Impact of vehicle usage on consumer choice of hybrid electric vehicles
Lin He\textsuperscript{a}, Wei Chen\textsuperscript{a,*}, Guenter Conzelmann\textsuperscript{b}

\textsuperscript{a}Integrated Design Automation Laboratory, Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA
\textsuperscript{b}Center for Energy, Environmental, and Economic Systems Analysis, Argonne National Laboratory, Argonne, IL 60439, USA

\textbf{Abstract}

We analyze the vehicle usage and consumer profile attributes extracted from both National Household Travel Survey and Vehicle Quality Survey data to understand the impact of vehicle usage upon consumers’ choices of hybrid electric vehicles in the US. In addition, the key characteristics of hybrid vehicle drivers are identified to determine the market segmentations of hybrid electric vehicles and the critical attributes to include in the choice model. After a compatibility test of two datasets, a pooled choice model combining both data sources illustrates the significant influences of vehicle usage upon consumers’ choices of hybrid electric vehicles. Even though the data-bases have in the past been used independently to study travel behavior and vehicle quality ratings, here we use them together.

\section{1. Introduction}

Alternative fuel vehicles have drawn increasing attention, because of their potential to reduce greenhouse-gas emissions and utilize renewable energy sources. The use of alternative fuel vehicles, such as plug-in hybrid electric vehicles (PHEVs) is expected to grow significantly in the near future; in particular, US President Obama has called for half of all the cars purchased by the federal government to be PHEVs by 2012 and to have a million PHEVs on the road by 2015. Ample literature can be found in the transportation research domain that deals with potential environmental impacts of PHEVs.

While PHEVs are just now becoming available, hybrid electric vehicles have been sold to consumers since the late 1990s. In terms of future developments, however, understanding consumers’ choices of HEV is challenging because their preference construction process involves many aspects beyond comparing vehicle specifications. For instance, mileage per gallon, heavily depends on total vehicle use and its form; e.g., local versus highway driving. As a result, consumers who drive primarily on local roads are expected to prefer HEV more than those who mainly drive on highways. In addition to its impact on product performance, vehicle use may also influence the preference structure. Understanding the impact of vehicle usage can be as important as understanding the other heterogeneous aspects, such as the demographic and socio-economic attributes of consumers.

\section{2. Methodology: hierarchical choice modeling}

To capture the impact of vehicle usage attributes upon consumers’ choices of hybrid electric vehicles, a hierarchical choice modeling framework is employed. As seen in Fig. 1, the vehicle choice modeling framework utilizes a hierarchical modeling structure (Hoyle and Chen, 2009) that considers the impact of vehicle usage attributes $E$ by including them as the input of the choice model at the top level together with their bottom-up influence on the vehicle performance attributes which also serve as input of the choice model. With discrete choice analysis (Train, 2003), the choice utility is derived by...
assuming that the individual’s \((n)\) true choice utility, \(u\), for a design alternative, \(i\), consists of an observed part \(W\), and an unobserved random disturbance \(\varepsilon\) (unobserved utility).

\[
\begin{align*}
    u_{ni} &= W_{ni} + \varepsilon_{ni} = W(\beta : S_n, A_{ni}, E_n) + \varepsilon_{ni} \\
    P_{ni} &= \frac{\exp(W_{ni})}{\sum_j \exp(W_{nj})}
\end{align*}
\]  

The observed part of utility for respondent \(n\) and for alternative \(i\), \(W_{ni}\), is expressed as a function of consumer profile attributes \(S_n\), consumer desired vehicle attributes \(A_{ni}\), vehicle usage attributes \(E_n\), and the \(\beta\) coefficients, which are estimated by observing choices respondents make. Typical consumer profile attributes \(S\) include gender, age, household income, etc., while local/highway driving condition and miles driven daily are two of the commonly used vehicle usage attributes \(E\). Consumer desired vehicle attributes \(A\) refer to key vehicle features that influence consumers’ choice in selecting a vehicle. The inclusion of consumer profile attributes \(S\) and vehicle usage attributes \(E\), in addition to consumer-desired vehicle attributes \(A\), in the estimation of demand to capture the heterogeneity of consumer preference and their vehicle usage, is the key component in the hierarchical choice modeling framework.

Level I in Fig. 1 is the choice probability prediction function in multinomial logit model with Type I extreme value error distribution (Train, 2003), and level II represents the deterministic portion of the choice utility as a function of \(S\), \(A\), and \(E\). At Level III, the bottom of hierarchy, a separate prediction model is established to link \(A\) with \(X\), \(S\), and \(E\).

\[
A_{ni} = A(\alpha : S_n, X_i, E_n)
\]  

While product attributes are often fixed as constants for different consumers in conventional choice modeling, their dependence on consumers and usage context is considered here. For example, mileage per gallon (mpg) is one of the key vehicle attributes. Even though vehicle manufacturers provide target mpg measures under city and highway driving condition for each of their car models, the actual mpg value varies significantly from consumer to consumer because of the heterogeneous usage scenarios and consumer driving habits. Similarly, consumer ratings of vehicle performances are also influenced by individual profile attributes such as gender and age.

In Eq. (2), the coefficients \(\alpha\) can either be established by physical relations or determined through modeling. For quantitative attributes, the above model can be expressed by physical equations. When ratings are used to measure qualitative attributes, an ordered logit model can be used due to its capability of handling discrete data (He et al., 2011). By establishing a relation with vehicle design variables \(X\) through the hierarchical modeling framework, the obtained choice model can be used to support engineering design decisions (Wassenaar et al., 2005).

3. Data

While vehicle usage \(E\) plays an important role in consumers’ choices because both product performance and consumer preference change under various usage conditions, questions about the relationship between vehicle usage and consumer profile attributes remain. To address some of these, National Household Travel Survey (NHTS) data is used; this includes
demographic characteristics of households, people, vehicles, and detailed information on daily travel in the US for all purposes by all modes (FHWA, 2009). The data are collected from a sample of US households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and household attributes. The 2009 data set includes information on 150,147 households, 308,901 people, 309,163 vehicles, and 1167,321 trips.

The vehicle usage attribute miles driven daily, included in NHTS 2009 data, is of interest because it has a significant impact on choices of hybrid electric vehicles and plug-in hybrid electric vehicles (Shiau et al., 2010). Meanwhile, gender, age, household income, number of children living together, and education level are among commonly used consumer profile attributes. The data tells us that male respondents drive 44 miles on average (including long distance trips), 54% more than their female counterparts. Overall, the miles driven daily increases from five at age 16, peaks around 45 miles at ages 35–45, and decreases slowly afterwards. There is an increasing trend in miles driven daily with the increase of household income, although little can be said about the relationship between miles driven daily and number of children. Miles driven daily slowly increases with the increase of education level but drops at the end for advanced degree.

Observations of one-to-one relationship between vehicle usage and consumer profile attributes can provide guidance for analysis of variance (ANOVA) (Tamhane and Dunlop, 2000) whereby miles driven daily is the dependent attribute with the five consumer profile attributes as independent attributes. The results indicate that all consumer profile attributes are statistically significant but only contribute to about 10.1% of the total sum of squares suggesting that it should be treated as an additional dimension of consumer classification measure. Similar to the consumer profile attributes, vehicle usage attributes are traits of the consumers that can be used to categorize consumers into groups. For example, miles driven daily can be used to classify consumers into short, medium, and long distance drivers. Hence, both vehicle usage and consumer profile attributes are considered in market segmentation analysis.

4. Analysis

While consumer profile attributes should be included in consumer preference studies, including all consumer and vehicle usage attributes in the choice modeling process may cause problems in parameter estimation and lead to difficulties in interpretation. In many cases, significant attributes identified directly by choice model estimation are unstable and sensitive to the list of attributes included in the estimation. To build a reliable model, a reduced set of attributes is identified. We start with a list of consumer profile and vehicle usage attributes extracted from NHTS. Individual-level attributes include demographic and socio-economic attributes of the survey respondents, such as gender, age, race, education level, occupation category, working status, marital status, age of youngest child, and miles driven in the past 12 months. Miles driven in the past 12 months belongs to vehicle usage attributes, while all others are consumer profile attributes. Three attributes, working status (working or retired), marital status (single or married), age group of youngest child (no child, age 0–5, age 6–15, age 16–21) are derived from life cycle classification in the household file of NHTS 2009. Household-level attributes include information about households, such as household income, household size, number of adults, number of workers, number of drivers, number of vehicles, home ownership, home type, size of metropolitan area, urban/rural status of the home address, and life cycle classification.

4.1. Market segmentation through principal component analysis

Multi-collinearity between consumer attributes can create problems of identification of hybrid owners’ characteristics and understanding the influence of vehicle usage on consumers’ choices. To alleviate this, principal component analysis is used to reduce the dimensionality of the consumer attributes. Forty explanatory attributes of consumer profile and vehicle usage attributes extracted are used for principal component analysis and 18 factors were extracted explaining 76.4% of variance in the data and each having eigenvalues greater than unity. The five dominant attributes forming factor drivers and vehicles are household size, number of adults, drivers, vehicles, and non-single marital status. Dominant attributes in factor main stream family include individuals between age 35 and 44, youngest children under 15, and household size, and therefore indicate young to middle-aged households with children. Similarly, high income and education is heavily influenced by professionals with advanced degrees and high income. The factor long distance worker represents those who drive long distances on a daily basis; a high correlation between this factor and blue collar occupation suggests that blue collar workers tend to travel more than their peers.

4.2. Characteristics of hybrid drivers

The hybrid attribute (one for HEVs, zero otherwise) is used in categorizing and identifying potential HEV shoppers. Individual consumers and households are labeled based on the hybrid attribute of the vehicles they drive: if a consumer drives a HEV, he/she is labeled as hybrid driver (HD); if not, he/she is labeled as conventional driver (CD). A series of t-tests is used to investigate to see if a significant difference exists between HDs and CDs in the 18 factors identified in the principal component analysis. The results indicate that high income & education has a highly significant positive impact on hybrid drivers suggesting that household income and education level may contribute to the choice behavior. Long distance worker also has a higher mean factor score for the hybrid drivers, which may result from its positive correlation with gender and miles driven in the past 12 months suggesting that people’s attitude toward hybrid electric vehicles may relate to their gender and their
vehicle usage. Moreover, minorities such as African American and Asian tend to have more hybrid drivers. On the other hand, young with child seems to have a lower mean factor score for hybrid drivers than for conventional drivers, indicating that consumer profiles, such as age, marital status, and number of children may play a role in consumers’ choices.

5. Results

5.1. Compatibility of NHTS with Vehicle Quality Survey data

While NHTS data provides detailed information about households, individuals, vehicles, and daily trips, information about the choice set each consumer considered during the vehicle purchasing process is not available. Vehicle Quality Survey (VQS) data collected by JD Power and Associates does contain the details of the vehicles considered by each respondent, and is suited for building choice model. To ensure that the choice model built combining NHTS and VQS data reflects vehicle choice behavior of consumers, the compatibility of NHTS (2009) and VQS (2007) data is examined (Fig. 2).

NHTS and VQS share common vehicle usage attribute $E_{\text{com}}$ miles driven daily, while each of them has their own vehicle usage attributes: trip purpose in NHTS and local/highway indicator in VQS. The common consumer profile attributes $S_{\text{com}}$ shared by the datasets include: gender, age, household income, number of children, education level, race, and marital status. Individually, NHTS provides information about home ownership, home style, urban/rural location, etc., while VQS collects consumers’ height and weight.

The consumer profile attributes of interest include: gender, age, household income, number of children, education level, marital status, and race. The vehicle usage attribute of interest is miles driven daily (miles driven in the past 12 months in NHTS). All these attributes are discrete data. The consumer profile attributes, household income, and education level, are ordinal, while gender and race are categorical. The coding of ordinal and categorical data, for example household income brackets, in both datasets is adjusted to attain consistency. The numerical attributes are compared using Student t-test (Table 1). For each attribute, the mean attribute values for hybrid drivers and conventional drivers are given. Four attributes, age, household income, education level, miles driven daily (miles driven in past 12 months in NHTS divided by 365), are shown to have a significant positive impact on differentiating hybrid drivers from conventional drivers. The difference between HDs and CDs in number of children is not significant in the NHTS data. Comparing the mean values of each attribute, we see VQS data has much higher means for household income and education. This may be because the VQS samples mainly cover new vehicle owners in a specific calendar year, while the NHTS samples cover a broader range of vehicle owners in US.

As for the composition of NHTS and VQS sample populations, the percentage of hybrid drivers and conventional drivers are compared within subgroups of gender, race and marital status. For instance in the gender distribution, about 94.5% of the male respondent in the NHTS are CDs, while the percentage of HDs in the VQS sample population is much smaller. This observation is confirmed by the other two consumer profile attributes, race, and marital status. Further, no significant difference in percentage of hybrid drivers is noted between male and female populations. In the NHTS sample, African Americans exhibit a higher percentage of hybrid drivers, compared with the other three race groups. However, the VQS sample differs: the black community has a much lower percentage of hybrid drivers in VQS. No clear difference is seen between the married or single individuals.

5.2. HEV choice modeling based on pooled data

NHTS and VQS data are pooled for choice model estimation. While VQS is more suited for choice modeling purpose, NHTS is more representative of residents of the US. Even though a complete matching of attributes is desirable, there is no need to have the exact same set of attributes for model construction and prediction as the utility function underlying the choice model is expected to capture the impact of both vehicle usage attributes $E$ and consumer profile attributes $S$ across multiple datasets.

![Fig. 2. Vehicle usage and consumer profile attributes in NHTS and VQS.](image-url)
In the 2007 Vehicle Quality Survey, vehicle purchase data from 90,000 nation-wide respondents of over 300 vehicles in the market are collected, including data for 11 HEV models. Further, respondents’ demographic attributes and their vehicle usage are recorded. For model estimation, data collected from 8025 respondents, who responded explicitly regarding three other vehicles considered in their choice set in addition to the vehicle they chose, are considered. As for the NHTS 2009 data, 15,973 individuals with 2007 model-year vehicles are selected for pooled choice model estimation. Since this data provides no information about other vehicles considered by the respondents, three vehicles other than the one purchased are randomly selected from a set of 262 car models based on a uniform distribution to compose an individual choice set of four vehicles.

Fifteen consumer-desired vehicle attributes \(A\) are selected including six common consumer-desired attributes \(A^{com}\), price, vehicle origin, vehicle size, vehicle type, mileage per gallon (mpg), hybrid electric vehicle indicator, and nine VQS-specific \(A^2\), rating scores given by the respondents. The attribute “price” is the money respondents paid, excluding tax, license, trade-in value, etc. Vehicle origins are categorized as domestic, European, Japanese, and Korean; vehicle sizes are grouped into compact, midsize, large, and premium; vehicle type includes mini, car, sport utility vehicles (SUV), minivan, van, multi-activity vehicles (MAV), and pickup. The hybrid electric vehicle indicator, coded as one for hybrids, and zero for conventional vehicles, reflects consumers’ attitude toward new hybrid technology. In VQS, nine aspects of the vehicle, including exterior attractiveness, interior attractiveness, storage and space usage, audio/entertainment/navigation system, seats, heating ventilation and air conditioning, driving dynamics, engine and transmission, and visibility and driving safety, are rated in an ascending scale from one to ten. These discrete ratings are included in the choice modeling procedure under \(A\), because they are considered to be a good measure of consumers’ perception of qualitative as well as quantitative vehicle attributes.

As for the vehicle usage attributes \(E\), two most commonly considered vehicle usage attributes for HEV are included in the choice model: \(local/highway\) \(indicator\) and \(miles\ driven\ daily\). It should be noted that the \(local/highway\) indicator is imputed by comparing the combined mpg published by the US Environmental Protection Agency (EPA, 2008) and the estimated mpg given by survey respondents in the VQS data. The indicator is a continuous parameter, ranging from zero for local driving to one for highway driving, and assumed to reflect the general driving conditions the respondents face, therefore the vehicle usage. Hence, as part of the level III in the hierarchical choice modeling framework, consumer desired attribute \(A_2, mileage\ per\ gallon\), is:

\[
A_2 = \frac{1}{\frac{1}{mpg_{city}} + \frac{1}{mpg_{highway}}} \tag{3}
\]

where \(mpg_{city}\) and \(mpg_{highway}\) belong to the vehicle design variable \(X\). The other vehicle usage attribute considered is the \(miles\ driven\ daily\), a popular descriptor of consumers’ travel pattern. The data is derived from the recorded miles driven in the first 3 months from the market survey.

Gender, age, household income, number of children under age 20 living together, education level, race, and marital status, are included as consumer profile attributes \(S\). Numerous combinations of consumer attributes are tested in the choice modeling process. From the final choice model estimation results, only two consumer profile attributes, household income and education level, are statistically significant. All consumer profile attributes are included in the ordered logit regression for predicting the performance rating scores. All correlations between consumer profile and vehicle usage attributes are between –0.34 and 0.32, justifying the inclusion of multiple attributes in choice model.

The utility function used is shown in Eq. (4), where interactions between \(A, E, \) and \(S\) are explicitly modeled. \(W_1, W_2\), and \(W_{pooled}\) stand for utility function in NHTS, VQS, and pooled data. The attributes in choice modeling include common consumer-desired attributes \(A^{com}\), VQS-specific \(A^2\), common vehicle usage attributes \(E^{com}\), VQS-specific \(E^2\), and common consumer profile attributes \(S^{com}\). A scale parameter \(\mu\) is introduced in the pooled utility, \(W_{pooled}\), to account for variation

\[\mu = \text{constant} \tag{4}\]

\[\mu = \text{constant} \tag{4}\]

\[\mu = \text{constant} \tag{4}\]

\[\mu = \text{constant} \tag{4}\]

\[\mu = \text{constant} \tag{4}\]
difference of error terms from two datasets. The structure of pooled utility function for choice modeling is similar to that used in nested logit model (Koppelman and Bhat, 2006). Alternative specific constants (ASC) for each of the car models are not included in the utility function. While this may decrease the goodness-of-fit of the model, it allows choice prediction of newly introduced vehicle.

\[
W_1 = \beta_{1A} A_{com} + \beta_{1AS} A_{com} \cdot S_{com} + \beta_{1AE} A_{com} \cdot E_{com}
\]

\[
W_2 = (\beta_{2c} A_{com} + \beta_{2} A^2)(1 + \gamma_{2S} S_{com} + \gamma_{2E} E_{com} + \gamma_{2E}^2)
\]

\[
W_{pooled} = W_1 + \frac{1}{\mu} W_2
\]

From Table 2, we see that most choice coefficients are significant and that the price/income coefficient is negative as expected. The second attribute 100/miles per gallon reflects the amount of gasoline needed to drive 100 miles. The negative parameter for 100/miles per gallon indicates that vehicle usage attribute has a negative impact on consumers’ preference; people primarily driving on highways tend to care more about the mpg value. Moreover, the attitude toward HEVs itself has a positive coefficient, showing that people driving locally tend to favor HEVs. Similarly as we expected, highway drivers’ preference towards HEVs are weaker; indicated by a negative coefficient for local/highway and HEV indicator interactions. The positive coefficient of HEV indicator and education level interaction suggests that people with higher education are more likely to prefer HEVs.

6. Conclusions

We used both NHTS and VQS data to examine consumer preferences for hybrid electric vehicles. Our ANOVA results suggest that vehicle usage is an additional dimension of overall consumer classification because consumers’ profile attributes

### Table 2

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price(k)/income</td>
<td>-0.1194*</td>
</tr>
<tr>
<td>100/miles per gallon</td>
<td>-0.3338*</td>
</tr>
<tr>
<td>100/miles per gallon + local/highway indicator (VQS only)</td>
<td>-0.5613*</td>
</tr>
<tr>
<td>100/miles per gallon + miles driven daily*</td>
<td>-0.0012*</td>
</tr>
<tr>
<td>Vehicle origin (domestic as base)</td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>0.4729*</td>
</tr>
<tr>
<td>Japanese</td>
<td>0.1544*</td>
</tr>
<tr>
<td>Korean</td>
<td>0.2816*</td>
</tr>
<tr>
<td>Vehicle size (compact as base)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-0.1855*</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.0167*</td>
</tr>
<tr>
<td>Premium</td>
<td>0.1240*</td>
</tr>
<tr>
<td>Vehicle type (Car as base)</td>
<td></td>
</tr>
<tr>
<td>MAV</td>
<td>-0.3463*</td>
</tr>
<tr>
<td>Mini</td>
<td>0.1222*</td>
</tr>
<tr>
<td>Minivan</td>
<td>-0.2939*</td>
</tr>
<tr>
<td>Pickup</td>
<td>-0.0110</td>
</tr>
<tr>
<td>SUV</td>
<td>-0.1680*</td>
</tr>
<tr>
<td>VAN</td>
<td>-1.8446*</td>
</tr>
<tr>
<td>Minivan</td>
<td>0.0713*</td>
</tr>
<tr>
<td>Hybrid electric vehicle</td>
<td></td>
</tr>
<tr>
<td>Hybrid electric vehicle + local/highway indicator (VQS only)</td>
<td>-1.9586*</td>
</tr>
<tr>
<td>Hybrid electric vehicle + miles driven daily</td>
<td>-0.0010</td>
</tr>
<tr>
<td>Hybrid electric vehicle + education level</td>
<td>0.0518</td>
</tr>
<tr>
<td>Rating (VQS only)</td>
<td></td>
</tr>
<tr>
<td>Exterior attractiveness</td>
<td>0.0131*</td>
</tr>
<tr>
<td>Interior attractiveness</td>
<td>0.1069*</td>
</tr>
<tr>
<td>Storage and space usage</td>
<td>0.1297*</td>
</tr>
<tr>
<td>Audio</td>
<td>0.0363*</td>
</tr>
<tr>
<td>Seats</td>
<td>0.0333*</td>
</tr>
<tr>
<td>HVAC</td>
<td>0.0273*</td>
</tr>
<tr>
<td>Driving dynamics**</td>
<td>0.0495*</td>
</tr>
<tr>
<td>Engine and transmission</td>
<td>0.0737*</td>
</tr>
<tr>
<td>Visibility and safety</td>
<td>0.0154*</td>
</tr>
</tbody>
</table>

* Significant at 0.01.
** Driving dynamics surveys consumers’ perception of ride smoothness in normal driving, quietness over harsh bumps, responsiveness/effort of steering system and braking, handling/stability on curves or winding roads and in adverse conditions.
cannot fully explain differences in vehicle use. Using principal component analysis, 40 consumer profile and vehicle usage attributes were grouped into 18 distinctive factors such as drivers and vehicles, mainstream family, working class, high income and education, etc. Key characteristics of hybrid electric vehicle drivers were identified by comparing the mean factor scores within the hybrid and conventional driver groups, which laid the foundation for segmenting the market based on both consumers profiles and vehicle usage. A discrete choice model then allowed analysis of consumers’ preferences for HEVs.

Acknowledgments

Grant support from National Science Foundation (CMMI-0700585 and DUE-0920047) and from the Initiative for Sustainability and Energy at Northwestern are greatly appreciated.

References