

Discussion Paper No. 12-017

**Forecasting Adoption of
Ultra-low-emission Vehicles Using the
GHK Simulator and Bayes Estimates
of a Multinomial Probit Model**

Ricardo A. Daziano and Martin Achtnicht

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 12-017

**Forecasting Adoption of
Ultra-low-emission Vehicles Using the
GHK Simulator and Bayes Estimates
of a Multinomial Probit Model**

Ricardo A. Daziano and Martin Achnicht

Download this ZEW Discussion Paper from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/dp12017.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other
economists in order to encourage discussion and suggestions for revisions. The authors are solely
responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non-technical Summary

In the European Union, transport is the largest consumer of oil products and second largest emitter of carbon dioxide (CO₂); within the sector, road transport dominates in both regards. 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.

Using stated-preference data on vehicle choice from a Germany-wide survey of potential light-duty-vehicle buyers using computer-assisted personal interviewing, in this paper we analyze market shares of different automotive technologies produced by a discrete choice model with flexible substitution among different fuel types. Effectively, we propose a methodology to use the estimates of a probit model to produce both market-share forecasts as well as Bayesian confidence intervals for the forecasted shares. These forecasts are simulated from the posterior distribution of a Bayesian model and account for uncertainty. Having better tools to address uncertainty is particularly relevant in the context of modeling consumer response to emerging energy-efficient technologies.

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. Because limited fuel availability is a major obstacle to consumer adoption of low-emission vehicles, we analyze the specific effect of increasing the density of the network of service stations for charging electric vehicles as well as for refueling hydrogen-fueled vehicles. 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.

Das Wichtigste in Kürze

Der Transportsektor ist innerhalb der Europäischen Union der größte Konsument von Ölprodukten und der zweitgrößte Emittent von Kohlenstoffdioxid (CO_2), wobei der Straßenverkehr in beiderlei Hinsicht dominierend ist. Die Verlagerung der Nachfrage hin zu besonders emissionsarmen oder -freien Fahrzeugen wird im Allgemeinen als Möglichkeit angesehen, einen nachhaltigen Personenverkehr voranzutreiben. Auch wenn neue Niedrigemissions-Technologien, wie beispielsweise batteriebetriebene Elektrofahrzeuge, herkömmlichen Technologien in puncto Energieeffizienz und Abgasemissionen deutlich überlegen sind, so verhindern doch verschiedene Probleme deren umfassende Einführung. Zum einen sind Elektrofahrzeuge in der Anschaffung viel teurer als vergleichbare Benziner. Zum anderen werfen beschränkte und schwankende Reichweiten sowie fehlende Tankstellen für Konsumenten Fragen nach der Zuverlässigkeit solcher Fahrzeuge auf.

Unter Verwendung von Daten einer deutschlandweiten Befragung von potentiellen Pkw-Käufern, analysieren wir in diesem Papier die Marktanteile unterschiedlicher Fahrzeugtechnologien auf Basis hypothetischer Kaufentscheidungen. Wir schlagen hierbei eine Methodik vor, aus der Posterior-Verteilung der Bayesgeschätzten Parameter eines flexiblen Probitmodells sowohl Prognosen über Marktanteile als auch die zugehörigen Bayes'schen Konfidenzintervalle zu generieren, und damit deren Unsicherheit auf geeignete Art und Weise zu erfassen. Diese Vorgehensweise ist insbesondere bei der Modellierung der Nachfrage nach neuen, noch nicht am Markt etablierten Technologien von großer Bedeutung.

In einem Basisszenario versuchen wir die gegenwärtige Marktsituation für deutsche Autokäufer darzustellen. Dazu nutzen wir, je nach Art des Kraftstoffs, unterschiedliche Durchschnittswerte für bestimmte Fahrzeugattribute wie Kaufpreis oder CO_2 -Ausstoß. Anschließend untersuchen wir, wie sich die Nachfrage nach strom- und wasserstoffbetriebenen Fahrzeugen im Vergleich zum Basisszenario verändert, wenn das jeweils zugehörige Ladestationen- oder Tankstellennetz ausgebaut wird. Der Marktanteil von Elektrofahrzeugen würde sich laut unseren Ergebnissen zum Beispiel mehr als verdreifachen, wenn diese an jeder Tankstelle aufgeladen werden könnten.

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

Ricardo A. Daziano* and Martin Achnicht†

February 2012

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. 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, 2011).

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. However, using frequentist econometrics, flexibility comes at a cost. 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. The GHK recursive probability simulator has been proposed and tested for deriving a frequentist estimator of the parameters of the model (Geweke, 1991; Hajivassiliou and McFadden, 1998; Keane, 1994). A problem that emerges is that the maximum simulated likelihood estimator using the GHK simulator is computationally very expensive and becomes infeasible for both medium and large-scale problems. Conversely, 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. 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 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 in empirical applications of discrete choice models, point estimates of the choice probabilities are virtually never reported with confidence intervals. Finding confidence intervals for nonlinear transformations of the structural parameters – such as the choice probabilities – is a highly complex problem in frequentist econometrics. A related contribution of this work is the empirical application of the recursive probability simulator using the Bayes estimates of the posterior distribution of the parameters of a probit model of vehicle choice. In effect, 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, 2011; Achtnicht et al., 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 methodological contribution that we overview above – the combination of the Bayes probit 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.

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

2 Review of two approaches for estimation of the multinomial probit model

2.1 Probit choice probabilities

Consider the following multinomial probit model

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

$(J \times 1) \quad (J \times K)(K \times 1) \quad (J \times 1)$

$$y_i = j \text{ iff } U_{ij} = \max_{j'} U_{ij'}, \quad (2)$$

$(1 \times 1) \quad j \text{ iff } U_{ij} = \max_{j'} U_{ij'}$

where individual i chooses alternative $j \in \{1, \dots, J\}$ if the random utility vector \mathbf{U}_i has a maximal element in the j -th cell; the deterministic component is assumed linear in the 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 $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \Sigma), \forall i$; and y_i is a choice indicator that reveals preferences.

Because of the ordinal nature of the utility function, only parameters of the model in differences with respect to an arbitrary base alternative can be identified. Consider the estimable model in differences with respect to alternative j :

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

$(J-1 \times 1) \quad (J-1 \times K)(K \times 1) \quad (J-1 \times 1)$

$$y_i = \begin{cases} j & \text{iff } \Delta_j U_{ij'} < 0, \forall j' \neq j \\ j' & \text{iff } \Delta_j U_{ij'} > \max\{0, \Delta_j U_{i,-j}\}, \forall j' \neq j \end{cases}, \quad (4)$$

where Δ_j is a matrix difference operator; and $\Delta_j U_{i,-j}$ represents the set of all elements of \mathbf{U}_i with the exception of U_{ij} .

The choice probabilities that are derived from the estimable version of the probit model take the following form:

$$P_{ij} = \int_{-\infty}^{(\mathbf{x}_{iJ} - \mathbf{x}_{ij})' \boldsymbol{\beta}} \cdots \int_{-\infty}^{(\mathbf{x}_{i1} - \mathbf{x}_{ij})' \boldsymbol{\beta}} f(\Delta_j \varepsilon_{i1}, \dots, \Delta_j \varepsilon_{iJ}) d\Delta_j \varepsilon_{i1} \cdots d\Delta_j \varepsilon_{iJ}, \quad (5)$$

where P_{ij} is the probability of individual i choosing the alternative j , and where

²Such that $U_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \varepsilon_{ij}$.

$f(\Delta_j \boldsymbol{\varepsilon}_i)$ is the density of the multivariate normal $\mathcal{N}(\mathbf{0}_{(J-1)}, \Delta_j \Sigma \Delta_j')$:

$$f(\Delta_j \boldsymbol{\varepsilon}_i) = \frac{1}{(2\pi)^{\frac{J-1}{2}} |\Delta_j \Sigma \Delta_j'|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \boldsymbol{\varepsilon}_j' \Delta_j' \Delta_j \Sigma \Delta_j' \Delta_j \boldsymbol{\varepsilon}_j \right\}. \quad (6)$$

Note that the probit choice probability P_{ij} is an integral of dimension $J - 1$ without a closed form. Numerical integration of the choice probabilities, including Gaussian quadrature methods, is feasible only for low dimensions.³

2.2 The GHK simulator

The GHK simulator is based on Monte Carlo integration via recursive truncation of the normal distribution (Geweke, 1991; Hajivassiliou and McFadden, 1998; Keane, 1994). The simulator is continuous and differentiable, produces probabilities that lie strictly between zero and one, and manages low choice probabilities well. The recursive truncation of the simulator comes from the model in differences with transformed error terms using the Cholesky decomposition of $\Delta_j \Sigma \Delta_j'$. Let \mathbf{C} be the Cholesky root of $(\Delta_j \Sigma \Delta_j')^{-1}$. Then, $\Delta_j \mathbf{U}_i = \Delta_j \mathbf{X}_i \boldsymbol{\beta} + \mathbf{C} \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}$ is a vector of $(J - 1)$ iid standard normal terms.

For the chosen alternative, $P_{ij} = \Pr(\Delta_j U_{ijj'} < 0, \forall j' \neq j)$. Since \mathbf{C} is a lower triangular matrix, P_{ij} results in an expression that involves the recursive product of truncated normal distributions. For example, consider a choice situation with 3 alternatives where, without loss of generality, the third alternative was chosen. In this case, the probit choice probability P_{i3} can be rewritten as

$$P_{i3} = \Phi \left(-\frac{\Delta_3 \mathbf{x}'_{i1} \boldsymbol{\beta}}{c_{11}} \right) \int_{\eta_{1n} = -\infty}^{-\frac{\Delta_3 \mathbf{x}'_{i1} \boldsymbol{\beta}}{c_{11}}} \Phi \left(-\frac{\Delta_3 \mathbf{x}'_{i2} \boldsymbol{\beta} + c_{21} \eta_{1n}}{c_{22}} \right) \frac{\phi(\eta_{i1})}{\Phi \left(-\frac{\Delta_3 \mathbf{x}'_{i1} \boldsymbol{\beta}}{c_{11}} \right)} d\eta_{1n}, \quad (7)$$

where c_{11} , c_{21} , and c_{22} are the elements of the Cholesky root, and where $\Phi(\cdot)$ is the CDF of a standard normal.

In equation (7), the dimensionality of the choice probability integral of equation (5) has been reduced. However, there is still an integral to be evaluated. The GHK

³In general, numerical integration is feasible for up to three dimensions.

simulator proposes an approximation to (7) by means of its empirical expectation:

$$\tilde{P}_{i3} = \Phi\left(-\frac{\Delta_3 \mathbf{x}'_{i1} \boldsymbol{\beta}}{c_{11}}\right) \sum_{s=1}^S \Phi\left(-\frac{\Delta_3 \mathbf{x}'_{i2} \boldsymbol{\beta} + c_{21} \eta_{i1}^{(s)}}{c_{22}}\right), \quad (8)$$

where $\eta_{i1}^{(s)}$ is a random draw from a truncated normal distribution, truncated at $-\Delta_3 \mathbf{x}'_{i1} \boldsymbol{\beta} / c_{11}$. In general, the GHK simulator of P_{ij} takes the form

$$\tilde{P}_{ij} = \Phi\left(-\frac{\Delta_j \mathbf{x}'_{i1} \boldsymbol{\beta}}{c_{11}}\right) \sum_{s=1}^S \Phi\left(-\frac{\Delta_j \mathbf{x}'_{i2} \boldsymbol{\beta} + c_{21} \eta_{i1}^{(s)}}{c_{22}}\right) \dots \Phi\left(-\frac{\Delta_j \mathbf{x}'_{iJ} \boldsymbol{\beta} + [\text{vech} \mathbf{C}]' \boldsymbol{\eta}^{(s)}}{c_{J-1, J-1}}\right). \quad (9)$$

2.3 Maximum simulated likelihood estimator

The GHK simulator can be used to derive a maximum simulated likelihood estimator (MSLE) of the parameters of the model. Then, a frequentist solution to the estimation problem is $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})_{\text{MSLE}} = \arg \max \sum_i \ln \tilde{P}_{ij_i}$, where j_i is the alternative actually chosen by individual i . Note that optimization of the simulated loglikelihood requires differentiation of the simulated choice probabilities. The GHK simulator is continuous and differentiable, which is an advantage for finding the optimum. Another advantage of the MSLE with the GHK simulator is that only the choice probability of the chosen alternative is considered.⁴ However, the loglikelihood of a multinomial probit model is not globally concave, making the search for the optimum a nontrivial task. In addition, the simulator needs to be run at every iteration of the optimization process. As a result, MSLE still is computationally expensive for medium scale problems even with a restricted covariance matrix.⁵ The associated computational cost is high enough to prevent the general use of the probit model in empirical applications. Additionally, consistency of the estimator requires not only a large sample, but also $S \rightarrow \infty$. In fact, even though the GHK simulator is unbiased for the choice probabilities, for a finite S the MSLE is biased.

⁴A simulator based on the method of simulated moments requires evaluation of the choice probabilities for the whole choice set.

⁵In fact, nonconvergence is a rather common problem of the maximum simulated likelihood estimator of the probit parameters.

2.4 A Bayes estimator using Gibbs sampling

Based on the work of Albert and Chib (1993), McCulloch et al. (2000) proposed a 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 y_i , the indirect utility has the following truncated normal distribution:

$$\pi(\Delta_j \mathbf{U}_i | y_i) \sim \begin{cases} \mathcal{N}(\Delta_j \mathbf{X}_i \boldsymbol{\beta}, \Delta_j \Sigma \Delta_j') 1(\Delta_j U_{ij'} < 0, \forall j' \neq j) \\ \mathcal{N}(\Delta_j \mathbf{X}_i \boldsymbol{\beta}, \Delta_j \Sigma \Delta_j') 1(\Delta_j U_{ij'} > \max\{0, \Delta_j U_{i,-j}\}, \forall j' \neq j) \end{cases} . \quad (10)$$

This conditional distribution is the core of the Gibbs sampler for the multinomial probit model (see McCulloch et al., 2000). By augmenting the data, samples 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.⁶ In effect, at iteration (g) of the Gibbs sampler, a draw $\Delta_j \mathbf{U}_i^{(g)} \sim \pi(\Delta_j \mathbf{U}_i | y_i)$ is generated according to the truncated normal distribution of equation (10). This draw enters equation (3), which becomes the following ordinary regression problem

$$\Delta_j \mathbf{U}_i^{(g)} = \Delta_j \mathbf{X}_i \boldsymbol{\beta} + \Delta_j \boldsymbol{\varepsilon}_i, \quad (11)$$

where the dependent variable is no longer latent. Thus, a Bayes estimator of $\boldsymbol{\beta}$ is found by generating a draw from the following multivariate normal distribution

$$\boldsymbol{\beta}^{(g)} \sim \mathcal{N} \left((\check{V}_{\boldsymbol{\beta}}^{-1} \check{\boldsymbol{\beta}} + (\mathbf{C}^{(g)'} \mathbf{X})' \mathbf{C}^{(g)'} \mathbf{X})^{-1} (\check{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}' \mathbf{C}^{(g)} \mathbf{C}^{(g)'} \Delta_j \mathbf{U}^{(g)}), \right. \\ \left. (\check{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{C}^{(g)'} \mathbf{X}' (\mathbf{C}^{(g)'} \mathbf{X})^{-1}) \right), \quad (12)$$

where $\check{\boldsymbol{\beta}}$ and $\check{V}_{\boldsymbol{\beta}}$ are the parameters of the prior distribution $p(\boldsymbol{\beta}) \sim \mathcal{N}(\check{\boldsymbol{\beta}}, \check{V}_{\boldsymbol{\beta}})$. Note that equation (12) assumes that the Cholesky decomposition of the covariance

⁶In general, discrete choice models require special econometric tools because the dependent variable is latent. If the utility function were observable, standard regression techniques could be applied. Data augmentation in the Bayesian setting provides observations of the utility.

matrix of the model in differences is known. The Cholesky root $\mathbf{C}^{(g)}$ can be obtained from the draw (g) of the covariance matrix of the model $\Delta_j \Sigma \Delta_j'$. Taking into consideration the identification restrictions of the probit model (Dansie, 1985; Bunch, 1991), Nobile (2000) proposed to generate draws of $\Delta_j \Sigma \Delta_j'$ by sampling from an inverted-Wishart distribution

$$(\Delta_j \Sigma \Delta_j')^{(g)} \sim IW(\check{\nu} + N, \Delta_j \check{\Sigma} \Delta_j' + \sum_{i=1}^N \Delta_j \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \Delta_j')|_{c_{11}=1}, \quad (13)$$

where the $\check{\nu}$ and $\Delta_j \check{\Sigma} \Delta_j'$ are the parameters of the inverted-Wishart prior $p(\Delta_j \Sigma \Delta_j') = IW(\check{\nu}, \Delta_j \check{\Sigma} \Delta_j')$.

2.5 Combining the Bayes estimates and the GHK simulator

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 Σ 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. It is the combination of the GHK simulator and the maximization of the loglikelihood function that becomes computationally infeasible for medium and large-scale problems. However, the GHK simulator remains a feasible approximation of the choice probabilities when the parameters of the model are given.⁷

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

⁷In addition, good statistical properties of the GHK simulator are present for the computation of the choice probabilities. Unlike the MSLE with the GHK simulator, the simulated estimator of the choice probabilities is unbiased.

the Gibbs sampler outlined in the previous subsection.

Consider the Markov chain Monte Carlo sample generated by the Bayes estimator at iteration (g) of the probit Gibbs sampler. Suppose in addition that a qualitative change in the attributes is captured by the vector $\mathbf{X}^{(1)}$. The simulated element (g) of the posterior distribution of the choice probability P_{ij} is then

$$\tilde{P}_{ij}^{(g)} = \Phi \left(-\frac{\Delta_j \mathbf{x}_{i1}^{(1)'} \boldsymbol{\beta}^{(g)}}{c_{11}^{(g)}} \right) \times \sum_{s=1}^S \Phi \left(-\frac{\Delta_j \mathbf{x}_{i2}^{(1)'} \boldsymbol{\beta}^{(g)} + c_{21}^{(g)} \eta_{i1}^{(s)}}{c_{22}^{(g)}} \right) \dots \Phi \left(-\frac{\Delta_j \mathbf{x}_{iJ}^{(1)'} \boldsymbol{\beta}^{(g)} + [\text{vech} \mathbf{C}^{(g)}]' \boldsymbol{\eta}^{(s)}}{c_{J-1, J-1}^{(g)}} \right). \quad (14)$$

Thus, using the samples $\boldsymbol{\beta}^{(g)}$ and $\mathbf{C}^{(g)}$ of the Gibbs sampler, and repeating this procedure $\forall g$, 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. This posterior distribution can be used to account for uncertainty in the determination of the choice probabilities and market shares through the derivation of credible intervals, which are the Bayesian counterpart of confidence intervals.

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, 2011; Ziegler, 2011; 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

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 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.

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.⁸ Respondents were asked to assume that the presented hypothetical alternatives only differed with regard to these at-

⁸The 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 (2011), which uses the same data set.

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)

tributes, but were otherwise identical. 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).

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.⁹ The final fractional factorial design of the choice experiment, which was generated using Sawtooth software, required respondents to evaluate six choice sets.

⁹According 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.

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

Variable	Bayesian probit	
	Estimate	Standard error
Purchase price [€1000]	-0.0150***	0.0020
Fuel costs [€/100 km]	-0.0316***	0.0029
Fuel availability [%]	0.0053***	0.0005
Engine power [HP]	0.0026***	0.0004
CO ₂ emissions [g/km]	-0.0016***	0.0002
Gasoline	0.0049	0.0787
Hybrid	-0.0685	0.0832
LPG/CNG	-0.2208**	0.1172
Biofuels	-0.2261***	0.1021
Hydrogen	-0.0714	0.0820
Electric	-0.2737***	0.1020
Observed choices		4186
Individuals		598
Simulated loglikelihood		-6113.707
Pseudo ρ^2		0.125

Note: Asterisks denote statistical significance at the *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ level.

4 Model specification and estimation

The result of Bayesian estimation of a discrete model is the posterior distribution of the parameters. However, the Bayesian framework also offers an answer to the point estimation problem. In the following table we present the Bayes point estimates of the probit model, which correspond to the mean of the posterior distribution. Since the Bayes estimator of a probit model is a Gibbs sampler, the point estimates are the empirical mean of the draws of the Markov chain.

We used a chain of 50,000 iterations which took roughly 200 minutes. Note that we attempted to solve the frequentist maximum simulated likelihood estimator with a full covariance matrix, but the estimator failed to converge.¹⁰

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

¹⁰When imposing an independent and homoskedastic covariance matrix, both the Bayes estimator and the frequentist estimator produced undistinguishable point estimates.

Table 4: Posterior quantile estimates.

Variable	Bayesian probit				
	2.5%	5%	50%	95%	97.5%
Purchase price	-0.0191	-0.0184	-0.0149	-0.0117	-0.0111
Fuel costs	-0.0374	-0.0365	-0.0316	-0.0268	-0.0259
Fuel availability	0.0043	0.0044	0.0053	0.0062	0.0064
Engine power	0.0019	0.0020	0.0026	0.0032	0.0033
CO ₂ emissions	-0.0021	-0.0020	-0.0016	-0.0013	-0.0013
Gasoline	-0.1493	-0.1256	0.0039	0.1344	0.1608
Hybrid	-0.2412	-0.2112	-0.0642	0.0615	0.0828
LPG/CNG	-0.4776	-0.4232	-0.2142	-0.0432	-0.0143
Biofuels	-0.4421	-0.4036	-0.2190	-0.0735	-0.0443
Hydrogen	-0.2374	-0.2119	-0.0696	-0.0598	0.0861
Electric	-0.4825	-0.4498	-0.2680	-0.1137	-0.0842

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. Electric vehicles turn out to be the least preferred. Note that the alternative specific constant captures the effect of the omitted variables. The experimental design omitted variables such as driving range, which is negatively perceived in the case of electric vehicles due to the limitations of current electric vehicle batteries. Thus, driving range anxiety is one explanation for the low alternative specific constant of electric vehicles (cf. Achtnicht et al., 2012).

In the following table we present 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.

For a multinomial probit model, the parameter space is completed with the

Table 5: Point estimates of the covariance matrix in differences.

Variable	Δ_D Gas	Δ_D Hybrid	Δ_D LPG/CNG	Δ_D Biofuels	Δ_D Hydrogen	Δ_D Electric
$\Delta_{\text{Diesel}}\text{Gasoline}$	1.00					
$\Delta_{\text{Diesel}}\text{Hybrid}$	0.36** (0.19)	1.19*** (0.35)				
$\Delta_{\text{Diesel}}\text{LPG/CNG}$	0.78*** (0.16)	0.86*** (0.29)	1.58*** (0.38)			
$\Delta_{\text{Diesel}}\text{Biofuels}$	0.47*** (0.14)	0.49*** (0.23)	0.93*** (0.27)	1.05*** (0.28)		
$\Delta_{\text{Diesel}}\text{Hydrogen}$	0.60*** (0.13)	0.42* (0.25)	0.87*** (0.25)	0.82*** (0.23)	1.14*** (0.29)	
$\Delta_{\text{Diesel}}\text{Electric}$	0.55*** (0.13)	0.57*** (0.23)	0.91*** (0.26)	0.66*** (0.20)	0.60*** (0.20)	1.08*** (0.27)

Note: Model in difference with respect to Diesel. Standard errors in parentheses. Asterisks denote statistical significance at the *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ level.

nuisance parameters associated with the elements of the covariance matrix of the model in differences. 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.

5 Forecasting

5.1 Experimental market shares

It is well known that parameters of a simple logit model are such that the observed and predicted attribute average are the same. Thus, in a logit model with alternative-specific constants the predicted market shares reproduce by construction the observed market shares. Because this property of the logit model does not extend to the probit model, we are interested in determining whether the probit is able to reproduce the market shares. In the following table 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.

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

Table 6: Experimental and predicted market shares

Vehicle type	Observed shares [%]	Point estimates	Bayes quantile estimates of the market shares [%]				
			2.5%	25%	50%	75%	97.5%
Gasoline	19.5	19.5	18.2	19.1	19.5	19.9	20.8
LPG/CNG	12.2	12.2	11.2	11.9	12.2	12.6	13.3
Hybrid	12.7	12.7	11.6	12.3	12.7	13.0	13.7
Electric	8.7	8.6	7.8	8.3	8.6	8.9	9.5
Biofuels	11.0	10.9	9.9	10.6	10.9	11.2	11.9
Hydrogen	15.1	15.1	14.0	14.7	15.1	15.5	16.2
Diesel	20.9	21.0	19.7	20.5	21.0	21.5	22.5

the marginal utilities, 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 200 repetitions of the GHK simulator. 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.

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 an 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.

Assumed values for purchase price and in-use CO₂ emissions¹¹ are taken from the 2015 scenario of the research project “Trends in Energy Markets until 2030 – Energy Forecast 2009”.¹² Fuel consumption data, also taken from this project,

¹¹Emissions occurring during fuel production were not taken into account here.

¹²This project aimed to establish a consistent set of realistic scenarios for the long-term evolution of energy-resource supply and demand in Germany, using of the TIMES PanEU energy

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

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 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.¹³ Thus, combining both the base

system model.

¹³It 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). Understanding how consumers react to qualitative improvements in the service station network is necessary for planning the corresponding infrastructure investments.

scenario of table 7 and the 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.¹⁴

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. To obtain these 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 A 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.

As expected, a more dense service station network clearly increases consumer adoption of the low-emission technologies. For instance, a charging infrastruc-

¹⁴When 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.

ture 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.¹⁵ In this ideal situation, and for the hypothesized adoption scenario, the market share 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 models, the elasticity of demand with respect to changes in fuel availability is not constant.¹⁶ 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.¹⁷

¹⁵The base situation assumes a density of 3.5%.

¹⁶In fact, using a linear specification, initial infrastructure investments have a low impact on the penetration of the energy efficient vehicles.

¹⁷Note 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.

Table 8: Forecasted market shares

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 stastically significant at the $p < 0.01$ level.

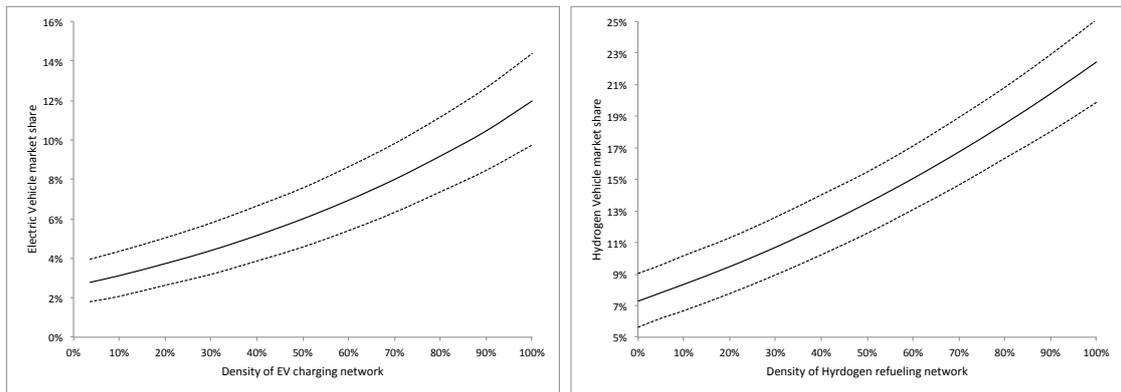


Figure 1: Point estimates and 95% credible interval bounds of the market shares of electric (left) and hydrogen (right) vehicles.

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 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.¹⁸

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.¹⁹

6 Summary and conclusion

In this paper we have shown how Bayesian econometrics allows modelers to revisit estimation of the multinomial probit model. There are several advantages of using a Bayes estimator instead of the maximum simulated likelihood estimator. Not only is the Bayes estimator feasible for higher dimensions, but it also permits the modeler to consider a full covariance matrix. In contrast, due to the computational cost of the frequentist estimator, practical applications of the GHK recursive probability simulator within the maximum simulated likelihood estimator have on the other hand typically used restricted covariance structures. We have shown that the full flexibility of the 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. Additionally, 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 the Bayes estimates can be used to obtain the posterior distribution of the choice probabilities as well. 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 probit choice probabilities and market shares can then be used

¹⁸Hybrid vehicles exhibit the largest absolute decrease.

¹⁹However, the provision of fast charging stations is necessary for ensuring reasonable charging times.

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 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.

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.

References

- Achtnicht M. (2011). German car buyers' willingness to pay to reduce CO₂ emissions. *Climatic Change*, doi: 10.1007/s10584-011-0362-8.
- Achtnicht M., Bühler G., Hermeling C. (2012). The impact of fuel availability on 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.
- 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 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.
- 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).
- 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.
- KBA (Kraftfahrt-Bundesamt) (2009). Fachartikel: Halter der Fahrzeuge. 2. November 2009. Flensburg.

- Keane M. (1994). A computationally practical simulation estimator for panel data. *Econometrica* 62, 95–116.
- 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.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). Conjoint analysis on objects with environmentally correlated attributes: The questionable importance of representative design. *Journal of Consumer Research* 16, 490–497.

Ziegler A. (2011). Individual characteristics and stated preferences for alternative energy sources and propulsion technologies in vehicles: A discrete choice analysis. In Revision for Transportation Research Part A.

A 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