhat prohibitive due to its computational cost (cf. Connors et al., 2012), despite the encouraging results found in some studies – including Geweke et al. (1997), who analyze a problem with 20 dimensions, and Ziegler (2012), where the most flexible model estimated involves 36-dimensional integrals. Convergence failure adds to this perception, which may be explained by the problem of weak identification encountered in panel (multi-period) multinomial probit models (Ziegler and Eymann, 2001), or in a single-period probit with individual covariates only (Keane, 1992). On the other hand, the Bayes probit estimator is analytically straightforward (Albert and Chib, 1993) and has proven to perform better in estimation than maximum simulated likelihood (Geweke et al., 1994, 1997). The basic idea is that data augmentation allows for treating the model as an ordinary regression. Even though several authors have analyzed the Bayes estimator of the multinomial probit model (McCulloch and Rossi, 1994, 2000; Bolduc et al., 1997; McCulloch et al., 2000; Nobile, 2000), applications in transportation are rather limited (Kim et al., 2003). In fact, Bayesian discrete choice, especially in modeling travel behavior, lags well behind Bayesian developments in other fields.

In this paper we contribute to the empirical literature of statistical inference in discrete choice modeling by showing how the Bayes estimates of a multinomial probit model can be combined with the GHK simulator to compute the posterior distribution of the probit choice probabilities, and how this posterior provides measures of uncertainty regarding the true choice probabilities. Note that point estimates of the choice probabilities are virtually never reported with confidence intervals in practice. Finding confidence intervals for nonlinear transformations of the structural parameters – such as the choice probabilities – is a highly complex problem in frequentist econometrics. Thus, in this paper we argue that finding credible intervals for the market shares is the key benefit of the use of Bayes estimators in fully identified models. Regarding perceived complexity and computational cost of the maximum simulated likelihood estimator of a multinomial probit, we show that both the frequentist and Bayes estimators produce results that are very similar for medium-scale problems with static data. However, the Bayes estimator is superior to the maximum likelihood estimator because it allows the researcher to directly analyze the posterior distribution of functions of the structural parameters of the model, including choice probabilities and market shares.

In addition, we add to the literature on consumer adoption of energy-intensive durable goods by addressing uncertainty in the forecasts of the model via the provision of exact credible intervals of the choice probabilities and market shares. The specific empirical case study in this paper is based on stated-preference data on vehicle choice in Germany. Different analyses have been performed using the data coming from this Germany-wide survey of potential light-duty-vehicle buyers using computer-assisted personal interviewing (Achtnicht, 2012; Achtnicht et al., 2012; Ziegler, 2012). For example, Achtnicht et al. (2012) analyze the effect of fuel availability on demand for alternative-fuel vehicles using the same choice data, focusing on marginal probability effects as well as on the determination of willingness to pay for increased fuel availability as derived from a standard conditional logit model.¹ However, the present study is the first to use this data for both deriving and analyzing market-share forecasts. Thus, we apply the combination of the Bayes probit

¹Note that the conditional (multinomial) logit model imposes proportional substitution patterns, a restriction that we withdraw in this paper.

estimates and the recursive probability simulator for constructing a Markov chain of market shares. We then find credible intervals of the market shares of differing scenarios of service stations for the new energy-efficient vehicle technologies.

The rest of the paper is organized as follows. In section 2 we discuss more details about the frequentist and Bayes estimators of the multinomial probit model, and introduce a method for combining the GHK simulator and the Bayesian Gibbs sampler. The vehicle choice data is described in section 3. Results of the estimation of the parameters of the multinomial probit model are displayed in section 4. In section 5 we use the estimates to produce forecasts to analyze the effect of increasing the network of service stations for charging electric vehicles as well as for refueling hydrogen. Section 6 concludes.

2 Frequentist and Bayesian inference in the multinomial probit model

2.1 Multinomial probit choice probabilities

Consider the following multinomial probit model of individual i choosing alternative $j_i \in \{1, \dots, J\}$

$$\mathbf{U}_i \underset{(J \times 1)}{=} \underset{(J \times K)(K \times 1)}{\mathbf{X}_i \boldsymbol{\beta}} + \underset{(J \times 1)}{\boldsymbol{\varepsilon}_i} \quad (1)$$

$$y_i \underset{(1 \times 1)}{=} j_i \text{ iff } U_{ij_i} = \max_j U_{ij}, \quad (2)$$

where choices are based on maximization of the random utility vector \mathbf{U}_i ; the deterministic component of utility is assumed linear in the vector of unknown parameters of the model $\boldsymbol{\beta}$; \mathbf{X}_i is a matrix of exogenous hedonic attributes with row j equal to \mathbf{x}'_{ij} ;² the error term has a multivariate normal distribution $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \forall i$; and y_i is a choice indicator that reveals preferences.

Because of the discrete maximization that represents choice (see Train, 2009), only parameters of the model in differences with respect to an arbitrary base alter-

²Such that the scalar utility of alternative j can be written as $U_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$.

native can be identified. Consider the estimable model in differences with respect to alternative j :

$$\begin{matrix} \Delta_j \mathbf{U}_i \\ (J-1 \times 1) \end{matrix} = \begin{matrix} \Delta_j \mathbf{X}_i & \boldsymbol{\beta} \\ (J-1 \times K) & (K \times 1) \end{matrix} + \begin{matrix} \Delta_j \boldsymbol{\varepsilon}_i \\ (J-1 \times 1) \end{matrix}, \Delta_j \boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}_{(J-1)}, \Delta_j \boldsymbol{\Sigma} \Delta_j') \quad (3)$$

$$\begin{matrix} y_i \\ (1 \times 1) \end{matrix} = \begin{matrix} j_i \text{ iff } \Delta_j U_{ij_i} = \max_j \Delta_j U_{ij}, \end{matrix} \quad (4)$$

where Δ_j is a $(J - 1 \times J)$ matrix difference operator with elements defined as

$$[\Delta_j]_{lm} = \begin{cases} -1 & \text{if } m = j \\ 1 & \text{if } l = m \text{ and } m \neq j \\ 0 & \text{otherwise} \end{cases} . \quad (5)$$

The choice probability P_{ij} of individual i choosing (any) alternative j takes the following form:

$$\begin{aligned} P_{ij} &= \Pr(U_{ij} \geq U_{ij'}, \forall j' \neq j) = \Pr(\varepsilon_{ij'} - \varepsilon_{ij} \leq (\mathbf{x}_{ij'} - \mathbf{x}_{ij})' \boldsymbol{\beta}, \forall j' \neq j) \\ &= \int_{-\infty}^{(\mathbf{x}_{iJ} - \mathbf{x}_{ij})' \boldsymbol{\beta}} \cdots \int_{-\infty}^{(\mathbf{x}_{i1} - \mathbf{x}_{ij})' \boldsymbol{\beta}} f(\Delta_j \boldsymbol{\varepsilon}_i) d\Delta_j \boldsymbol{\varepsilon}_i, \end{aligned} \quad (6)$$

where

$$f(\Delta_j \boldsymbol{\varepsilon}_i) = \frac{1}{(2\pi)^{\frac{J-1}{2}} |\Delta_j \boldsymbol{\Sigma} \Delta_j'|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \boldsymbol{\varepsilon}_i' \Delta_j' \Delta_j \boldsymbol{\Sigma} \Delta_j' \Delta_j \boldsymbol{\varepsilon}_i \right\}. \quad (7)$$

Note that the multinomial probit choice probability P_{ij} is an integral of dimension $J - 1$ that lacks a closed form. Numerical integration of the choice probabilities, including Gaussian quadrature methods, is feasible only for low dimensions.³

2.2 Maximum simulated likelihood estimator

Large scale multinomial probit models can be estimated using simulation-based inference. Probit estimates can be found using the method of simulated moments (McFadden, 1989), the method of simulated scores (Hajivassiliou and McFadden, 1998), or

³In general, numerical integration is feasible for up to three dimensions, i.e. a model with four alternatives.

maximum simulated likelihood. A maximum simulated likelihood estimator (MSLE) of the parameters of the model can be derived as $(\hat{\beta}, \hat{\Sigma})_{\text{MSLE}} = \arg \max \sum_i \ln \tilde{P}_{ij_i}$, where j_i is the alternative actually chosen by individual i and \tilde{P}_{ij_i} is a simulated choice probability.⁴ An advantage of the MSLE is that only the choice probability of the chosen alternative is considered.⁵ Any choice probability simulator can be used to calculate the simulated likelihood function, but an importance sampler performs better than a frequency simulator. Among the class of importance samplers, the smooth-recursive-conditioning GHK simulator due to Geweke (1991); Hajivassiliou and McFadden (1998) and Keane (1994) is the most popular choice in empirical work. The GHK simulator is continuous and differentiable, which is an advantage for finding the optimum. However, the loglikelihood of a multinomial probit model is not globally concave, making more involved the search for the optimum. In addition, the simulator needs to be run at every iteration of the optimization process. Additionally, consistency of the estimator requires not only a large sample, but also a large number of replications. In fact, even though the GHK simulator is unbiased for the choice probabilities, for a finite number of repetitions of the simulator the MSLE is biased.

2.3 Combining multinomial probit Bayes estimates and the GHK simulator for forecasting

Based on the work of Albert and Chib (1993), McCulloch and Rossi (1994) proposed a multinomial probit Bayes estimator that is analytically straightforward and avoids the problems of the MSLE. The Bayes estimator exploits the distribution of the reduced form of the structural system defined by equations (3) and (4). In effect, the choice indicator in equation (4) truncates the distribution of the random utility of equation (3). Thus, conditional on the cone defined by y_i , the indirect utility has a truncated normal distribution. This conditional distribution is the core of the Gibbs sampler of McCulloch and Rossi (1994): by augmenting the data, samples

⁴Note that optimization of the simulated loglikelihood requires differentiation of the simulated choice probabilities.

⁵A simulator based on the method of simulated moments requires evaluation of the choice probabilities for the whole choice set.

of the utility function are drawn and then used as observations of the dependent variable of equation (3); then, parameters of the model can be estimated using the Gibbs sampler for an ordinary regression. In sum, at iteration (g) of the Gibbs sampler, a truncated normal draw $\Delta_j \mathbf{U}_i^{(g)}$ is generated. This draw enters equation (3), which becomes the following ordinary regression problem $\Delta_j \mathbf{U}_i^{(g)} = \Delta_j \mathbf{X}_i \boldsymbol{\beta} + \Delta_j \boldsymbol{\varepsilon}_i$, where the dependent variable is no longer latent. Because the original Gibbs sampler uses prior distributions on unidentified parameters (see Nobile, 2000), McCulloch et al. (2000) updated their estimator taking into account the necessary normalization constraints. In this paper we use the sampler proposed in Imai and van Dyk (2005), which also considers priors on the identified parameters but with a better rate of convergence than the method of McCulloch et al. (2000).

Although the estimation problem seems to be dominant in theoretical research, the problem of prediction is as relevant for empirical research (see Connors et al., 2012). Forecasting with discrete choice models involves analyzing the choice probabilities after a qualitative change. Thus, once the posterior distributions of the taste parameters $\boldsymbol{\beta}$ and the nuisance parameters $\boldsymbol{\Sigma}$ have been found, forecasting with the model requires evaluation of the choice probabilities at different levels of the hedonic attributes. In the case of the multinomial probit model, this evaluation can be computed using the GHK recursive probability simulator. Note that use of the GHK simulator for forecasting and estimation is different. Whereas for forecasting the parameters of the model are given by the estimates, for estimation the parameters are unknown. For instance, it is the combination of the GHK simulator and the maximization of the loglikelihood function that produces simulation bias. However, the GHK simulator remains an unbiased estimator of the choice probabilities when the parameters of the model are given.

In this paper, to derive the posterior distribution of the probit choice probabilities we propose to use the GHK simulator to postprocess the Gibbs sampling estimates (cf. Edwards and Allenby, 2003). Specifically, we propose running the GHK simulator for every sample of the posterior generated at every iteration of the sampler of Imai and van Dyk (2005).

Consider the Markov chain Monte Carlo sample of both $\boldsymbol{\beta}^{(g)}$ and $\Delta_j \boldsymbol{\Sigma}^{(g)} \Delta_j'$ generated by the Bayes estimator at iteration (g) of the multinomial probit sampler.

Let \mathbf{C} be the Cholesky root of $(\Delta_j \Sigma \Delta_j')^{-1}$, with elements $[\mathbf{C}]_{lm}^{(g)} = c_{lm}^{(g)}$ for each sample of the posterior of the nuisance parameters $\Delta_j \Sigma^{(g)} \Delta_j'$. Suppose in addition that a qualitative change in the attributes is captured by the matrix $\mathbf{X}_i^{(1)}$.⁶ The simulated element (g) of the posterior distribution of the choice probability P_{ij} is then

$$\tilde{P}_{ij}^{(g)} = \frac{1}{S} \Phi \left(-\frac{(\mathbf{x}_{i1}^{(1)} - \mathbf{x}_{ij}^{(1)})' \boldsymbol{\beta}^{(g)}}{c_{11}^{(g)}} \right) \times \sum_{s=1}^S \Phi \left(-\frac{(\mathbf{x}_{i2}^{(1)} - \mathbf{x}_{ij}^{(1)})' \boldsymbol{\beta}^{(g)} + c_{21}^{(g)} \boldsymbol{\eta}_i^{(s)}}{c_{22}^{(g)}} \right) \dots \Phi \left(-\frac{(\mathbf{x}_{iJ}^{(1)} - \mathbf{x}_{ij}^{(1)})' \boldsymbol{\beta}^{(g)} + [\text{vech} \mathbf{C}^{(g)}]' \boldsymbol{\eta}_i^{(s)}}{c_{J-1, J-1}^{(g)}} \right), \quad (8)$$

where $\boldsymbol{\eta}$ is a vector of (J-1) iid standard normal terms such that $\mathbf{C} \boldsymbol{\eta}_i = \Delta_i \boldsymbol{\varepsilon}_i$, $\Phi(\cdot)$ is the CDF of a standard normal, and $\boldsymbol{\eta}_i^{(s)} = \{\eta_{i1}^{(s)}, \dots, \eta_{i, J-1}^{(s)}\}$ denotes realization (s) $\in \{1, \dots, S\}$ of a vector of independent random draws of the elements in $\boldsymbol{\eta}_i$. Equation (8) is the GHK approximation of the recursive decomposition of choice probability by means of its empirical expectation.

Thus, using the samples $\boldsymbol{\beta}^{(g)}$ and $\mathbf{C}^{(g)}$ of the Gibbs sampler, and repeating this procedure for all g , it is possible to build a sequence of iterative random draws that forms an irreducible and ergodic Markov chain converging at an exponential rate to the posterior distribution of the choice probabilities (and market shares). The Bayes point estimate of the choice probability P_{ij} is $\hat{P}_{ij} = (1/G) \sum_{g=1}^G \tilde{P}_{ij}^{(g)}$, where G is the total number of repetitions in the recursive sampler of Imai and van Dyk (2005). In addition, the posterior distribution can be used to account for uncertainty in the determination of the choice probabilities and market shares through the derivation of high posterior density (HPD) credible intervals, which are the Bayesian counterpart of confidence intervals.

Finally, note that the posterior choice probabilities that are derived using the GHK simulator are not standard predictive posteriors, in the sense that the calculation does not directly take into account the choice indicators. Predictive poste-

⁶Row j of matrix $\mathbf{X}_i^{(1)}$ is given by $\mathbf{x}_{ij}^{(1)}$, such that the new deterministic utility of alternative j experienced by individual i is $\mathbf{x}_{ij}^{(1)'} \boldsymbol{\beta}$.

rior probabilities can be derived by Monte Carlo approximation of $P_{ij}(\mathbf{X}_i^{(1)}, y_i) = \int P_{ij}(\mathbf{X}_i^{(1)}, y_i, \boldsymbol{\beta}, (\mathbf{C}\mathbf{C}')^{-1})p(\boldsymbol{\beta}, (\mathbf{C}\mathbf{C}')^{-1}|\mathbf{y})d\boldsymbol{\theta}$, where $p(\boldsymbol{\beta}, (\mathbf{C}\mathbf{C}')^{-1}|\mathbf{y})$ is the posterior distribution, and $\boldsymbol{\theta} = (\boldsymbol{\beta}', [\text{vech}\mathbf{C}'])'$.

3 Vehicle choice data

The stated preference data used in this paper comes from a Germany-wide survey of potential car buyers that was administered between August 2007 and March 2008 as a computer-assisted personal interview (CAPI). The survey was designed to garner insights into consumer preferences for alternative-fuel vehicles (see Achtnicht, 2012; Ziegler, 2012; Achtnicht et al., 2012). A total of approximately 600 interviews were conducted at various car dealerships and branch offices of TÜV, the German agency responsible for certifying vehicle roadworthiness. The respondents were picked randomly, but had to be of legal age and possess a valid driver's license. The sample comprises individuals from different regions in Germany (eastern and western Germany, urban and rural areas) and various demographic and socioeconomic groups (in terms of age, gender, education, income, etc.). It thus provides a broad cross-section of the target population, i.e. potential car buyers in Germany, although it is not entirely representative. Compared with the official data available from KBA (2009) and MiD (2010), it seems that more educated individuals are over-represented, whereas women and individuals aged 40 to 49 years are under-represented in the sample; see table 1 for more details.

Table 1: Sample demographics.

Survey question	Sample (N=598)	Population
Gender		
Male	74.6	69.0
Female	25.4	31.0
Age		
29 or below	20.7	17.7
30–39	21.1	19.9
40–49	20.2	28.2
50–59	17.7	19.4
60 or above	20.2	14.8
Education		
Secondary modern school degree	17.1	24.0
High school degree	31.1	33.2
University of applied sciences entrance qualification	8.0	9.5
Higher education entrance qualification, university or college degree	43.5	31.3
(Yet) without school degree or others	0.3	2.0
Household’s monthly net income		
€1,000 or below	3.3	
€1,000–2,000	18.4	
€2,000–4,000	37.1	
€4,000 or above	22.6	
Not stated	18.6	

Source: KBA (2009); MiD (2010); own calculations

Note: The population shares for gender and age are based on car owner data including all registrations of new and used cars in Germany in 2008 (KBA, 2009). The population shares for education represent the distribution among people with a car-driver’s license, based on a representative survey on mobility in Germany (MiD, 2010). To the authors’ knowledge, there are no data on the income distribution of the target population (i.e. potential car buyers from Germany) available.

In the survey, respondents participated in a choice experiment involving various alternative-fuel vehicles. In each choice set, respondents were presented with seven hypothetical vehicles and asked to select the car they preferred most. The alternatives were characterized by the following six attributes: purchase price; fuel costs per 100 km; engine power; CO₂ emissions per km; fuel availability (given by the service station network size); and fuel type.⁷ Respondents were asked to assume that the

⁷The 7×6 choice set design used in this survey was relatively demanding for respondents. However, based on the results of a pretest, the survey team at that time concluded that the experimental design was appropriate and not overly challenging. For a more detailed discussion of the issue of choice complexity, see Achtnicht (2012), which uses the same data set.

presented hypothetical alternatives only differed with regard to these attributes, but were otherwise identical (e.g., in terms of vehicle size and design). Table 2 gives details on the attribute levels. To examine potential alternative-specific effects related to fuel type, each fuel was included once in each choice set (thus “labeling” the choice experiment). The attributes “purchase price” and “engine power” were customized. Respondents were asked beforehand to describe the vehicle they intended to buy, indicating upper and lower bounds for price and horsepower, which were then averaged and used as an individual reference or pivot. This pivot or customization approach is common in the transportation literature and it increases the relevancy of attribute levels and choice scenarios (e.g., Hensher, 2010; Hensher et al., 2005).

Table 2: Attributes and attribute levels for the vehicle choice experiment.

Attribute	Levels
Fuel type	Gasoline, Diesel, Hybrid, LPG/CNG, Biofuel, Hydrogen, Electric
Purchase price	75%, 100%, 125% of reference ^a (in €)
Engine power	75%, 100%, 125% of reference ^a (in hp)
Fuel costs per 100 km	€5, €10, €20
CO ₂ emissions per km	no emissions ^b , 90 g, 130 g, 170 g, 250 g
Fuel availability	20% ^c , 60%, 100% of service station network

^a average of the lower and upper bounds for the next car indicated by the respondent

^b only applied to non-fossil fuel types (i.e. biofuel, hydrogen, and electric)

^c not applied to conventional fuel types (i.e. gasoline and diesel)

In the choice experiment, the attribute levels varied independently between alternatives and choice sets. This ensured that each attribute’s impact on choice selection could be isolated. However, in order to avoid the inclusion of unrealistic scenarios, only positive emissions were allowed for fossil fuels (i.e. gasoline, diesel, CNG/LPG), and the lowest fuel availability level (i.e. 20%) was excluded for conventional-fuel alternatives.⁸ The final fractional factorial design of the choice experiment, which

⁸According to Moore and Holbrook (1990), the degree to which attribute-level combinations are realistic is of less practical importance than is sometimes feared. Moore and Holbrook analyzed the effect of unrealistic stimuli on consumer judgements in terms of perceived realism and predictive power in three experiments in a car choice context. Their results provide evidence that the choice likelihoods are not affected by differences in scenario realism.

was generated using Sawtooth software, required respondents to evaluate six choice sets.

4 Model specification and estimation

The result of Bayesian estimation of a discrete choice model is the posterior distribution of the parameters. However, the Bayesian framework also offers an answer to the point estimation problem. In table 3 we present the Bayes point estimates of the multinomial probit model, which correspond to the mean of the posterior distribution. Since the Bayes estimator of a multinomial probit model is a Gibbs sampler, the point estimates are the empirical mean of the draws of the Markov chain. We also present the results of the maximum simulated loglikelihood estimator using the GHK simulator (MSLE-GHK).

Table 3: Point estimates of the multinomial probit model of vehicle choice.

Variable	MSLE-GHK		Gibbs sampler	
	Estimate	Standard error	Estimate	Standard error
Purchase price [€1000]	-0.0141***	0.0019	-0.0131***	0.0018
Fuel costs [€/100 km]	-0.0297***	0.0030	-0.0272***	0.0029
Fuel availability [%]	0.0050***	0.0005	0.0046***	0.0005
Engine power [HP]	0.0025***	0.0004	0.0023***	0.0003
CO ₂ emissions [g/km]	-0.0015***	0.0002	-0.0014***	0.0002
LPG/CNG	-0.2575***	0.0828	-0.2214***	0.0797
Hybrid	-0.1197	0.0801	-0.0903	0.0704
Electric	-0.2916***	0.0862	-0.2714***	0.0829
Biofuel	-0.2576***	0.0853	-0.2351***	0.0848
Hydrogen	-0.0895	0.0643	-0.1053	0.0659
Diesel	-0.0760	0.0714	-0.0663	0.0661
Observed choices			3588	
Individuals			598	
Simulated loglikelihood		-6117.9		-6113.7
Pseudo ρ^2		0.125		0.125
CPU time [sec]		1384.7		1081.7

Note: Asterisks denote statistical significance at the *** p< 0.01, ** p< 0.05, * p< 0.1 level. CPU time in a personal computer with a 2.93 GHz Quad-Core processor and 8 GB RAM.

For the Gibbs sampler we used a chain of 50,000 iterations which took roughly

18 minutes. For MSLE-GHK, we used 250 repetitions for each evaluation of the choice probabilities. Using the BHHH approximation of the Hessian, the MSLE-GHK converged after 14 iterations of the search algorithm (taking about 23 minutes). Lower standard errors in the Bayes estimates are found with respect to MSLE-GHK for all vehicle attributes and almost all alternative specific constants. We note that the frequentist estimator was sensitive to the selection of the base alternative. Convergence problems – due to a singular BHHH estimate – were detected for specific choices of the base.

Because we assumed a linear specification of the indirect utility, the parameters of the model represent marginal utilities that can be described as fixed taste parameters. Buyers of new vehicles obtain less satisfaction when a car comes with an elevated price tag. More expensive variable costs – which are related to fuel costs – also reduce utility. An interesting result is the negative marginal utility of carbon dioxide emissions. This result shows that prospective buyers care about the environmental externalities of personal transportation, and they prefer vehicles that produce less pollution. More power is a desired feature, as can be seen from the associated positive marginal utility. Another appreciated attribute is availability of fuel. If the specific fuel is readily available, then some of the reliability issues of low-emission vehicles are resolved and consumers are more satisfied. In discrete choice modeling, these levels of satisfaction or dissatisfaction translate into higher or lower choice probabilities. For example, consumers are more likely to choose a car that is relatively cheap, with inexpensive fuel or an energy-efficient engine that reduce operating costs, with a dense station network, good horsepower, and reduced CO₂ emissions.

If all of the considered attributes were the same among all of the alternatives, then gasoline and diesel vehicles would be preferred. This is indicated by the alternative specific constants. Electric vehicles turn out to be the least preferred. It is possible that respondents associated electric vehicles with long charging times, short ranges, or other disadvantages (cf. Achtnicht et al., 2012). Note that the alternative specific constants capture the average effect of the omitted variables.

Table 4 presents the quantile estimates of each marginal utility. These quantiles are a summary of the joint posterior distribution. Note that the values that concentrate 95% of the mass can be used to determine the credible intervals for each taste

parameter, i.e. the 2.5% quantile represents the lower bound and the 97.5% quantile the upper bound of the 95% credible interval. These values can be used directly for hypothesis testing.

Table 4: Bayes posterior quantile estimates.

Variable	Quantile estimates (Gibbs sampler)				
	2.5%	5%	50%	95%	97.5%
Purchase price	-0.0170	-0.0143	-0.0130	-0.0118	-0.0097
Fuel costs	-0.0333	-0.0294	-0.0273	-0.0252	-0.0221
Fuel availability	0.0037	0.0042	0.0046	0.0049	0.0056
Engine power	0.0016	0.0020	0.0022	0.0025	0.0030
CO ₂ emissions	-0.0018	-0.0015	-0.0014	-0.0013	-0.0011
LPG/CNG	-0.3849	-0.2755	-0.2196	-0.1695	-0.0769
Hybrid	-0.2354	-0.1284	-0.0788	-0.0323	0.0528
Electric	-0.4834	-0.3370	-0.2751	-0.2220	-0.1265
Biofuel	-0.4106	-0.2778	-0.2211	-0.1672	-0.0726
Hydrogen	-0.2376	-0.1387	-0.0912	-0.0480	0.0238
Diesel	-0.2064	-0.0991	-0.0514	-0.0081	0.0632

For a multinomial probit model, the parameter space is completed with the nuisance parameters associated with the elements of the covariance matrix of the model in differences. (Table 5 contains point estimates and standard deviations of the elements of the matrix $(\mathbf{C}\mathbf{C}')^{-1}$.) We allowed flexible substitution patterns through a fully flexible covariance structure, and from the point estimates it is possible to see the presence of heteroskedasticity and different correlation levels.

Table 5: Bayes point estimates of the covariance matrix $(\mathbf{CC}')^{-1}$.

Variable	LPG/CNG _{gas}	Hybrid _{gas}	Electric _{gas}	Biofuels _{gas}	Hydrogen _{gas}	Diesel _{gas}
LPG/CNG _{gas}	1.00					
	-					
Hybrid _{gas}	0.45*** (0.11)	0.69*** (0.19)				
Electric _{gas}	0.41*** (0.12)	0.44*** (0.15)	0.78*** (0.22)			
Biofuels _{gas}	0.29** (0.14)	0.31*** (0.12)	0.50*** (0.15)	0.65*** (0.18)		
Hydrogen _{gas}	0.43*** (0.10)	0.38*** (0.13)	0.43*** (0.0.15)	0.30*** (0.12)	0.69*** (0.19)	
Diesel _{gas}	0.41*** (0.12)	0.14 (0.12)	0.35*** (0.14)	0.25** (0.12)	0.31*** (0.12)	0.77*** (0.20)

Note: Model in difference with respect to gasoline. (The cell (Hybrid_{gas}, Diesel_{gas}) represents element $[\Delta_{gas}\Sigma\Delta'_{gas}]_{hybrid,diesel}$.) Standard errors (posterior standard deviations) in parentheses. Asterisks denote statistical significance at the *** p< 0.01, ** p< 0.05, * p< 0.1 level.

5 Forecasting

5.1 Experimental market shares

It is well known that parameters of a simple conditional (multinomial) logit model are such that the observed and predicted attribute average are the same. Thus, in a conditional logit model with alternative-specific constants the predicted market shares reproduce by construction the observed market shares. Because this property of the conditional logit model does not extend to the multinomial probit model, we are interested in determining whether the multinomial probit is able to reproduce the market shares. In table 6 the observed and predicted market shares are displayed. Note that in the case of the stated-preference data, the observed or experimental market shares are the percentages that are directly derived from the stated choices.

Table 6: Experimental and predicted market shares [%]

Vehicle type	Observed	MSLE-GHK	Bayes-GHK	Bayes-GHK quantile estimates					Predictive
	shares	point est.	point est.	2.5%	25%	50%	75%	97.5%	probs.
Gasoline	19.5	19.7	19.5	18.2	19.1	19.5	19.9	20.8	19.7
LPG/CNG	12.2	12.3	12.2	11.2	11.9	12.2	12.6	13.3	12.2
Hybrid	12.7	12.7	12.7	11.6	12.3	12.7	13.0	13.7	12.7
Electric	8.7	8.7	8.6	7.8	8.3	8.6	8.9	9.5	8.6
Biofuels	11.0	10.9	10.9	9.9	10.6	10.9	11.2	11.9	10.9
Hydrogen	15.1	15.1	15.1	14.0	14.7	15.1	15.5	16.2	15.0
Diesel	20.9	20.8	21.0	19.7	20.5	21.0	21.5	22.5	20.8

As discussed in section 2, a clear advantage of the Bayes estimator is that the sample of the posterior distribution, simulated via Markov chain Monte Carlo methods, can be used to generate the posterior distribution of any function of the original parameters of the model. Since the choice probabilities are a function of the marginal utilities, to derive the Bayes-GHK estimates we determined first the posterior distribution of the choice probabilities and then the posterior distribution of the aggregate choices in the form of market shares. More specifically, for every observation in the sample and for every MCMC draw of the Bayes estimator, we ran 250 repetitions of the GHK simulator. (The computational cost of the GHK simulator in forecasting is very low; evaluation of the 7 choice probabilities for a single individual took 0.006 seconds.) Not only are the experimental market shares within the 95% credible interval, but the point estimates also replicate the observed values almost perfectly. Note that credible intervals for the market shares are tight, although the coefficient of variation of the credible intervals of some of the alternative-specific constants is relatively high. We complete table 6 with the point estimates of both MSLE-GHK and the predictive posterior probabilities. Computational efficiency of the MSLE-GHK is the same than for the Bayes-GHK point estimates (both use the GHK simulator). The difference is that the MSLE-GHK is evaluated just once at the MSLE point estimates, whereas the Bayes estimates are evaluated at every draw of the posterior distribution of the marginal utilities. (The latter evaluation provides information for deriving the credible intervals and quantile estimates.) In the case of the posterior predictive probabilities, a single evaluation for one individual took 0.117 seconds.

5.2 Adoption scenarios

A problem with experimental market shares of stated-preference studies, as opposed to observed shares in real markets of revealed-preference data, is that stated choices are a response to the experimental attribute variation. Thus, little can be said about the competitiveness of the different alternatives. Because of these limitations, we decided to examine the behavior of a representative individual faced with a scenario of vehicle attributes intended to represent an average of the current vehicle choice situation in Germany. This base scenario is summarized in table 7. Both sources and assumptions are discussed below.

Table 7: Base scenario: average vehicle choice in Germany

Vehicle type	Attributes				
	Purchase price [€]	Fuel costs [€/100km]	Fuel availability [%]	Engine power [HP]	CO ₂ emissions [g/km]
Gasoline	19558	7.86	100	100	143
LPG/CNG	21240	4.69	42	100	116
Hybrid	22739	5.90	100	100	107
Electric	34897	4.00	3.5	100	0
Biofuels	19895	7.34	2.3	100	20
Hydrogen	27474	5.00	0.1	100	0
Diesel	20735	6.38	100	100	146

Assumed values for purchase price and in-use CO₂ emissions⁹ are taken from the 2015 scenario of the research project “Trends in Energy Markets until 2030 – Energy Forecast 2009”.¹⁰ Fuel consumption data, also taken from this project, was used to derive average fuel costs. Here we assume a gasoline price of €1.31 per liter, a diesel price of €1.16 per liter, and that one liter of LPG costs about half as much, and biofuel about two thirds as much as gasoline, as is currently the case in Germany. Because the vehicle-choice survey was conducted mainly in 2007, all monetary values of table 7 are adjusted to 2007 euros using the German consumer price index provided by the German Federal Statistical Office. Note further that

⁹Emissions occurring during fuel production were not taken into account here.

¹⁰This project aimed to establish a consistent set of realistic scenarios for the long-term evolution of energy-resource supply and demand in Germany, using the TIMES PanEU energy system model.

our assumptions regarding biofuel are based on E85, which consists of 85% ethanol and 15% gasoline. The assumed fuel costs for hydrogen and electric cars are based on results from the “GermanHy” study (BMVBS, 2009), financed by the German Federal Ministry of Transport, and a recent McKinsey study (McKinsey, 2010). The fuel availability data reflects the German status quo. Today, there are approximately 15,000 service stations (including freeway service stations) in Germany. Based on an online search, we found that LPG/CNG can be refueled at 6,280/892 service stations, biofuel (E85) at 345, hydrogen at 8, while for electric cars there are 512 charging stations available. However, we could not find any reliable average data for current or expected engine power. Therefore, we decided to ignore possible differences in engine power and use 100 HP for each fuel type. Of course, all assumed figures are tentative and should be treated with caution.

Consumer adoption of ultra-low emission vehicles depends on adequate provision of refueling or recharging infrastructure.¹¹ Thus, combining both the base scenario of table 7 and the multinomial probit Bayes point estimates of tables 3 and 5 we produce forecasts to analyze the effect of increasing the density of the service station network. In particular, we study the effects on market shares of increases in the density of service stations required for charging electric vehicles as well as of those for refueling hydrogen-powered vehicles.¹²

Table 8 summarizes the posterior distribution of the aggregate choice probabilities of the different vehicle types. We start with the base scenario, i.e. we produce market shares that represent choices of representative consumers when faced to the attribute levels of table 7. Given the results of the previous subsection, the reported results correspond to the Bayes estimates using the GHK simulator (Bayes-GHK).

¹¹It has been argued that the lack of an appropriate service station infrastructure is a major barrier for the adoption of ultra-low emission vehicles (Bunch et al., 1993; Daziano and Bolduc, 2011; Achtnicht et al., 2012; Daziano and Chiew, 2012). Understanding how consumers react to qualitative improvements in the service station network is necessary for planning the corresponding infrastructure investments.

¹²When analyzing differing scenarios of fuel availability, Achtnicht et al. (2012) focus on marginal probability effects rather than on forecasting market shares, and thus use the same price, fuel costs, engine power, and CO₂ emissions for all vehicles in their scenarios. Additionally, whereas we assume flexible substitution patterns via a multinomial probit model, the marginal probability effects in Achtnicht et al. (2012) are based on a conditional logit model.

Therefore, to obtain the market shares for each MCMC draw of the joint posterior of the probit parameters and given the average attribute levels we calculate the choice probabilities of each alternative. For this, 200 repetitions of the GHK simulator are performed. Once this procedure has been repeated for every sample, we obtain the joint posterior distribution of the market shares. In the table the posterior distribution is summarized presenting its mean, which is equivalent to the point estimate, and its standard deviation, which can be used as an analog to frequentist standard error.

As a second step, we vary the density of the charging network for electric vehicles. In table 8 results for densities equal to 10%, 30%, 50%, 70%, and 100% are reported, holding everything else constant (and equal to the base scenario). For example, the market shares of the upper 10%-column are given by the choice probabilities of a representative consumer facing the same attribute levels as in the base scenario, except that the density of electric charging infrastructure has gone up from 3.5% to 10%. The joint posterior of the market shares for this situation is obtained following the same procedure used for the base scenario. Then, we perform the same exercise for the density of the hydrogen fueling network. In Appendix B the information of table 8 is supplemented with a summary of the posterior distribution of the market shares of more extensive scenarios varying service station density for both electric and hydrogen vehicles.

Table 8: Forecasted market shares (Bayes-GHK estimates)

Vehicle type	Base shares [%]	Density of the Electric Vehicle charging network				
		10%	30%	50%	70%	100%
Gasoline	21.98 (1.22)	21.94 (1.22)	21.79 (1.22)	21.57 (1.19)	21.29 (1.18)	20.75 (1.15)
LPG/CNG	11.43 (0.97)	11.41 (0.97)	11.40 (0.93)	11.33 (0.91)	11.16 (0.89)	10.71 (0.88)
Hybrid	23.97 (1.46)	23.86 (1.45)	23.42 (1.36)	22.96 (1.32)	22.39 (1.28)	21.23 (1.26)
Electric	2.78 (0.55)	3.12 (0.58)	4.40 (0.66)	6.00 (0.77)	8.00 (0.89)	11.97 (1.19)
Biofuels	6.01 (0.74)	5.95 (0.72)	5.62 (0.68)	5.37 (0.65)	5.08 (0.63)	4.55 (0.64)
Hydrogen	7.28 (0.88)	7.24 (0.87)	7.17 (0.82)	7.00 (0.80)	6.79 (0.78)	6.37 (0.77)
Diesel	26.55 (1.65)	26.48 (1.63)	26.15 (1.63)	25.77 (1.60)	25.29 (1.59)	24.42 (1.60)

Vehicle type	Base shares [%]	Density of the Hydrogen refueling network				
		10%	30%	50%	70%	100%
Gasoline	21.98 (1.22)	21.79 (1.22)	21.29 (1.18)	20.70 (1.15)	20.00 (1.13)	18.77 (1.14)
LPG/CNG	11.43 (0.97)	11.46 (0.93)	11.18 (0.90)	10.85 (0.86)	10.47 (0.85)	9.78 (0.86)
Hybrid	23.97 (1.46)	23.57 (1.37)	22.85 (1.32)	22.01 (1.28)	21.03 (1.25)	19.37 (1.23)
Electric	2.78 (0.55)	2.76 (0.53)	2.64 (0.50)	2.50 (0.48)	2.35 (0.46)	2.10 (0.44)
Biofuels	6.01 (0.74)	5.81 (0.70)	5.65 (0.68)	5.46 (0.67)	5.25 (0.65)	4.86 (0.65)
Hydrogen	7.28 (0.88)	8.35 (0.89)	10.70 (0.94)	13.50 (1.00)	16.74 (1.10)	22.44 (1.33)
Diesel	26.55 (1.65)	26.26 (1.64)	25.70 (1.64)	24.99 (1.62)	24.16 (1.67)	22.68 (1.74)

Note: Standard errors in parentheses. All estimates statistically significant at the $p < 0.01$ level.

As expected, a more dense service station network clearly increases consumer adoption of the low-emission technologies. For instance, a charging infrastructure that matches the density of standard gasoline stations produces a remarkable 331% increase in the market share of electric vehicles as compared with the base scenario.¹³ In this ideal situation, and for the hypothesized adoption scenario, the market share

¹³The base situation assumes a density of 3.5%.

of electric vehicles is forecasted to reach almost 12% of the market, with the upper bound of the 95% credible interval at 14.38% and the lower bound at 9.75%. Due to an expected lower purchase price for hydrogen vehicles as well as to the difference in alternative-specific constants, the base market share is larger for hydrogen vehicles than for electric vehicles. From a base level of 7.28%, even with an extremely low fuel availability density of 0.1%, the market share of hydrogen vehicles goes up to 22.44% with a fully competitive refueling network, with upper and lower bounds of the 95% credible interval of the market shares at 19.88% and 25.13%, respectively.

As a general characteristic of discrete choice models, the elasticity of demand with respect to changes in fuel availability is not constant.¹⁴ This can be seen in the curves shown in figure 1, which depict the point estimate and 95% credible interval of the market shares under increased fuel availability.¹⁵

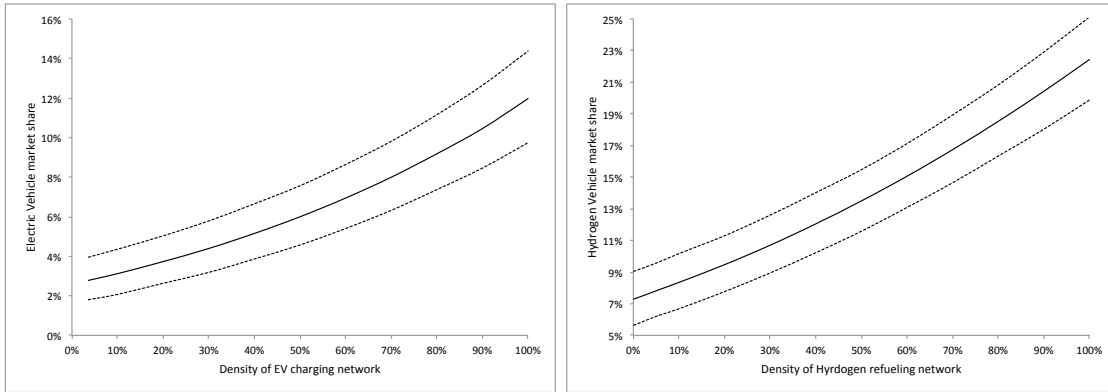


Figure 1: Point estimates and 95% credible interval bounds of the market shares of electric (left) and hydrogen (right) vehicles.

The average elasticity of the market share of electric vehicles is 0.07 in the 0-10% density interval, 0.65 in the 40-50% interval, and 1.30 in the 90-100% interval. In the case of hydrogen vehicles, the average elasticity is 0.07 in the 0-10% density

¹⁴In fact, using a linear specification, initial infrastructure investments have a low impact on the penetration of the energy efficient vehicles.

¹⁵Note that in discrete choice models point estimates of the market shares are usually reported without confidence intervals, whereas our suggested method for postprocessing the Bayes estimators has proven to facilitate the derivation of credible intervals.

interval, 0.48 in the 40-50% interval, and 0.89 in the 90-100% interval. These measures are relevant for planning both public and private investments in the infrastructure necessary to promote and ensure adequate consumer adoption of energy-efficient vehicle technologies.

Another relevant outcome of the modeling strategy adopted in this paper is the different degree of competition among vehicle types, which is a result of our assumption of a multinomial probit model with full covariance matrix. For example, the increase of the market share of electric vehicles from the base 2.78% with a charging network density of 3.5% to 6.00% when the density achieves 50% is explained by a decrease of 1.87% in the market share of gasoline vehicles, 0.90% of LPG/CNG, 4.21% of hybrids, 10.58% of biofuels, 3.79% of hydrogen, and 2.95% of diesel. If we repeat the same exercise for hydrogen vehicles, the deeper penetration of hydrogen when the refueling density achieves 50% is accompanied by a decrease of 5.85% in the market share of gasoline vehicles, 5.12% of LPG/CNG, 8.17% of hybrids, 10.07% of electric, 9.19% of biofuels, and 5.86% of diesel. It is noteworthy that when electric vehicles become more competitive, the largest relative changes occur in the consumer switch from hybrids and biofuel to electricity propelled vehicles.¹⁶

Although our results suggest a potentially large penetration of hydrogen vehicles, in practice hydrogen vehicles are not yet commercially available and the required infrastructure investments are larger than those needed for charging electric vehicles. Not only will new, dedicated fueling stations be needed for fuel-cell vehicles, but also substantial investments in production, distribution, and storage of hydrogen fuel. At the other extreme, however, electric batteries can, if necessary, be charged using regular outlets.¹⁷

6 Summary and conclusion

In this paper we have shown how Bayesian econometrics allows modelers to revisit estimation of the multinomial probit model, not only for the point estimation prob-

¹⁶Hybrid vehicles exhibit the largest absolute decrease.

¹⁷However, the provision of fast charging stations is necessary for ensuring reasonable charging times.

lem but also for forecasting. Although the frequentist probit maximum simulated likelihood estimator with the GHK simulator is feasible for static discrete choice problems,¹⁸ there are clear advantages of using a Bayes estimator instead of the maximum simulated likelihood estimator. Not only is the Bayes estimator gradient- and hessian-free, but it is also free of simulation bias.¹⁹ In addition, we have shown that the full flexibility of the multinomial probit model can be exploited in practice through the Bayes estimator, and that once the model has been estimated the GHK simulator can be used for evaluating the choice probabilities. Because the result of the Bayesian estimation process is not merely a point but rather the whole posterior distribution of the parameters, we show that a key feature of the Bayes estimates is the derivation of the posterior distribution of the choice probabilities. In the procedure proposed here the GHK recursive probability simulator is run for every draw of the posterior distribution of both the parameters of interest and the nuisance parameters of the multinomial probit model, i.e. to construct a Markov chain of samples of the posterior of the choice probabilities we propose to postprocess the parameter posterior via Monte Carlo simulation. The resulting posterior distribution of the multinomial probit choice probabilities and market shares can then be used to obtain credible intervals that account for uncertainty regarding the true value of the random data generating process. Having better tools to address uncertainty is particularly relevant in the context of modeling consumer response to emerging energy-efficient technologies.

In our case study of consumer adoption of ultra-low-emission vehicles in Germany, we first showed that the Bayes point estimates of the market shares reproduce the shares given by the stated choices. Then, we produced forecasts for a representative individual based on a scenario of vehicle attributes that aims at representing an average of the current vehicle choice situation in Germany. Because limited fuel availability is a major obstacle to consumer adoption of low-emission vehicles, we have also analyzed the effect of increasing the density of the network of service stations

¹⁸We actually show that the frequentist estimates are a very good approximation of the Bayesian results for medium-scale problems. Computational cost for both estimators is of the same order.

¹⁹The probit Bayes estimator has also the potential to overcome the MSLE-GHK convergence problems in panel versions of the probit model due to weak identification. We leave for further research the application of Bayes estimators of a multiperiod multinomial probit model in forecasting.

for charging electric vehicles as well as for refueling hydrogen-fueled vehicles. The result is the posterior distribution of the choice probabilities that represent adoption of energy-efficient technologies in the context of a more competitive infrastructure. For example, our results indicate that if availability of charging is increased to its maximum, electric vehicles would experience a greater than three-fold increase in market penetration.

However, the mere adoption of ultra-low-emission vehicles is not sufficient to make personal transport more sustainable. The electricity needed to power electric cars and to produce hydrogen through electrolysis has to be generated somehow. Many of the climate-damaging emissions would then be shifted from the transport sector to the energy sector. Given the current electricity mix in Germany, with roughly 20% renewables, about 550g of CO₂ are emitted per generated kWh.²⁰ The production of biofuels is also emission-intensive, especially if rainforests are cut down to gain cropland, and has a negative impact on food prices and security. Hence, the road to greener transport is rocky and it will not be paved by simply expanding the fueling station infrastructure.

Acknowledgements

We thank the Institute for Energy Economics and the Rational Use of Energy (IER) in Stuttgart for providing the scenario data used as basis for the forecasting. In addition, this research is based upon work supported by the National Science Foundation Faculty Early Career Development CAREER Award No. CBET-1253475.

References

- Achtnicht M. (2012). German car buyers' willingness to pay to reduce CO₂ emissions. *Climatic Change* 113(3): 679-697.
- Achtnicht M., Bühler G., Hermeling C. (2012). The impact of fuel availability on

²⁰Note that emissions from the German energy sector are covered by the European Union Emissions Trading System.

- demand for alternative-fuel vehicles. *Transportation Research Part D: Transport and Environment* 17(3): 262-269.
- Albert J., Chib S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88: 669-679.
- Ben-Akiva M., Bolduc D. (1996). Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure, Working Paper, Massachusetts Institute of Technology.
- Bhat C.R. (2011). The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. *Transportation Research Part B* 45(7): 923-939.
- BMVBS (German Federal Ministry of Transport) (2009). GermanHy – “Woher kommt der Wasserstoff in Deutschland bis 2050?”. Berlin.
- Bolduc D., Ben-Akiva M. (1991). A Multinomial Probit Formulation for Large Choice Sets, *Proceedings of the 6th International Conference on Travel Behaviour* 2: 243-258.
- Bolduc D., Fortin B., Gordon S. (1997). Multinomial probit estimation of spatially interdependent choices: an empirical comparison of two new techniques. *International Regional Science Review* 20: 77-101.
- Brownstone D., Bunch D.S., Golob T.F., Ren W. (1996). Transactions choice model for forecasting demand for alternative-fuel vehicles. *Research in Transportation Economics* 4: 87-129
- Brownstone D., Train K.E. (1999). Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics* 89: 109-129
- Brownstone D., Bunch D.S., Train K.E. (2000). Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles. *Transportation Research Part B: Methodological* 34(5): 315-338

- Bunch D.A. (1991). Estimability in the Multinomial Probit Model. *Transportation Research Part B* 25: 1-12.
- Bunch D., Bradley M., Golob T., Kitamura K. (1993). Demand for Clean-Fuel Vehicles in California: A Discrete-Choice Stated Preferences Pilot Study. *Transportation Research Part A: Policy and Practice* 27(3): 237-253
- Connors R.D., Hess S., Daly A. (2012). Analytic approximations for computing probit choice probabilities. *Transportmetrica* DOI: 10.1080/18128602.2012.702794
- Daganzo C. (1979). *Multinomial Probit: the Theory and its Application to Demand Forecasting*. Academic Press, New York.
- Dansie B.R. (1985). Parameter Estimability in the Multinomial Probit Model. *Transportation Research B* 19(6): 526-528.
- Daziano R.A., Bolduc D. (2011). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica* (forthcoming).
- Daziano R.A., Chiew E. (2012). Electric vehicles rising from the dead: data needs for forecasting consumer response toward sustainable energy sources in personal transportation. *Energy Policy* 51: 876-894
- Edwards Y., Allenby G. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research* 40: 321-334.
- Geweke J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints, in E. M. Keramidas, ed., *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, Interface Foundation of North America, Inc., Fairfax: 571-578.
- Geweke J.F., Keane M.P., Runkle D.E. (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics* 76: 609-632.

- Geweke J.F., Keane M.P., Runkle D.E. (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics* 80: 125-165.
- Hajivassiliou V., McFadden D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica* 66: 863-896.
- Hensher D.A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B* 44, 735-752.
- Hensher D.A., Beck M.J., Rose J.M. (2011). Accounting for preference and scale heterogeneity in establishing whether it matters who is interviewed to reveal household automobile purchase preferences. *Environmental and Resource Economics* 49: 1-22
- Hensher D.A., Rose J.M., Greene W.H. (2005). *Applied choice analysis: A primer*. Cambridge University Press, Cambridge.
- Horne M., Jaccard M., Tiedemann K. (2005). Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. *Energy Economics* 27: 59-77
- Imai K., van Dyk D.A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics* 124: 311-334.
- KBA (Kraftfahrt-Bundesamt) (2009). Fachartikel: Halter der Fahrzeuge. 2. November 2009. Flensburg.
- Keane, M. (1992). A Note on Identification in the Multinomial Probit Model. *Journal of Business and Economic Statistics* 10: 193-201
- Keane M. (1994). A computationally practical simulation estimator for panel data. *Econometrica* 62: 95-116.
- Kim Y., Kim T.Y., Heo E. (2003). Bayesian estimation of multinomial probit models of work trip choice. *Transportation*, 30: 351-365.
- McCulloch R., Rossi P. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64: 207-240.

- McCulloch R., Rossi P. (2000). Bayesian analysis of the multinomial probit model, in R. Mariano, T. Schuermann, and M. Weeks, eds., *Simulation-Based Inference in Econometrics*, Cambridge University Press, New York.
- McCulloch R.R., Polson N.G., Rossi, P.E. (2000). Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99: 173-193.
- McFadden D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5): 995-1026.
- McFadden D.L., Train K.E. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics* 15(5):447-470.
- McKinsey & Company (2010). A portfolio of power-trains for Europe: a fact-based analysis. The role of Battery Electric Vehicles, Plug-in Hybrids and Fuel Cell Electric Vehicles.
- MiD (2010). *Mobilität in Deutschland 2008: Tabellenband*. Bonn and Berlin.
- Moore W.L., Holbrook M.B. (1990). Conjoint analysis on objects with environmentally correlated attributes: The questionable importance of representative design. *Journal of Consumer Research* 16: 490-497.
- Munizaga M.A., Alvarez Daziano R. (2005). Testing mixed logit and probit by simulation. *Transportation Research Record: Journal of the Transportation Research Board* 1921: 53-62.
- Nobile A. (2000). Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics*, 99: 335-345.
- Thurstone L. (1927). A Law of Comparative Judgement. *Psychological Review* 34: 273-286.
- Train K. (2009) *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, second edition.

Ziegler A. (2012). Individual characteristics and stated preferences for alternative energy sources and propulsion technologies in vehicles: A discrete choice analysis. *Transportation Research Part A: Policy and Practice* 46(8): 1372-1385.

Ziegler A., Eymann A. (2001). Zur Simulated Maximum Likelihood Schätzung von Mehrperioden-Mehralternativen-Modellen. *Allgemeines Statistisches Archiv* 85: 319-342

A Details of the probit Gibbs sampler (adapted from Imai and van Dyk, 2005)

- Start at any given point $\Delta_j \mathbf{U}_i^{(0)}$, $\beta^{(0)}$, and $(\Delta_j \Sigma \Delta_j')^{(0)}$ in the parameter space, and set $\alpha^{(0)} = 1$.
- Consider the prior distributions $p(\beta) \sim \mathcal{N}(\check{\beta}, \check{V}_\beta)$ and

$$p(\Delta_j \Sigma \Delta_j') \propto |\Delta_j \Sigma \Delta_j'|^{-(\nu+J)/2} [\text{trace}(\mathbf{S}(\Delta_j \Sigma \Delta_j')^{-1})]^{-\nu(J-1)/2},$$

where S is the prior scale of $\Delta_j \Sigma \Delta_j'$ and ν is the prior for the degrees of freedom of the covariance.

- Compute the Cholesky root $\mathbf{C}^{(0)}$ of $(\Delta_j \Sigma \Delta_j')^{-1}$
- For $g \in \{1, \dots, G\}$

1. For all i , if $y_i = j$, draw $\Delta_j \mathbf{U}_i^{(g)}$ from the truncated normal distribution

$$\mathcal{N}\left(\Delta_j \mathbf{X}_i \beta^{(g-1)}, (\mathbf{C}^{(g-1)} \mathbf{C}^{(g-1)'})^{-1}\right) 1(\Delta_j U_{ij'} < 0, \forall j' \neq j),$$

otherwise draw $\Delta_j \mathbf{U}_i$ from the truncated normal distribution

$$\mathcal{N}\left(\Delta_j \mathbf{X}_i \beta, (\mathbf{C}^{(g-1)} \mathbf{C}^{(g-1)'})^{-1}\right) 1(\Delta_j U_{ij'} > \max\{0, \Delta_j U_{i,-j}\}, \forall j' \neq j).$$

2. Update the scale parameter $\alpha^{(g)^2}$ using the distribution

$$\alpha_0^2 \text{trace}(\mathbf{S} \mathbf{C}^{(g-1)} \mathbf{C}^{(g-1)'}) / \chi_{\nu(J-1)}^2.$$

3. Compute the Cholesky root $\mathbf{C}^{(g)}$ of \mathbf{W}/w_{11} , where w_{11} is the first element in the diagonal of the matrix \mathbf{W} , which in turn is a draw from the Wishart distribution

$$\text{Wishart}\left(\nu + N, \alpha_0^2 \mathbf{S} + \alpha^{(g)^2} \sum_{i=1}^N (\Delta_j \mathbf{U}_i^{(g)} - \Delta_j \mathbf{X}_i \beta^{(g-1)}) (\Delta_j \mathbf{U}_i^{(g)} - \Delta_j \mathbf{X}_i \beta^{(g-1)})'\right).$$

4. Draw $\beta^{(g)}$ from the normal distribution

$$\mathcal{N}\left((\check{V}_\beta^{-1} \check{\beta} + (\mathbf{C}^{(g)'} \mathbf{X})' \mathbf{C}^{(g)'} \mathbf{X})^{-1} (\check{V}_\beta^{-1} + \mathbf{X}' \mathbf{C}^{(g)} \mathbf{C}^{(g)'} \Delta_j \mathbf{U}^{(g)}), (\check{V}_\beta^{-1} + \mathbf{C}^{(g)'} \mathbf{X}' (\mathbf{C}^{(g)'} \mathbf{X})^{-1})\right).$$

5. Step 4 completes the sampler for estimation. For forecasting, update the GHK approximation $\tilde{P}_{ij}^{(g)}, \forall j$ using equation (8).
6. Update $g = g + 1$ and go back to step 1.

Note that this sampler is for a static discrete choice model. Analysis of a sampler that accounts for correlations due to taste persistence or memory effects across the choice sets is left for further research.

B Forecasts

Table 9: Summary of the posterior distribution of the Electric Vehicle share

Charging network density	Mean [%]	Bayes quantile estimates of the market shares [%]				
		2.5%	25%	50%	75%	97.5%
3.5%	2.78	1.80	2.40	2.75	3.13	3.95
10%	3.12	2.07	2.72	3.09	3.49	4.36
20%	3.73	2.64	3.30	3.70	4.13	5.04
30%	4.40	3.19	3.94	4.36	4.82	5.79
40%	5.16	3.86	4.66	5.12	5.61	6.66
50%	6.00	4.57	5.47	5.97	6.50	7.57
60%	6.95	5.40	6.38	6.91	7.49	8.64
70%	8.00	6.32	7.38	7.97	8.58	9.82
80%	9.18	7.37	8.52	9.15	9.83	11.16
90%	10.46	8.46	9.72	10.44	11.15	12.65
100%	11.97	9.75	11.15	11.94	12.76	14.38

Table 10: Summary of the posterior distribution of the Hydrogen Vehicle share

Refueling network density	Mean [%]	Bayes quantile estimates of the market shares [%]				
		2.5%	25%	50%	75%	97.5%
0.1%	7.28	5.64	6.67	7.25	7.85	9.05
5%	7.82	6.20	7.21	7.79	8.39	9.57
10%	8.35	6.68	7.73	8.32	8.94	10.17
20%	9.47	7.77	8.84	9.45	10.07	11.30
30%	10.70	8.95	10.05	10.67	11.32	12.60
40%	12.05	10.23	11.38	12.02	12.69	14.02
50%	13.50	11.59	12.81	13.47	14.17	15.47
60%	15.06	13.09	14.35	15.01	15.75	17.12
70%	16.74	14.65	15.98	16.72	17.47	18.93
80%	18.53	16.34	17.74	18.50	19.30	20.83
90%	20.43	18.03	19.60	20.42	21.25	22.91
100%	22.44	19.88	21.53	22.42	23.32	25.13