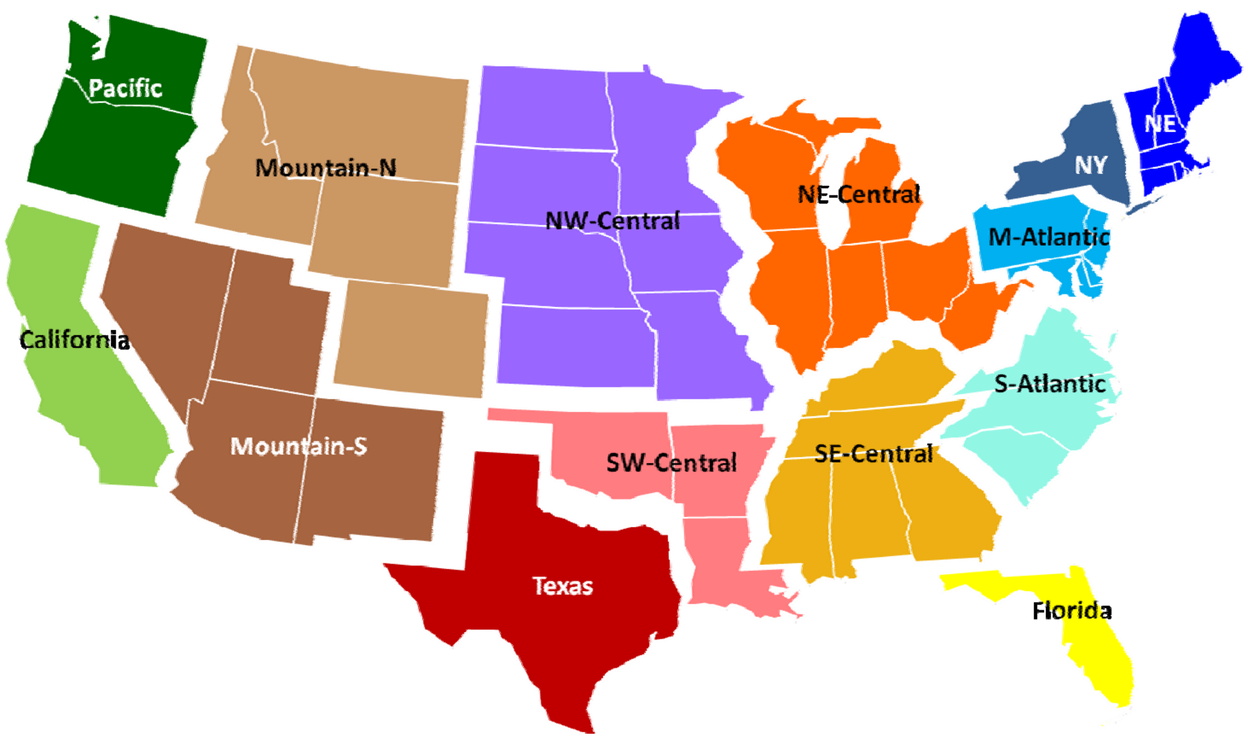
ELECTRIC POWER RESEARCH INSTITUTE

PRISM 2.0: Mixed Logit Consumer Vehicle Choice Modeling Using Revealed Preference Data

3002001455



PRISM 2.0: Mixed Logit Consumer Vehicle Choice Modeling Using Revealed Preference Data

3002001455

Technical Update, September 2013

EPRI Project Manager

J. Merrick

ELECTRIC POWER RESEARCH INSTITUTE
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ PO Box 10412, Palo Alto, California 94303-0813 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com

THE FOLLOWING ORGANIZATIONS, UNDER CONTRACT TO EPRI, PREPARED THIS REPORT:

Energy and Environmental Research Associates, LLC.

Environmental and Sustainable Technologies Lab., University of Michigan

This is an EPRI Technical Update report. A Technical Update report is intended as an informal report of continuing research, a meeting, or a topical study. It is not a final EPRI technical report.

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

ACKNOWLEDGMENTS

The following organizations, under contract to the Electric Power Research Institute (EPRI), prepared this report:

Energy and Environmental Research Associates, LLC.
2 Babcock Farms Lane
Pittsford, NY 14534

Principal Investigator
J. Winebrake

Environmental and Sustainable Technologies Lab - The University of Michigan
2250 G. G. Brown Bldg.
2350 Hayward St.
Ann Arbor, MI 48104

Principal Investigator
S. Skerlos in collaboration with V.Raichur

This report describes research sponsored by EPRI.

This publication is a corporate document that should be cited in the literature in the following manner:

PRISM 2.0: Mixed Logit Consumer Vehicle Choice Modeling Using Revealed Preference Data. EPRI, Palo Alto, CA: 2013, 3002001455.

ABSTRACT

Predicting the penetration of electric vehicles into the automotive market is challenging because these vehicles do not exist in the market today and therefore consumer reaction is largely unknown. One way to estimate consumer demand for electric vehicles is to model the attribute bundles of vehicles that are present in the market today and predict market share using state-of-the-art discrete choice demand models.

This research develops a choice-based demand model to extract consumer preferences from data available on historic vehicle purchases (revealed preference data). The report begins with an introduction to the discrete choice models in Section 1, followed by a literature review leading to the choice of a mixed logit model for this analysis. Section 2 describes the methods used and presents sample results. Section 3 presents the results from several counterfactual analyses intended to "stress test" the model. The results indicate that the mixed logit model characterized in this study may be an effective tool for predicting market demand for future vehicle designs and configurations.

Keywords

Consumer preferences
Discrete choice models
Mixed logit
PRISM 2.0
US-REGEN (U.S. Regional Economy, Greenhouse Gas, and Energy model)

EXECUTIVE SUMMARY

The purpose of this project is to advance EPRI's ability to estimate electric vehicle market penetration using quantitative consumer demand functions. These functions, when used in combination with vehicle attribute assumptions and in conjunction with other EPRI modeling systems, will allow EPRI to evaluate the economic, energy, and environmental impacts of electric vehicle technologies and policies.

In this work we discuss the development of a mixed logit vehicle demand model using historical revealed preference data. The model can be used to estimate market shares given vehicle attributes and customer demographics. We present the model structure as well as the model's parameter coefficients, and demonstrate how the model can be used for instance to estimate market shares under different scenarios for fuel price and hybrid vehicle prices. The main reason to create the model was for Veritas to use in a companion project that modeled future electric vehicle penetration into the market under alternative scenarios. The four test scenarios that were applied to test the mixed logit model in this work are as follows:

1. In-sample 2008: In this case we use data on vehicle attributes and household demographics for the year 2008 to estimate the coefficients of the utility function. Using these coefficients we calculate the market shares of vehicles for the same year to evaluate the accuracy with which the model can reproduce market shares. We find that the predicted and actual market shares match closely with correlation coefficient of 99%.
2. Out-of-sample 2007: In this second case we predict the vehicle market shares for the year 2007 using the coefficients estimated with data for the year 2008. We had to reduce the choice set used to estimate the coefficients so that it includes *only* vehicles that were available in the market in *both* 2007 and 2008. With the reduced choice set the correlation between estimated and actual market shares for year 2008 (in-sample) reduced from 99% to 92%. The estimated market shares for year 2007 were found to match actual market shares with 87% correlation. This leads to a conclusion that the model performs well even for out-of-sample predictions.
3. Counterfactual 1 – Gas prices: Following the successful analysis of in-sample and out-of-sample cases, we study two counterfactuals. In the first case we use change in gas prices as a trigger (holding all other factors constant) and study the change in market shares as compared to the year 2008. As expected, correlation between predicted and actual market shares decreases as gas prices are increased from 2008 levels. Among the top 15 vehicle models that lost market shares are mostly low mpg and larger footprint vehicle models. Among the top 15 models that gained market shares are higher mpg and smaller footprint vehicles. Hybrid sedans were among the top 3 models to gain market shares.
4. Counterfactual 2 – Hybrid vehicle prices: In the second counterfactual we focus on hybrid vehicles and study the influence of change in prices of these vehicles (compared to 2008) on the overall market shares. As expected, hybrids gain market shares as prices decrease. We

make a more interesting observation by comparing the market shares for conventional and hybrid versions of the same vehicle (Ford Escape for example). We observe that hybrid version's market share is twice that of the conventional version when both versions cost the same. The lower $/mile attribute of hybrid version is thus responsible for its greater market share even when both versions cost the same.

Through these scenarios we confirmed the model's ability to predict market shares for out-of-sample cases. The results therefore can be used to estimate consumer willingness to pay for certain attributes of electric vehicles (e.g., cost per mile to drive the car and acceleration) and therefore provides an important input to begin thinking about how the broader suite of electric vehicle attributes beyond those observed in today's market may diffuse into the market place as studied in the Veritas project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1
INTRODUCTION

Forecasting the demand for new products requires information about consumers' preferences for products that do not currently exist in the marketplace. Researchers have attempted to address this challenge by designing stated preference (SP) experiments to measure consumers' preferences over hypothetical alternatives including products that do not currently exist in the market. Stated preference experiments have been subject to considerable criticism by economists and other researchers since consumers react differently to hypothetical experiments than they would when facing the same alternatives in a real market. This problem is particularly acute when the new products incorporate public good attributes such as "zero-pollution" electric vehicles (EVs). Respondents may misrepresent their choices in SP experiments to strategically signal their preference for provision of the public good (less pollution), although in reality they would not spend extra money on purchasing an EV (1).

An alternative to the SP approach is the use of revealed preference (RP) data that looks at actual consumer behavior in the marketplace and from these data attempts to extract consumer preferences for certain product attributes. This study takes this RP approach, building a consumer demand model exclusively on RP data - in this case historical observed vehicle purchases. While RP approach has limitations regarding its ability to capture vehicle attributes not currently observable in the market, the approach can adequately capture basic trade-offs that exist in today's market and will persist even with the introduction of electric vehicles. As a result, a thoughtful RP can be an important input to understanding future EV adoption

Attribute Bundle Based Modeling

In the RP model we develop, we estimate consumer preferences for vehicle attributes such as price, fuel economy, acceleration (0 to 60 second time), and vehicle type, among others. We also include in our analysis the effect of demographics on preferences related to specific types of vehicles; for example, the relationship between living in a rural location and owning a pickup truck, or the relationship between having children and owning an SUV. Finally, based on these preferences we make predictions about vehicle choices made by any consumer. These predictions are tested against out-of-sample market choices made by real consumers.

Although EVs are not currently on the road in sufficient numbers to derive a statistical model of consumer preference for their attributes, it is possible to obtain an estimate of EV market penetration using choice models because most of the important EV-related *attributes* are in the market today. Such attributes include vehicle price, size, fuel economy, price of fuel, and acceleration. Other attributes, such as whether the vehicle is a hybrid electric vehicle (HEV), captures to some extent the degree to which consumers are willing to pay for "green technology" or "early adoption" in vehicles, though it is likely to be an underestimate of this willingness-to-pay for a fully electric vehicle. As a first estimate, we can assume that some of the value of perceived "greenness" of owning a future EV is captured by the value of owning an HEV today.

Discrete choice models can be used to analyze and predict a decision maker's choice of one alternative from a finite set of alternatives. Such models have been used successfully in the passenger vehicle market, and in large measure were developed in response to the need to understand consumer preferences for vehicle attributes. Discrete choice models are generally derived under the assumption of utility maximizing behavior by the decision maker. In other words, a decision maker chooses the alternative that generates his/her highest utility. This utility is known to the decision maker but not to the researcher. The researcher observes some attributes of the alternatives as faced by the decision maker and some attributes of the decision maker and can specify a function that relates these observed factors to the decision maker's utility. Usually, the utility value depends on parameters that are unknown to the researcher and are therefore estimated statistically.

Literature Review

The most popular discrete choice models used for forecasting demand of new products are so-called logit and nested logit models [(2), (3), (4) and (5)]. These models have some desirable attributes such as ease of estimation and calculation (6). However, these models have fatal flaws when it comes to understanding how future EVs will diffuse into the market. Their most significant flaw is their *independence from irrelevant alternatives* (also called their "iia" property). Logit models exhibit the *iia* property over all alternatives, while nested logit models exhibit it over alternatives within each nest. The *iia* property states that the ratio of the probabilities for any two alternatives is independent of the existence and attributes of other alternatives. As a result of this property, logit and nested logit models predict that a change in the attributes of one alternative (or the introduction/elimination of a new alternative) changes the probabilities of the other alternatives proportionately, such that the ratios of probabilities remain the same. This substitution pattern can be unrealistic in many choice situations. For example, the introduction of a pick-up truck with a larger cargo space is unlikely to affect the demand for a two-seater EV.

Several studies [(1), (6), (7) and (8)] have found that mixed logit models not only address the *iia* challenge, they provide sound and realistic estimates of consumer choice in the market. A mixed logit model is a superior representation of the market compared with logit and nested logit models mainly because of its heterogeneous representation of the market. For instance a logit model would predict that everyone in the market values 1 mpg of fuel economy identically. A nested logit model assumes everyone in a given market segment has the same preference for fuel economy. A mixed logit model captures heterogeneity within and between market segments and as such can capture important demographic interactions with vehicle features. Appendix 5.1 provides additional background on mixed logit models.

Basic Terminology

Here we describe some of the common terminology used in the literature and this report.

> *Decision maker* – the entity whose choice behavior is to be modeled. The decision maker can be an individual or a group of people depending on whether it is an individual or group decision to buy a certain product. In our analysis we select each *household* as a decision making entity. This is because the decision to buy a vehicle is a household matter taking into account factors such as number of children in the household, residential location, total income, etc.

Choice set – the set of alternatives available in the market from which the decision maker chooses. In vehicle choice analysis, a choice set includes various vehicle models available in the market in any given year. However, in this study, we wanted to keep our choice set general and not overly specific to vehicle models available in a particular year. Therefore we define each "vehicle alternative" in our choice set by the vehicle's manufacturer ("producer name") and segment ("vehicle segment"). Each element in our choice set is identified by the following: "producer-name_vehicle-segment". (More information about the justification for using "producer name" as an identifier is discussed in the next section). Hence, in this work a Ford Fusion would be identified in our choice set as "Ford_mid car" (where mid car is a mid-size vehicle). If there is more than one vehicle model manufactured by the same producer in any given segment, we take the average of attribute values for these models. These choices were made to cleanly interface with other EPRI research efforts that are currently on-going to inform EV market adoption over time.

2
METHODOLOGY

Mixed Logit Model Specification

Demand for automobiles is modeled here using a mixed logit representation of consumer preferences. The indirect utility U_{nj} that household $\boldsymbol{n}$ derives from purchasing vehicle alternative $\boldsymbol{j}$ is defined as:

$$\boldsymbol{U_{nj} = \delta_j + \mu' y_{nj} + \beta'_n x_j + \epsilon_{nj}} \quad \text{2-1}$$

$$\boldsymbol{\delta_j = \alpha' z_j + \xi_j} \quad \text{2-2}$$

The utility derived by a household $\boldsymbol{n}$ from purchasing vehicle $\boldsymbol{i}$ is separated into four categories as follows:

1. **Mean utility ($\boldsymbol{\delta_j}$).** The mean utility $\boldsymbol{\delta_j}$ acquired by each household only depends on the vehicle alternative, and therefore by definition does not depend on household characteristics. Consistent with the modeling approaches of [(1), (6) and (9)], we specify $\boldsymbol{\delta_j}$ as a function of vehicle attributes of interest z_j (e.g., price, mpg, acceleration, etc.) in linear combination with parameters $\boldsymbol{\alpha}$ which are specific to each vehicle attribute but do not change with vehicle alternative. Additionally, ξ_j captures the average utility for each unobservable vehicle alternative associated with vehicle attributes that are not included in z_j (e.g., noise, vibration, luxury feel, etc.).
2. **Observed Heterogeneity ($\boldsymbol{\mu' y_{nj}}$).** The observed heterogeneity is the component of the total utility that captures how different household demographics value specific vehicle attributes. It is called "observed" heterogeneity because both the vehicle attributes *and* household demographics are known. This term allows consumers with different observable characteristics (e.g., family size or rural location) to have different tastes for certain vehicle attributes, and thus specifies the extent to which vehicle choice varies with observable consumer demographics. Interactions between vehicle attributes and household characteristics are given by $\boldsymbol{y_{nj}}$ and are assumed in the model to affect utility homogenously across the population (e.g., the relationship of preferring an SUV as a function of family size).
3. **Unobserved Heterogeneity ($\boldsymbol{\beta'_n x_j}$).** The unobserved heterogeneity term is a reflection of the reality that other demographic (or consumer-specific) preferences affect the overall utility of a specific vehicle attribute (e.g., mpg or acceleration) in ways that are not measurable. As a result, this term helps explain why certain consumers have stronger preferences for some vehicle attributes than other consumers with the same measured demographics. Since $\boldsymbol{x_j}$ is always a subset of z_j, $\boldsymbol{\alpha}$ (in Equation 2-2) can be thought of as representing the average

vehicle utility coefficient with $\boldsymbol{\beta_n}$ capturing random variation around this average. It is common in this type of modeling to assume that the random variation in preference for $\boldsymbol{x_j}$ across the population follows a normal distribution with a mean of zero. In this work we have used the normal distribution for all vehicle attributes except for price, which is assumed to follow a lognormal distribution. As a result, the $\boldsymbol{\beta_n}$ coefficients can be thought of as representing the standard deviation ($\boldsymbol{\sigma}$) in the unobserved preference distributed either normally or lognormally about a mean value of zero.

4. **Disturbance Term ($\boldsymbol{\epsilon_{nj}}$).** The disturbance term is a random scalar that captures all remaining utility provided by vehicle $\boldsymbol{j}$ to household $\boldsymbol{n}$. As is the custom of mixed logit modeling, it is assumed that the disturbance term is Independent and Identically Distributed (*iid*). This assumption is the primary enabler permitting the estimations and computations presented in the *Coefficient Estimation Process* section.

Table 2-1 lists and defines the vehicle attributes and household demographic characteristics included in the mixed logit utility model developed for this report. Table 2-2 lists vehicle attributes and/or household demographics that enter the mixed logit utility model through variables $\boldsymbol{x_j}$, $\boldsymbol{y_{nj}}$ and $\boldsymbol{z_j}$.

Table 2-1
Vehicle and consumer demographic attributes included in the model

Variable	Meaning	Units
Price	Vehicle/Alternative MSRP	10,000 $
$/mile	Cost of fuel consumed per vehicle mile driven $\frac{\$}{\text{mile}} = \left(\frac{\text{gallons}}{\text{mile}}\right) * \left(\frac{\$}{\text{gallon}}\right)$	$/mile
Acceleration time inverse	Inverse of 0-60 acceleration time in seconds	second^{-1}
Footprint	Vehicle footprint	inch^2
Power*	Power rating of the vehicle	100 hp
Sport (two-seater)	Two-seater sports vehicle segment dummy	(1 if sport, else 0)
Truck	Truck segment dummy	(1 if truck, else 0)
SUV	SUV segment dummy	(1 if SUV, else 0)

* The attribute "power" is included mainly to represent pulling power of a vehicle. Aspects of the vehicle related to acceleration are covered by acceleration time inverse.

Table 2-1 (continued)
Vehicle and consumer demographic attributes included in the model

Variable	Meaning	Units
Minivan	Minivan segment dummy	(1 if minivan, else 0)
Hybrid	Hybrid vehicle dummy	(1 if hybrid, else 0)
Price/income	Ratio of vehicle price to household's total annual income	$/$
Minivan-children	Interaction between children in the household and minivan ownership Minivan = (1 if the vehicle is a minivan else 0) Children = (1 if the household has at least one child else 0)	Minivan * Children
SUV-children	Interaction between children in the household and SUV ownership SUV = (1 if the vehicle is a SUV else 0) Children = (1 if the household has at least one child else 0)	SUV * Children
Truck-rural	Interaction between living in rural area and Truck ownership Truck = (1 if the vehicle is a truck else 0) Rural = (1 if the household is located in a rural area else 0)	Truck * Rural

Table 2-2
Vehicle and consumer demographic attributes

Variable y_{nj}	Variable x_j	Variable z_j
Price/income	Price	Price
Minivan-children	$/mile	$/mile
SUV-children	Acceleration time inverse	Acceleration time inverse
Truck-rural	Footprint	Footprint
	Horsepower	Horsepower
	Hybrid	Hybrid
		minivan
		SUV
		Truck
		Sport

Data for vehicle alternative sales as a function of demographics was derived from the U.S. New Vehicle Customer Study (NVCS), which collects data monthly from households that purchased or leased new vehicles (10). The technical characteristics and prices (MSRP) of the vehicles themselves were acquired from a custom-built code for webscraping. Webscraping is a method of extracting specific information from websites. In this case, data on the vehicle attributes were extracted from cars.com (11).

Coefficient Estimation Process

The mixed logit formulation requires the estimation of four vectors of coefficients presented in Equations 2-1 and 2-2: $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\alpha}$. This estimation is carried out in two stages. The first stage uses maximum simulated likelihood to estimate $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and $\boldsymbol{\delta}$. Given the estimated coefficients from the first stage, the second stage uses instrument variable regression to estimate $\boldsymbol{\alpha}$**.** Information regarding the use of maximum simulated likelihood and instrument variable regression is provided below.

Logit Choice Probabilities

The probability that decision maker $\boldsymbol{n}$ chooses $\boldsymbol{i}$ is

$$P_{ni} = Prob\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}\ \forall\, j \neq i\right) \qquad \textbf{2-3}$$

$$= Prob\left(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}\ \forall\, j \neq i\right) \qquad \textbf{2-4}$$

Where, V_{ni} is the deterministic part of the utility function. If $\varepsilon_{ni}$is considered given, this expression is the cumulative distribution for each ε_{nj} evaluated at $\varepsilon_{nj} + V_{ni} - V_{nj}$, which is exp(-exp(-($\varepsilon_{ni} + V_{ni} - V_{nj}$))) since ε_{ni} is assumed to follow type 1 extreme value (Gumbel distribution). Since the ϵ values are independent (from their *iid* assumption), this cumulative distribution over all j ≠ i is the product of the individual cumulative distributions:

$$P_{ni}|\varepsilon_{ni} = \prod_{j\neq i} e^{-e^{-(\varepsilon_{ni}+V_{ni}-V_{nj})}} \qquad \textbf{2-5}$$

However, ε_{ni} is not given and so the choice probability is the integral of $P_{ni}|\varepsilon_{ni}$ over all values of ε_{ni} weighted by its density.

$$P_{ni} = \int \left(\prod_{j\neq i} e^{-e^{-(\varepsilon_{ni}+V_{ni}-V_{nj})}}\right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}}\, d\varepsilon_{ni} \qquad \textbf{2-6}$$

This results in a closed form expression for probability:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \qquad \textbf{2-7}$$

Maximum Simulated Likelihood Method

Since the logit probabilities take closed form, traditional maximum-likelihood procedures can be applied. The probability of person ***n*** choosing the alternative that he/she was actually observed to choose can be expressed as,

$$\prod_{\mathrm{i}} (\mathrm{P_{ni}})^{\mathrm{y_{ni}}} \qquad \textbf{2-8}$$

Where, $y_{ni} = 1$ if person ***n*** chose ***i*** and zero otherwise. Note that since $y_{ni} = 0$ for all non-chosen alternatives and P_{ni} raised to the power of zero is 1, this term is simply the probability of the chosen alternative.

Assuming that each decision maker's choice is independent of the other ***N*** decision makers, the probability of each person in the sample choosing the alternative that he/she was observed actually to choose is,

$$L(\theta) = \prod_{n=1}^{N} \prod_{i} (P_{ni})^{y_{ni}} \qquad \textbf{2-9}$$

Where, $\boldsymbol{\theta}$ is a vector containing the parameters of the model. The log-likelihood function is then,

$$LL(\theta) = \sum_{n=1}^{N} \sum_{i} y_{ni} lnP_{ni} \qquad \textbf{2-10}$$

Since the mixed logit probability is not deterministic, given that β is random, it is necessary to integrate over all possible values of $\boldsymbol{\beta_n}$ to determine the probability that household ***n*** chooses vehicle ***i***. As a result, the choice probability is the integral of $\boldsymbol{P_{ni}}(\boldsymbol{\beta_n})$ over all possible variables of $\boldsymbol{\beta_n}$ as given in Equation 2-11.

$$P_{ni} = \int \left(\frac{e^{\delta_j + \mu' y_{nj} + \beta_n' x_j}}{\sum_j e^{\delta_j + \mu' y_{nj} + \beta_n' x_j}} \right) f(\beta)\, d\beta \qquad \textbf{2-11}$$

For notational ease we define a vector of parameters **θ ≡ (μ, β)**. Equation 2-11 does not have an analytical solution. Therefore the result must be simulated. Steps involved in the simulation are as follows:

1. Random values of θ are drawn. We label these draws as θr with the superscript r = 1 referring to the first draw;
2. The probability $\boldsymbol{P_{ni}}$ in Equation 2-7 is calculated with this set of θr;
3. Steps 1 and 2 are repeated many times (approximately 500 times) and the results are averaged to determine the simulated probability as,

$$\hat{P}_{ni} = \frac{1}{R}\sum_{r=1}^{R} P_{ni}(\beta^r) \qquad \text{2-12}$$

where, R is the total number of draws.

The simulated probabilities $\hat{\boldsymbol{P}}_{\boldsymbol{ni}}$ are then multiplied together and the resulting values are compared with the observed vehicle $\boldsymbol{i}$ chosen by household $\boldsymbol{n}$. Previous researchers have found, due to computational difficulties, that this is better achieved by maximizing the sum of the log of likelihood function as presented by (8). The simulation proceeds by taking the natural log of simulated likelihoods and summing them as shown in Equation 2-10. The maximum likelihood estimate of **θ** is that value of **θ** that maximizes **SLL(θ).** In other words, it is the value that makes the observed data most probable.

$$SLL = \sum_{n=1}^{N}\sum_{j=1}^{J} ln\hat{P}_{nj} \qquad \text{2-13}$$

Berry Inversion/Contraction Method

In the numerical search for the maximum of the simulated log likelihood function (Equation 2-13), **δ** is calculated for each trial value of **θ**. Therefore the **δ** vector is estimated conditional on **θ** and is thus formally **δ(θ).** We use the contraction procedure developed in (4) where at any given value of **θ**, the formula in Equation 2-14 is applied iteratively until predicted shares equal observed market shares. Therefore **δ** is also a function of actual market shares S and thus defined as **δ(θ, S)**.

$$\boldsymbol{\delta_j^t(\theta, S) = \delta_j^{t-1}(\theta, S) + \ln\left(S_j\right) - \ln\left(\widehat{S_j}\left(\theta, \delta^{t-1}(\theta, S)\right)\right)} \qquad \text{2-14}$$

Where, $\widehat{\boldsymbol{S_j}}$ is the predicted market share obtained by calculating $\mathbf{P}_{ni}$ with parameters **θ** and **δ** and averaging $\mathbf{P}_{ni}$ over the **n** households in the sample.

Instrument Variable Regression

Once **θ** and **δ** are estimated we estimate the regression given by Equation 2-2 which relates the **δ** values (average vehicle utilities) to vehicle attributes. Unobservable vehicle quality variables ξ include vehicle attributes that are not observed, but that are likely to be correlated with price. For example, consider two vehicles *L* and *C* that are similar in all respects expect that *L* has a leather interior and *C* has cloth upholstery. Type of interiors is not an attribute that we observe. In a scenario where sales of *L* exceeded that of *C*, a simple regression analysis would conclude that people are less sensitive to higher prices. However, in reality they are sensitive to price but they have traded off between price and leather interior. Therefore, a simple regression of the **δ** vector on vehicle price and other attributes will estimate that consumers are less price sensitive than they actually are. In order to correct for this bias, we use the *Instrument Variable* (IV) regression approach. Further details on this approach can be found in (12). The first two instruments that we use were first used in (4). The latter two measures, which capture the extent to which other vehicles' non-price attributes differ from vehicle i's non-price attributes were first used in (13). Letting $\boldsymbol{d_{ji}}$ be the difference in an attribute, vehicle footprint, between vehicle j and i, we calculate four instruments for vehicle i for each attribute: the sum of $\boldsymbol{d_{ji}}$ over all j made by the same manufacturer, the sum of $\boldsymbol{d_{ji}}$ over all j made by competing manufacturers, the sum of $\boldsymbol{d_{ji}^2}$ over all j made by the same manufacturer, and the sum of $\boldsymbol{d_{ji}^2}$ over all j by competing manufacturers.

3
SAMPLE ESTIMATION RESULTS

Using household demographics, sales and vehicle attribute data from 2008, we estimated the coefficients for Equations 2-1 and 2-2. Table 3-1 presents the estimates associated with the demographic and vehicle attribute interaction (representing observed heterogeneity). As noted in the table all the estimates are statistically significant at 95% level. Estimates associated with the vehicle attributes representing unobserved heterogeneity are presented in Table 3-2.

Results show that price is the only attribute in this category that is significant at 95% level. This observation and its implications are explained in detail in the next chapter.

Table 3-3 presents the estimates for the regression in Equation 2-2. $/mile, horsepower, minivan and hybrid attributes are found to be significant at 95% level.

Table 3-1
Heterogeneous (Observed) Demand Parameters (μ)

Vehicle Attribute	Parameter	Standard Error
Price/income*	-5.05	0.25
Minivan-children*	0.89	0.14
SUV-children*	0.44	0.05
Truck-rural*	1.11	0.06

* These estimates are statistically significant at 95% level

Table 3-2
Heterogeneous (Unobserved) Demand Parameters (β)

Vehicle Attribute	Parameter	Standard Error
Price*	0.03	0.01
$/mile	1.17	1.32
Acceleration time inverse	0.41	1.24
Footprint	0.31	0.21
Horsepower	0.08	0.05
Hybrid	0.23	0.16

* These estimates are statistically significant at 95% level

Table 3-3
Homogeneous Demand Parameters (α)

Vehicle Attribute	Parameter	Standard Error
Price	-0.59	0.37
$/mile*	-17.31	7.83
Horsepower*	2.49	1.00
Acceleration time inverse	-19.99	11.71
Footprint	0.79	2.00
Minivan*	-2.37	0.52
SUV	-0.05	0.34
Truck	-1.18	0.60
Sport	-0.30	0.38
Hybrid*	-1.27	0.61
Constant*	-3.92	1.89

* These estimates are statistically significant at 95% level. This indicates which attributes are significant given the model structure and the specific data set used in estimating the parameters. These attributes were selected following an extensive literature survey due to their role in influencing consumer decisions. Some of these individual attributes may not be significant considering the current data set, but the results could be different for a different dataset. Also, the combined effect of these attributes could still be significant. Therefore we retain all these attributes in later analysis.

Using the Utility Model for Computing Market Shares

The mixed logit approach presented above models the choice behavior of specific households as a function of their demographic characteristics and attributes of the vehicles in the household's choice set. However, an important objective for EPRI is to better understand how the purchase behavior of groups of consumers and sales of specific vehicles might change as a result of changes in socio-demographic characteristics (e.g., income, family size, etc.) over time and/or changes in attributes of alternatives (e.g., fuel prices, mpg, etc.).

While this forecasting function was the primary role of a companion study conducted by Veritas using the results of this study, it was also important for the EERA/UM Team to look at such "counterfactual" scenarios to verify the model behavior and check the realism of the coefficients listed in Table 3-1, Table 3-2 and Table 3-3.

We performed several counterfactual scenarios to "stress test" the model and provide a sense of its realistic behavior.

The following three steps are required to take the mixed logit utility model described above and use it to compute market share using the estimated coefficients listed in Table 3-1, Table 3-2 and Table 3-3.

1. Calculate the utility derived by a consumer $\boldsymbol{n}$ by choosing vehicle $\boldsymbol{j}$:

$$\boldsymbol{U_{nj} = \delta_j + \mu' y_{nj} + \beta'_n x_j + \epsilon_{nj}}$$

Where, $\boldsymbol{\delta_j = \alpha' z_j + \xi_j}$

When $\boldsymbol{\epsilon_{nj}}$ is assumed to be independent and identically distributed, the term can be neglected from future calculations resulting in the need to compute only the non-stochastic part of the utility[1]:

$$\boldsymbol{V_{nj} = \delta_j + \mu' y_{nj} + \beta'_n x_j} \qquad \textbf{3-1}$$

2. Once the utility values for each combination of household and vehicle are determined, the probability of household $\boldsymbol{n}$ choosing vehicle $\boldsymbol{i}$ is computed as shown below in Equation 3-2.

$$\boldsymbol{P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}} \qquad \textbf{3-2}$$

3. The third and final step in the process is to calculate the market share of a particular vehicle. The aggregate number of vehicle alternative i sold is the sum of the probabilities of households $1, 2, 3 \ldots n$ choosing vehicle i. The market share S_i of vehicle i is then the ratio of aggregate number of vehicles sold to the total number of households considered.

$$\boldsymbol{S_i = \frac{\sum_n P_{in}}{n}} \qquad \textbf{3-3}$$

Results Shared with Veritas

The primary role of EERA/UM team was to develop a revealed preference based choice model for use in the dynamic electric vehicle adoption model developed by Veritas. Pursuant to this objective we have shared the results that would enable Veritas to forecast market shares following the steps described in the previous section. Among the results shared are the estimated coefficients as presented in Table 3-1, Table 3-2, and Table 3-3.

In addition to these coefficients, values for mean utility derived from unobserved attributes **(ξ)** are also required to calculate the mean utility **(δ).** ξ_j specific to each vehicle alternative is calculated as a difference $\boldsymbol{\delta_j - (\alpha^* Z_j)}$. The vector ξ_j was also shared with the Veritas team.

---

[1] $\boldsymbol{\beta'_n x_j}$ in Equation 3-1 in fact varies randomly across households. We approximate its deterministic value through simulation in order to calculate the non-stochastic part of the utility $\boldsymbol{V_{nj}}$. The estimated coefficients $\boldsymbol{\beta}$ (in Table 3-2) represent the standard deviation (σ) of distributions with zero mean. The approximated choice probabilities are obtained by drawing $\boldsymbol{\beta_n}$ for each household $\boldsymbol{n}$ from the distribution given by **f(0,σ)**, calculating $\boldsymbol{V_{nj}}$ and $\boldsymbol{P_{ni}}$ (see Step 2) for each value of $\boldsymbol{\beta_n}$ and averaging $\boldsymbol{P_{ni}}$ over the number of iterations performed during simulation.

4
MODEL VALADATION AND DISCUSSION

Using the market share calculation approach, we perform three basic checks of the model function. First, we compare model predictions for market share against the actual observed market shares ("in-sample") used to create the model. We would expect a high level of agreement in such a comparison. Our goal was to further evaluate the role of various model components (e.g., mean utility, observed heterogeneity, and unobserved heterogeneity) in achieving the accuracy of market share calculations.

Second, we compare model predictions for market share against "out-of-sample" vehicles. Specifically, we consider how well a market share model created with 2008 data estimates the actual market share data from 2007. We expect a good level of agreement, noting that the agreement cannot be perfect since the set of actual vehicle alternatives sold in 2007 was a bit different than in 2008, and there are unobserved year-to-year changes in preferences that all demographics will have for specific vehicle attributes.

After observing sound agreement between out-of-sample predictions based on 2008 vehicles and actual market shares in 2007, we looked at a couple of cases to determine the correlation between observed 2008 market shares and what market shares the model predicts under higher gas prices and lower hybrid vehicle prices than were actually observed in 2008. This final exercise is enlightening while also providing an additional check on the realism of the results.

In-Sample Market Share Calculations

In this section we compare the estimated and actual market shares for 2008, noting that the coefficients estimated in Table 3-1, Table 3-2, and Table 3-3 actually were derived from the same 2008 dataset (i.e., "in-sample"). Figure 4-1 presents a plot of actual v. estimated market shares. The estimated values match the predicted values with a correlation coefficient of 99%. Given strong agreement between estimated and actual values (in-sample), we set out to better understand which of the three estimated parameters [mean utility ($\boldsymbol{\delta}$), observed heterogeneity ($\boldsymbol{\mu}$) and unobserved heterogeneity ($\boldsymbol{\beta}$)] had the most important effect on the quality of the estimation. The effects were determined by removing the parameters from the model one-at-a-time. First, the unobserved heterogeneity ($\boldsymbol{\beta}$) was removed with results shown in Figure 4-2. There are about five vehicle alternatives (out of a total of 161 alternatives) for which there is some deviation between observed and estimated market shares. A comparison of estimated and observed market shares reveals that the estimates are nearly identical, again achieving a correlation coefficient of 0.99. This means that the random effects associated with the attributes considered in the model do not play an important role in estimating vehicle choices made by households for 2008 market shares. Therefore, we can safely remove the unobserved heterogeneity ($\boldsymbol{\beta}$) without much concern regarding its effect on model performance.

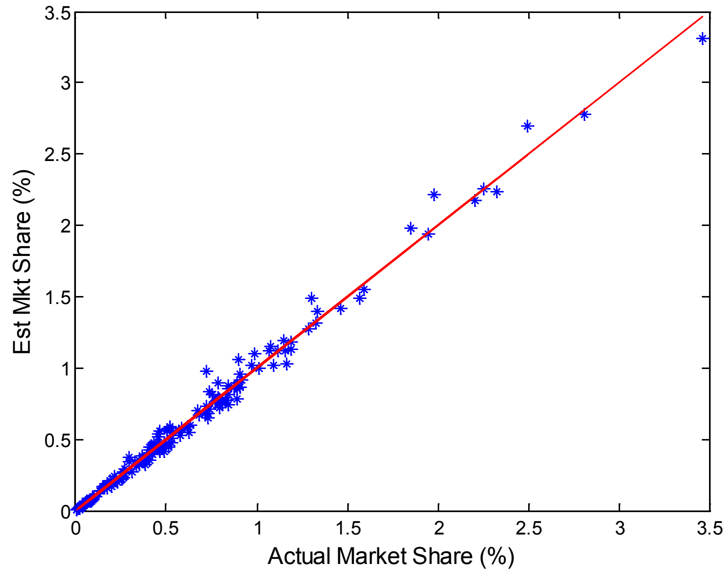


Figure 4-1
Plot of actual vs estimated market shares

The fact that **β** elements are not critical in ensuring a good match between actual and observed market shares is advantageous from the point of view of computation time. As discussed in the methodology section, some simulation work is associated with **β** calculations and that can take up to several minutes of computation time. While this computation time is not significant for our work in this report, it was particularly important for Veritas, where the bulk of the computational burden lies. The fact that the **β** elements could be removed meant the Veritas model could remain both high quality and computationally tractable after incorporation of our results.

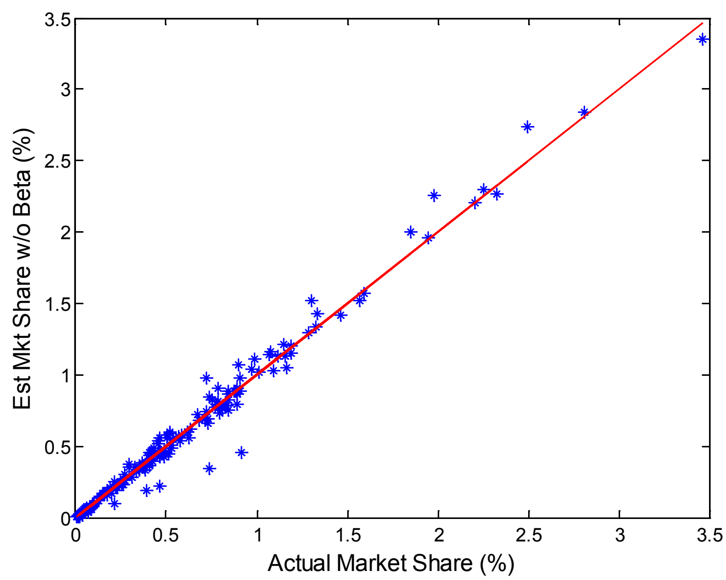


Figure 4-2
Market shares computed without β elements

After removing the unobserved heterogeneity (**β**) elements, the observed heterogeneity (**μ**) elements were removed – meaning that the utility derived by a consumer from choosing a particular vehicle was determined only by the mean utility (**δ**) associated with the vehicle. The resulting market share estimates are presented against the actual market shares in Figure 4-3. As evident from the figure, there is a very poor correlation between estimated and actual market shares in this scenario where both **β** and **μ** elements are removed. This result indicates that (a) an appropriate set of attributes was observed in the model definition; and, (b) additional attributes are not required to achieve a strong in-sample match of actual and estimated market shares.

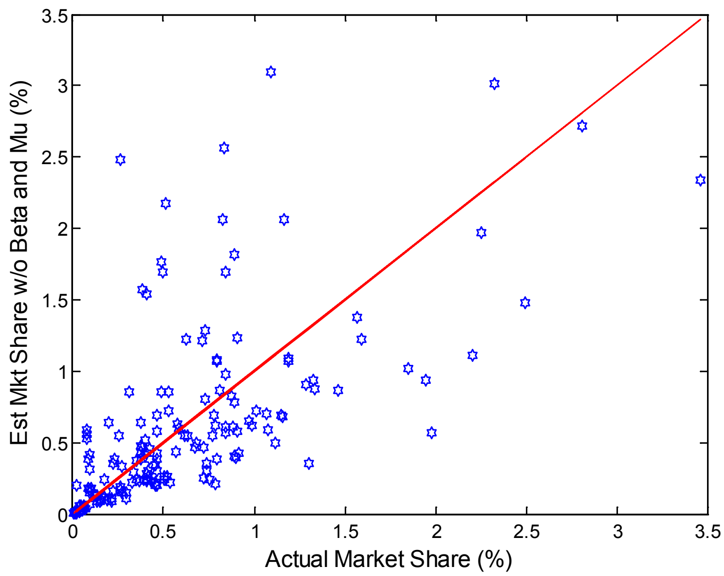


Figure 4-3
Market shares estimated without both β and μ elements

Investigation of Out-of-Sample Market Shares

The year 2007 was used as the out-of-sample test for the model (estimated using 2008 data). This test of out-of-sample market share predictions is challenged by the fact that some vehicles offered in 2007 were discontinued in 2008, and some vehicles that were offered in 2008 were not available in 2007. Therefore, we could not use the coefficients estimated using 2008 data (with all vehicle alternatives from 2008) to predict 2007 market shares without adjustment. Instead, we reduced the choice set used to estimate the 2008 market that it included *only* vehicles that were available in the market in *both* 2007 and 2008. This resulted in a reduction from 161 vehicle alternatives to 147 vehicles. Figure 4-4 shows that this has a significant impact on 2008 estimations, since about 10% of the vehicles (and households that bought these vehicles) have been removed from the sample used to estimate the coefficients. The correlation coefficient between estimated and actual market shares (in-sample for reduced choice set) drops from 99% to 92% before considering the change in model year to 2007. This reduction occurred because there were vehicle options available to consumers in 2008 that were intentionally being ignored due to the desire to use this model to estimate market shares in 2007.

Once the dataset was reduced, the coefficients from year 2008 were used along with household and vehicle attribute data from 2007 to arrive at expected market shares for 2007 using the model. Equations 4-1 and 4-2 describe the combination of these data.

$$U_{nj} = \delta_j + \mu^{08} y_{nj}^{07} + \beta_n^{08} x_j^{07} + \epsilon_{nj} \tag{4-1}$$

$$\delta_j = \alpha^{08} z_j^{07} + \xi_j^{08} \tag{4-2}$$

Here the ξ terms represent unobserved part of the mean utility δ, which are not explained by the observed vehicle attributes in Z_j. They are calculated as the difference between $\delta_j - (\alpha * Z_j)$, and are specific to each vehicle. The ξ terms are the reason that the vehicle alternatives in 2008 and 2007 choice set need to match exactly. It is also important to note that the values of observed demographic and vehicle attributes (income, price, etc.) can vary for the same vehicle between the two data sets.

Figure 4-5 shows the estimated 2007 market share using coefficients estimated on the reduced 2008 dataset as in Equations 4-1 and 4-2. It is seen that, compared with Figure 4-4, the correlation coefficient is reduced from 92% to 87%, which suggests that the model impact of the year-to-year change was roughly similar to the impact on the sample size caused by the year-to-year change in vehicle models. This is considered an outstanding result given that the 2007 market shares were indeed out-of sample and resulted in a roughly similar impact on the model as removing 10% of the data.

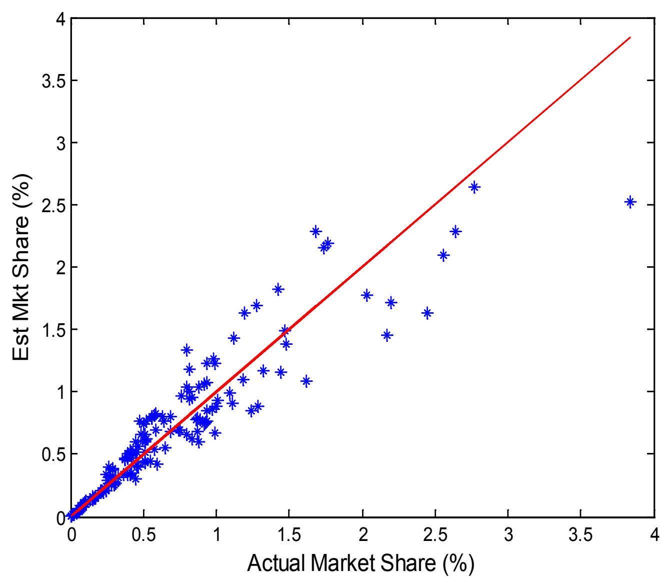


Figure 4-4
Estimated and actual market shares for 2008 using reduced data set to match vehicle alternatives that were available in 2007. Correlation Coefficient = 92%.

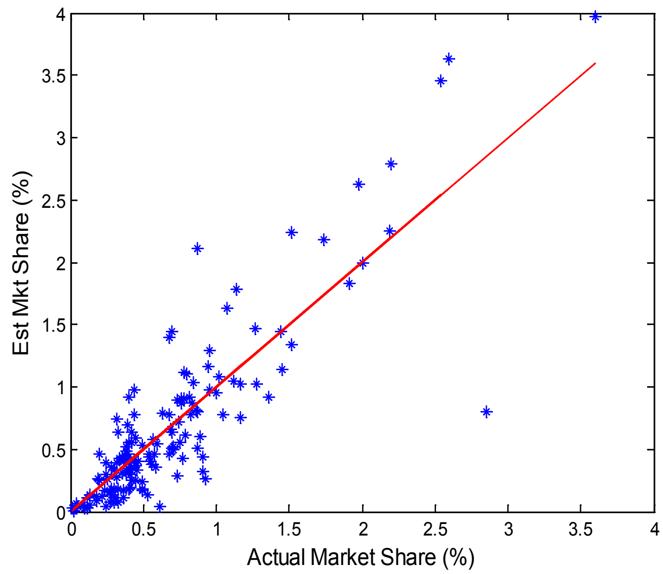


Figure 4-5
Estimated and actual market shares for the year 2007 (using coefficients estimated using data from year 2008). Correlation Coefficient = 87%

Counterfactual #1: Gas Prices

In the tests described above, we used a national average retail gas price of 3.266 $/gallon (14) to calculate $/mile. To experiment with the 2008 mixed logit model, we performed a counterfactual study where we steadily increased gas prices starting from the 2008 prices up to $5/gallon and report the impact on the overall vehicles purchased.

Assuming fuel economy of all vehicle alternatives remains the same, an increase in gas prices would be expected to lead to higher market shares of more fuel efficient vehicles. This is both intuitive, and because Table 3-3 showed that the $/mile parameter (i.e., the operating cost considering, in this case, only fuel cost) is a significant factor. The model coefficients from Table 3-1, Table 3-2 and Table 3-3 were used to estimate market shares of all 161 vehicle alternatives as the price of gas increased. The correlation between the actual observed market shares in 2008 with these counterfactual market shares is shown in Figure 4-6. As expected, the correlation decreases as gas prices increase, and it is observed that they do so at an accelerated rate starting at $4/gallon.

Table 4-1 lists the top 15 vehicle models that lost market share and the top 15 vehicle models that gained market share when the assumed gas prices were at 4.50 $/gallon. Among the top 15 models that lost market shares are mostly low mpg and larger footprint vehicle models from the SUV, truck, and luxury sedan segments. Among the top 15 models that gained market shares are higher mpg and smaller footprint vehicles from compact and midsize sedan segments. Hybrid sedans were among the top 3 models to gain market shares. While it is not possible to verify these results directly, they do have a strong appeal to intuition regarding market behavior.

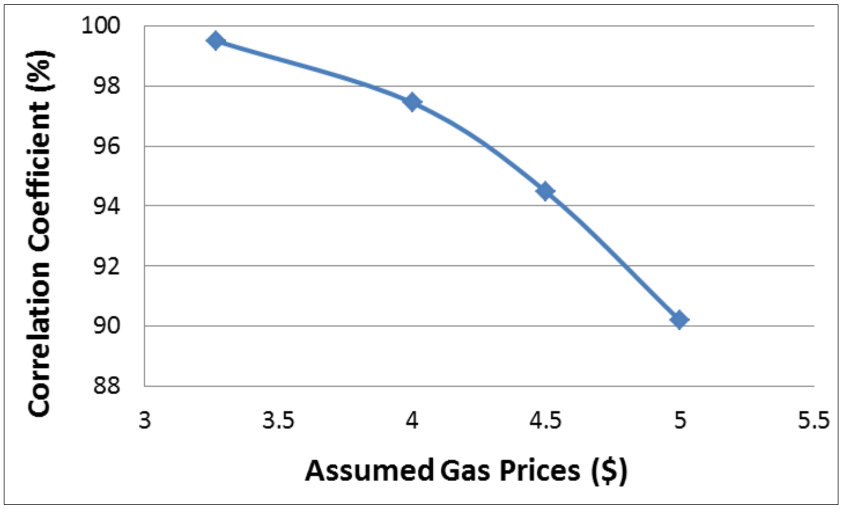


Figure 4-6
Plot of gas prices factor increase compared to 2008 prices vs correlation coefficient between actual and estimated market shares

Figure 4-6 also demonstrates that in order to observe a significant change in vehicle choices as a result of an increase in gas prices alone, the increase in gas prices has to be significant. For instance, only when gas prices reach 5.00 $/gallon do we see the correlation between estimated and 2008 actual market shares drop to 90%. This means that though fuel $/mile is a significant consideration for consumers, it is not the only or even the most significant deciding factor in their vehicle choices. Many consumers show a significant amount of preference towards other vehicle attributes such as vehicle segment, acceleration, price, and other factors.

Table 4-1
List of top 15 vehicle models that gained market shares and top 15 vehicle models that lost market shares assumed gas prices of 5.00 $/gallon

Top 15 Models that Lost Market Shares		Top 15 Models that Gained Market Shares	
Vehicle Models	% Change in Market Share	Vehicle Models	% Change in Market Share
Nissan Titan	-37%	Honda Civic Hybrid	73%
Ford Expedition	-35%	Toyota Camry Hybrid	66%
Land Rover LR3	-35%	Ford Escape Hybrid	47%
Nissan Armada	-35%	Honda Fit	36%
Ford F-150	-31%	Scion xD	35%
Toyota Sequoia	-28%	Honda Civic	33%
Cadillac Escalade	-28%	Chevrolet Cobalt	31%
HUMMER H3	-25%	Nissan Sentra	30%
Chrysler Aspen	-24%	Ford Focus	28%
Lexus GX470	-24%	Nissan Versa	27%
Chevrolet Avalanche	-23%	Toyota Highlander Hybrid	26%
Audi Q7	-23%	Saturn Astra	26%
Ford F-250	-22%	Hyundai Accent	26%
Lincoln MKX	-22%	Hyundai Elantra	26%
Infiniti EX35	-22%	Nissan Altima	25%

Counterfactual #2: Hybrid Vehicle Prices

In this counterfactual test we analyze the effect of a decrease in prices of hybrid vehicle alternatives on their market shares. We use vehicle attribute and household characteristics data and coefficient values for the year 2008 (as used in Section 3.1) for this analysis. Prices of hybrid vehicles were decreased by 20%, 30% and 40%.

Table 4-2 presents the actual market shares for these vehicles and market shares for the three price reduction scenarios. As expected, market shares steadily increase for hybrids as their prices decrease. We gain more insight into how consumers may trade-off between different attributes when we compare hybrid and conventional versions of the same vehicles as the prices of hybrids are reduced relative to conventional vehicles. For instance, the Ford Escape hybrid at a 27% price reduction costs the same as a conventional Ford Escape. However, the market shares for these vehicles under this scenario (where their prices are same) are not the same. The hybrid version's market share was approximately twice that of the conventional version's market share. Except for the $/mile rating, the hybrid version (0.10 $/mile) and the conventional version (0.14 $/mile) are more or less similar. The lower $/mile attribute of the hybrid version is thus responsible for its greater market share even when both versions cost the same.

Table 4-2
Variation in market shares of hybrid vehicle alternatives

Example Vehicle Alternative	2008 Baseline Price ($)	2008 Actual Mkt Share (%)	Mkt Share at 20% Price Reduction (%)	Mkt Share at 30% Price Reduction (%)	Mkt Share at 40% Price Reduction (%)
Ford Escape Hybrid	27797	0.91	1.61	2.04	2.75
GMC Yukon Hybrid	35345	0.01	0.03	0.04	0.05
Honda Civic Hybrid	22600	0.20	0.34	0.43	0.56
Mazda Tribute Hybrid	25310	0.02	0.03	0.04	0.05
Nissan Altima Hybrid	25480	0.06	0.11	0.14	0.19
Toyota Prius Hybrid	23350	0.70	1.18	1.48	1.96
Toyota Highlander Hybrid	34200	0.37	0.73	0.98	1.38
Lexus GS450h	55800	0.03	0.08	0.13	0.23
Lexus RX400h	42080	0.44	0.97	1.39	2.08

5
CONCLUSION

In this work we discuss the development of a discrete choice based vehicle demand model. Our modeling approach draws from the state-of-the art mixed logit models. This approach can model heterogeneity in consumer preferences for various vehicle attributes and is not limited by factors such as iia that are associated with traditional logit models. We provide the results from estimation process in this report and describe the process used to calculate market shares using the results from estimation.

Using the estimated coefficients along with data on vehicle attributes and household demographics, we estimate vehicle market shares for four different scenarios. For the in-sample scenario we use coefficients estimated for the year 2008. We calculate the market shares of vehicles for the same year to evaluate the accuracy with which the model can reproduce market shares. We find that the predicted and actual market shares match closely with correlation coefficient of 99%.

In the second scenario we predict the vehicle market shares for the year 2007 using the coefficients estimated with data for the year 2008. We had to reduce the choice set used to estimate the coefficients so that it includes only vehicles that were available in the market in both 2007 and 2008. With the reduced choice set the correlation between estimated and actual market shares for year 2008 (in-sample) reduced from 99% to 92%. The estimated market shares for year 2007 were found to match actual market shares with 87% correlation. This leads to a conclusion that the model perform well even for out-of-sample predictions.

We also present the analysis of two counterfactuals intended to “stress test” the model. In the first case we use change in gas prices as a trigger (holding all other factors constant) and study the change in market shares as compared to the year 2008. We identified the top 15 vehicle models that lost market shares and observed that they were mostly low mpg and larger footprint vehicle models. The top 15 models that gained market shares were higher mpg and smaller footprint vehicles. Hybrid sedans were among the top 3 models to gain market shares.

In the second counterfactual we focus on hybrid vehicles and study the influence of change in prices of these vehicles (compared to 2008) on the overall market shares. As expected, hybrids gain market shares as prices decrease. We make a more interesting observation by comparing the market shares for conventional and hybrid versions of the same vehicle (Ford Escape for example). We observe that hybrid version’s market share is twice that of the conventional version when both versions cost the same. The lower $/mile attribute of hybrid version is thus responsible for its greater market share even when both versions cost the same.

Through these scenarios we confirm the model’s ability to predict market shares for out-of-sample cases with great accuracy. The results indicate that the mixed logit model characterized for this study may be an effective tool for predicting electric vehicle market demand for future vehicle designs and configurations.

6
REFERENCES

1. *Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles .* **Brownstone , D., Bunch, D. and Train, K.** 2000, Transportation Research Part B, Vol. 34, pp. 315-338.

2. *Product Differentiation and oligopoly in International Markets: The Case of the U.S. Automobile Industry.* **Goldberg, P. K.** 4, 1995, Econometrica, Vol. 63, pp. 891-951.

3. *The Effects of the Corporate Average Fuel Efficiency Standards in the U.S.* **Goldberg , P. K.** 1, 1998, The Journal of Industrial Economics , Vol. 46, pp. 1-33.

4. *Automobile Prices in Market Equilibrium .* **Berry , S, Levinsohn, J and Pakes, A.** 4, 1995, Econometrica, Vol. 63, pp. 841-890.

5. *Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market.* **Berry, S., Levinsohn, J. and Pakes, A.** 1, 2004, Journal of Political Economy, Vol. 112, pp. 68-105.

6. *Forecasting New Product Penetration with Flexible Substitution Patterns.* **Brownstone, D. and Train, K.** 1, 1999, Econometrics, Vol. 89, pp. 109-129.

7. *Mixed MNL Models for Discrete Response.* **McFadden , Daniel and Train , Kenneth.** 5, 2000, Journal of Applied Economics, Vol. 15, pp. 447-470.

8. **Train, Kenneth.** *Discrete Choice Methods with Simulation.* s.l. : Cambridge University Press, 2009.

9. **Kate Whitefoot, Meredith Fowlie, and Steven Skerlos.** Product Design Response to Industrial Policy:. 2011.

10. **Maritz Research.** U.S. New Vehicle Customer Study (NVCS).

11. **cars.com.** Webscrape from cars.com.

12. **Heckman , James J, Urzua, Sergio and Vytlacil, Edward J.** Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics.* 2006, Vol. 88, pp. 389-432.

13. *Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers.* **Train, K. and Winston, C.** 4, 2007, International Economic Review, Vol. 48, pp. 1469-1496.

14. **U.S. Energy Information Agency.** Weekly Retail Gasoline and Diesel Prices. [Online] http://www.eia.gov/oil_gas/petroleum/data_publications/wrgp/mogas_history.html.

A
APPENDIX

Table A-1
Additional Literature Review

Serial Num.	Article Information		Summary
1	Title	Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry	The effect of two trade policies, voluntary export restraint (VER) and exchange rate pass-through, on prices change during 1983-87 has been discussed. On the demand side a nested-logit discrete choice model is adopted that is estimated using micro data from Consumer Expenditure Survey. The supply side of the automobile industry is modeled as an oligopoly with product differentiation. The nested-logit model deals with the issue of IIA to a great extent but not completely.
	Authors	Goldberg, P. K.	
	Journal	Econometrica, Vol. 63, No. 4, pp. 891-951	
	Year	1995	
2	Title	Automobile Prices in Market Equilibrium	The substitution behavior of moving from vehicle purchase to outside good, when price increased, is compared under standard logit and BLP's random coefficient logit model. The major contribution of this paper is that it provides a framework to utilize the existing aggregate consumer-level data and estimate the cost and demand parameters. BLP offered a useful method to deal with endogeneity and move it out of nonlinear choice models into linear regression. The automobile market data is collected in a 20-years period, 1971-1990, from Automotive News Market Data Book.
	Authors	Berry, S., Levinsohn, J., and Pakes, A.	
	Journal	Econometrica, Vol. 63, No. 4, pp. 841-890	
	Year	1995	
3	Title	The Effects of the Corporate Average Fuel Efficiency Standards in the US	This paper focuses on the effects of CAFE standards on automobile sales, prices, and fuel consumption. Author builds a discrete choice model of auto demand and a continuous model of vehicle utilization using data from Consumer Expenditure Survey (1984-1990). It also argues that nested logit models are better to use for modeling automobile demand than simple multinomial logit models. This is because the nested logit models consider the possibility that the consumer forgoes the purchase and includes information on past purchases.
	Authors	Goldberg, P. K.	
	Journal	The Journal of Industrial Economics, Vol. 46, No. 1, pp 1-33.	
	Year	1998	

Table A-1 (continued)
Additional Literature Review

<table>
<tr><td rowspan="4">4</td><td>Title</td><td>Mixed MNL Models for Discrete Response</td><td rowspan="4">This paper describes in great detail how the mixed multinomial logit model works, using as an example the problem of demand for alternative vehicles. They show that the restrictions on consumer behavior imposed by IIA can be relaxed by using the mixed logit specification for vehicle choice probabilities. Choice probabilities are estimated using simulation methods because under mixed logit approach the utility function does not have a convenient closed form.</td></tr>
<tr><td>Authors</td><td>McFadden, D., and Train, K</td></tr>
<tr><td>Journal</td><td>Journal of Applied Economics, Vol 15, Issue 5, pp 447-470</td></tr>
<tr><td>Year</td><td>2000</td></tr>
<tr><td rowspan="4">5</td><td>Title</td><td>Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles</td><td rowspan="4">This paper focused on comparing the multinomial logit model to the mixed logit model for data on California households’ revealed and stated preferences for automobiles. The paper argues that the mixed logit model is superior to the multinomial logit model in that it fits the data more accurately for this purpose. Most importantly, this paper discusses how critical it is to use both stated and revealed preferences of consumers. The stated preference data are critical for getting information about attributes not available in the marketplace, but the forecasts from this data can be implausible</td></tr>
<tr><td>Authors</td><td>Brownstone, D., Bunch, D., and Train, K.</td></tr>
<tr><td>Journal</td><td>Transportation Research Part B, Vol. 34, pp 315-338</td></tr>
<tr><td>Year</td><td>2000</td></tr>
<tr><td rowspan="4">6</td><td>Title</td><td>Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market</td><td rowspan="4">An essential part to this study is that they utilize not only a consumer’s first choice car (the one purchased), but also the second choice car that the consumer might have purchased if the first choice was not available. This information helps in determining just how important each characteristic is for each consumer. The study compares the results of using a logit model where only the first choice data is used and one where both the first and second choice data is used.</td></tr>
<tr><td>Authors</td><td>Berry, S., Levinsohn, J., Pakes, A</td></tr>
<tr><td>Journal</td><td>Journal of Political Economy, Vol. 112, No. 1, pp. 68-105</td></tr>
<tr><td>Year</td><td>2004</td></tr>
<tr><td rowspan="4">7</td><td>Title</td><td>Vehicle choice behavior and the declining market share of US automakers</td><td rowspan="4">This paper employed mixed logit demand model to study the relation between the consumer choice behavior and market share drops of the U.S. automakers in the past decade. It showed that the loss of U.S. automaker market shares can be explained by the vehicle attributes, such as retail price, power, weight, fuel consumption, body type, transmission type and reliability, where Japanese and European manufacturers have more improvements on attributes than U.S. manufacturers.</td></tr>
<tr><td>Authors</td><td>Train, K.E. and Winston, C.</td></tr>
<tr><td>Journal</td><td>International Economic Review, 48(4), pp. 1469-1496.</td></tr>
<tr><td>Year</td><td>2007</td></tr>
</table>

Table A-1 (continued)
Additional Literature Review

<table>
<tr><td rowspan="4">8</td><td>Title</td><td>Product Design Responses to Industrial Policy: Evaluating Fuel Economy Standards Using an Engineering Model of Endogenous Product Design</td><td rowspan="4">The article studied the impacts of a policy on design decisions which are intrinsically connected to the interaction between the policy, consumer demand, engineering tradeoffs and constraints, and the economic structure of the industry. The authors employed a mixed logit based discrete choice model following the development of Train and Winston (2007) supplemented by observed vehicle purchases and demographics with stated information about other considered vehicles.</td></tr>
<tr><td>Authors</td><td>Kate Whitefoot, Meredith Fowlie, and Steven Skerlos</td></tr>
<tr><td>Journal</td><td>Energy Institute at Haas Working Paper Series</td></tr>
<tr><td>Year</td><td>2011</td></tr>
</table>

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI's members represent approximately 90 percent of the electricity generated and delivered in the United States, and international participation extends to more than 30 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; and Lenox, Mass.

Together…Shaping the Future of Electricity

3002001455

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com