EVS30 Symposium
Stuttgart, Germany, October 9 - 11, 2017

Investigation of Consumer Preference and Dynamic Effects in the Emerging US Market for New Vehicle Energy Technologies using Longitudinal Consumer Survey Data

David S. Bunch[1], Rubal Dua[2]

[1]*Graduate School of Management, University of California, Davis, One Shields Ave., Davis, CA 95616, dsbunch@ucdavis.edu*

[2] *King Abdullah Petroleum Science and Research Center (KAPSARC)*

Summary

This paper reviews concepts from existing modelling approaches developed in the context of consumer demand for new conventional vehicles, and considers how these may be expanded and extended to incorporate consumer adoption processes for new energy technologies. A motivation for the work is access to data from a large-scale longitudinal survey of new vehicle buyers containing a substantial amount of (revealed preference) data on purchases of recently introduced plug-in vehicles (BEVs and PHEVs), as well as data on consumer attitudes and other factors. Initial results from a first-stage discrete choice model for the entire US new vehicle market that directly incorporates plug-in purchases are presented, along with consideration of how the data can be used to support more advanced forms of choice modelling.

Keywords: BEV, consumer, demand, modelling, PHEV

1 Introduction

In the not-too-distant past, policy makers responsible for addressing transportation-related issues (e.g., for personal vehicles) were primarily concerned about mitigating impacts of criterion pollutants, but also energy usage (through policies such as CAFE/GHG standards in the US) for reasons that included energy security. However, there is now a rapidly growing consensus at a global level that it is critically important for energy systems be decarbonized to levels well below what might possible from incremental improvements to existing technologies. Scenarios that meet desired levels typically require a major transition to new alternative energy technologies (e.g., plug-in vehicles, fuel cells). There are major policy questions surrounding how such a transition can be successfully supported, particularly in locales where the consumer response plays a major role in determining outcomes.

One frequent approach to policy analysis is to use quantitative models that produce projections of consumer sales under alternative scenarios. This approach has been used with some degree of success in the past, and there is a vast literature based on using so-called revealed preference (RP) data on real-world consumer purchases to produce such models. In more recent work, models have been developed with increasingly higher levels of detail to address a variety of policy and methodological requirements that arise in this highly differentiated product market. This success has been due at least in part to the fact that the fundamental market structure has been relatively stable.

However, even for this case there have been ongoing challenges. For example, there is still little consensus on exactly to what degree degree factors as such fuel efficiency and vehicle performance affect consumer market behavior. Projecting how new technologies with very different operating characteristics and use cases will fare when introduced into the existing system, and what the subsequent diffusion dynamics might look like, are much more challenging. Analyzing these questions has generally relied on alternative modeling approaches that (although based on the same underlying theories and methodologies as RP-data-based models) operate at a much lower level of detail (e.g., focusing on "fuel technology choice"), and typically rely much more heavily on theory-based assumptions supported by insights from stated preference (SP) data or other survey-based techniques that ask consumers to respond to hypothetical situations.

From a research perspective, we are now in a very interesting and quickly changing environment where these two "modeling worlds" must inevitably merge. After many years of policy initiatives and industry developments, vehicles using new energy technologies have finally been introduced into the marketplace, so that RP data are becoming available. At the same time, most of these purchases are being made by, e.g., "innovators" or "early adopters," so that extreme care must be taken when attempting to develop models for purposes of projection or extrapolation to the wider market.

This paper reports on current results from a recently initiated research program that specifically seeks to support model development during this transition period in a way that bridges the gap between more established work based on RP-data (largely limited to conventional vehicles, but addressing total market effects), versus more specialized studies based on either SP data and/or in-depth explorations exclusively limited to early purchasers of new technology vehicles. These efforts are feasible due to the availability of large-scale surveys of consumers in the US new vehicle market.

The paper first briefly reviews background and concepts required to bridge these two worlds, followed by an overview of the longitudinal survey data whose availability motivated this work. Some aspects of recent market dynamics captured by these data are demonstrated. Most of the progress thus far has focused on development of data sets and discrete choice model estimation results that use highly detailed vehicle definitions and household segmentation effects that have recently been developed for conventional vehicles, and extending them to include RP data on plug-in vehicle purchases using the aforementioned consumer survey data. We conclude with a discussion of how this work can be expanded more directly address dynamic effects and the role of consumer attitudes.

2 Consumer Preference, Vehicle Choice, and Dynamic Effects

Previous research has explored and confirmed that the household decision-making process for vehicle choice can be affected by a variety of factors. Many of these are based on the fundamental mobility needs of households, which vary by household type and are affected by demographic and locational factors. Moreover, much of their observed behaviour can be explained using models that are well grounded in economic theory. However, research also indicates that past experience, awareness and knowledge, attitudes and perceptions, and even social interactions and network effects can also play a role, particularly when it comes to their reaction to new types of products and technologies.

Initial efforts in this project have centred on development and estimation of discrete choice models for new vehicle markets that incorporate a high level of detail, both for definition and characterization of vehicle choice options, as well as household segments that incorporate many demographic and locational effects. These models are a logical starting point, relying on past research (including our own) that has been used successfully for policy analysis requiring a high degree of rigor, which is considered a pre-requisite for the goals of this project. In particular, addressing number and diversity of make/model offerings is potentially a very important factor in producing high-quality projections in a highly differentiated market. These preliminary modelling results (presented in section 4) comprise the bulk of the paper.

At the same time, a key goal is to move in the direction of also incorporating processes related to consumer adoption of new technologies. These processes typically emphasize key roles of played by development of awareness and knowledge levels, and he these and perceptions might dynamically evolve over time. In addition the diffusion literature has relied on an aspect of consumer segmentation that are primarily based on attitudes that might not necessarily be coincident with more easily observed demographic characteristics. Consumers are divided in anywhere from three to five "groups," where a three-group scheme is illustrative: Early Adopters (Innovators), Early Majority, and Late Majority. In the alternative-fuel vehicle literature, two effects are frequently mentioned as drivers of early adopter behaviour: interest in technological innovation, and a commitment to environmental friendliness. With regard to vehicles, such consumers would be more likely to aggressively seek out information, achieving a high level of awareness and knowledge regarding new offerings, which is a prerequisite for consideration and then purchase.

In contrast, late majority consumers in the vehicle space are likely to have low levels of awareness and knowledge. Only when new technology vehicles have achieved a sizeable share of the market so that these consumers would inevitably have direct experience of them might their awareness levels shift. Moreover, even if these consumers have some level of awareness of new technologies, they may be likely to view such offerings as "risky" until they have been established in the market long enough to have been "proven."

3 Overview of Longitudinal Consumer Survey Data

As noted in the introduction, a major factor in pursuing this work is the availability of Strategic Vision's New Vehicle Experience [NVES] Survey. It is now being conducted on an ongoing basis, where each Study Year (SY) coincides with the time period associated with a traditional vehicle model year (MY), i.e., October (MY-1) through September (MY). There are four waves per SY (one wave per quarter, although earlier versions did not necessarily cover all quarters). Our current focus is on expanding consumer models to effectively address demand for plug-in vehicles, for which initial US sales occurred in December 2010. The sampling frame is based on detailed vehicle sales/registration information from multiple sources (e.g., vehicle manufacturers, Polk) and essentially covers all new vehicles sold to consumers. The sample is generated using a complex process involving quotas, where quota cell definitions can vary in their level of detail (e.g., minimally, Make/Model). In contrast to simple random sampling, such an approach may oversample specific types of vehicles to ensure that sufficient data have been obtained to draw conclusions.

The sample sizes for this survey (roughly 250-300,000 per Study Year) are much larger than academic researchers working in this field have typically had access to. For example, the 2009 National Household Transportation Survey (NHTS) includes roughly 150,000 households generally be viewed as a random sample from the entire population (i.e., a relatively small proportion would have purchased new vehicles within the 14-month survey period). Other academic researchers using survey data from new car buyers are frequently limited to a few hundred observations collected over a period of a few months.

The NVES collects a large amount of information from each consumer, including: demographics, location, current vehicle holdings, details of the recent purchase (including trade-in information), and other vehicles considered for purchase. Additional data include responses to a variety of attitudinal, perception, and preference-related scaling questions. The amount of data per respondent can vary depending on which version of the survey they ultimately receive. And, of course, for surveys of this length the level of item non-response can vary widely. For this reason, work presented here is based on a subset of key demographic and locational variables, as well as three attitudinal indicators, that have been selected so as to minimize the number of respondents dropped from the analysis.

Before proceeding to the section on discrete choice modelling, we first review some of the properties of the NVES data with the potential to support future development of more advanced models that could aid in identification of structural dynamic effects. First, note that we are currently focusing on four fuel technology types: Gasoline, Diesel, Hybrid (a.k.a. Hybrid Electric Vehicle or HEV), Electric (BEV), and Plug-In Electric Hybrid (PHEV). (Note: Flex-Fuel vehicles have been subsumed under Gasoline for our purposes, and, although the survey contains data on a small number of natural gas vehicles, these have been omitted.)

When considering the current evolution and progress of the plug-in vehicle market, various researchers have tracked the progress of new model introductions, cumulative sales levels, and ways in which markets vary around the world. The NVES data support a similar approach when focusing on the US market as a whole, but with the potential of identifying individual-level behavioural effects due to the large sample sizes. Although it would be possible to generate a variety of graphs and charts reviewing the recent dynamics of the plug-in market using these data, to conserve space we instead provide a brief description: there were two initial plug-in offerings in the US market in December 2010: Chevrolet Volt (PHEV), and Nissan Leaf (BEV). The Tesla Model S was introduced in 2011. The total number of plug-in model offerings increased to about 20 in 2014 (roughly evenly divided between BEV and PHEV), and then to about 25 near the beginning of 2016. All of these vehicles are "passenger cars," with the only "large car" offering being the Tesla. The Volt and Leaf are midsize, and many of others fall into the smaller size categories (Mini-compact, Subcompact, Compact, and even a Two-Seater). (Note: Identifying size classes and other market segmentation considerations are addressed in section 4.)

In contrast, the number of gasoline "models" has generally hovered in the 275-325 range, covering a wide variety of vehicle types (all sizes and styles of cars, as well as small and large pickups, SUVs and minivans). It is also important to note that gasoline models are more highly differentiated, coming in many more distinct configurations (i.e., many combinations of engines, body styles, transmissions and drivetrains), so the number of distinguishable offerings for any particular model year is well over 1,000. The number of Hybrid models has also increased during this period (2010 to 2016), from 20 in 2010 to about 30 in late 2015. (At this time, we forego any discussion of diesel.)

In tracing out the sales patterns of plug-ins during this period, sales start at zero in 2010, and on a percentage increase basis, rise rather rapidly. In accordance with theory, these sales increases occur roughly in parallel with the introduction of new models. Still, as a share of the entire US market, these vehicles only very recently approached (or perhaps exceeded) the 1% threshold. However, the shares also vary widely across regions and locales within the US market, with California being noteworthy of consideration. The availability of the NVES data and the large sample sizes have the potential to support identification of dynamic effects based on both cross-sectional and inter-temporal variation in the US market.

To capture a range of locational effects, we assign each household to a geographic region based on a modified version of the US Energy Information Agency (EIA) Petroleum Administration for Defense Districts (PADDs). The standard version uses 5 regions (East Coast, Midwest, Gulf Coast, Rocky Mountain, West Coast); however, we split out California as a separate region. Fuel technology type sales shares by region are summarized in Tables 1 and 2 for SY2015 and SY2016, respectively.

Table 1: Estimated Sales Shares for Fuel Technology Types By Region (SY2015)

FuelTechType	EastCoast	MidWest	GulfCoast	RockyMtn	WestNoCal	California	Total US
Gas	96.70%	97.20%	97.40%	95.80%	93.40%	89.70%	95.70%
Diesel	0.70%	0.60%	0.70%	1.10%	1.70%	1.40%	0.80%
Hybrid	2.00%	1.90%	1.40%	2.20%	3.60%	6.60%	2.60%
BEV	0.50%	0.20%	0.40%	0.60%	0.90%	1.20%	0.50%
PHEV	0.20%	0.20%	0.10%	0.30%	0.50%	1.10%	0.30%
	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table 2: Estimated Sales Shares for Fuel Technology Types By Region (SY2016)

FuelTechType	EastCoast	MidWest	GulfCoast	RockyMtn	WestNoCal	California	Total US
Gas	96.50%	96.60%	96.80%	93.00%	92.70%	89.20%	95.20%
Diesel	0.70%	0.60%	1.10%	1.60%	1.50%	1.30%	0.90%
Hybrid	1.80%	2.10%	1.50%	2.10%	2.70%	4.00%	2.20%
BEV	0.30%	0.30%	0.30%	2.20%	2.00%	2.60%	0.80%
PHEV	0.60%	0.40%	0.30%	1.00%	1.10%	2.80%	0.90%
	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

These side-by-side comparisons illustrate the existence differences across regions, as well as the accelerating but uneven dynamics that have occurred over the past year.

Finally, we provide a specific example that relates to the potential for incorporating consumer attitude measures from NVES into quantitative models. As discussed in section 2, theory suggests that when new technologies are introduced to the market, the initial purchasers of these technologies will be early adopters/innovators. These consumers are viewed as being particularly attracted to products that are technologically innovative. However, as sales continue over time and other consumer segments become aware and more knowledgeable about such products, they become perceived as "less risky" and "more mainstream," so that some consumers in these groups consider them as purchase options. The NVES survey includes a 5-point importance scale question for technological innovation (5 = Extremely important, 1 = Not at all important). Theory suggests that, as new technologies diffuse into the market, the average scores among the consumers choosing to purchase them should therefore decline. This effect is demonstrated in Figure 1.

The average importance scores for purchasers of BEVs and PHEVs are extremely high in 2011 compared to the other fuel technologies. Among the three incumbent technologies, Hybrid averages are higher than Diesel, which are in turn higher than Gasoline. The averages for the incumbents seem to have converged in 2014 through 2016. At the same time, averages for both BEVs and PHEVs systematically decline over time, and even begin to converge with those of the incumbents in SY2016. In section 4.6, we explore the implications of incorporating this importance rating (as well as two others) in the context of discrete choice modelling, which we turn to next.

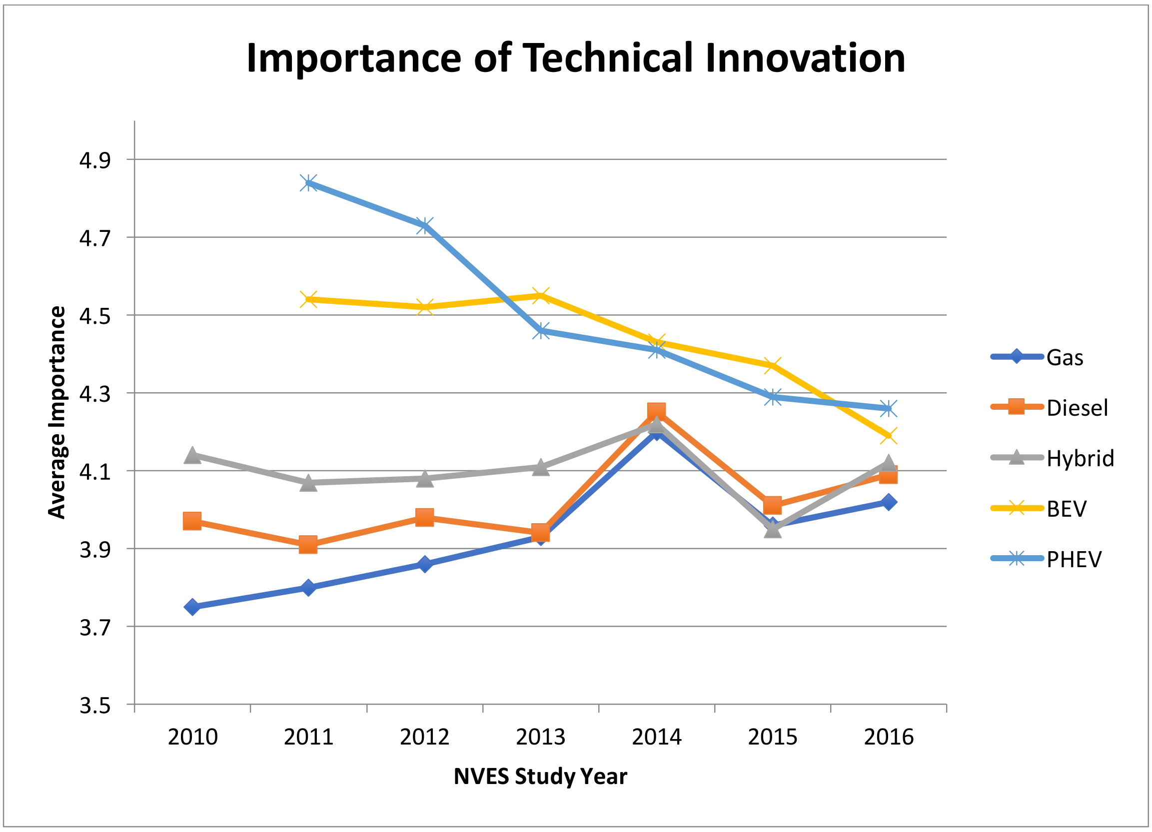

Figure 1: Average Importance of Technological Innovation by Fuel Technology Type

4 Discrete Choice Models of the US New Vehicle Market

This section presents the results of discrete choice model development and estimation efforts using NVES data for a specific quarter: Quarter 2 (April-June) of 2015.

4.1 Overview of Discrete Choice Modelling

Although a complete treatment of discrete choice models is beyond the scope of this paper, approach considered here is based on a random utility maximization (RUM) framework, where choice is assumed to arise from a process in which an individual consumer (i) evaluates the "utility" of each option in a set C of J competing alternatives, and (ii) chooses the option c with the largest utility. Based on behavioural theory and other considerations, the analyst specifies a utility function that assumes consumer preferences are based on attributes of the options, and that preferences can vary based on characteristics of consumers/households. Let x_j be a vector vehicle attributes for vehicle j, $j = 1,\ldots, J$, and d_n a vector of consumer characteristics for consumer n. Consumer n's utility for (vehicle) option j is modelled as:

$$U_{jn} = V(x_j, d_n; \beta) + \varepsilon_{jn}, \; j = 1,\ldots, J \tag{1}$$

where β is a vector of parameters that represent consumer preferences. The choice probability of option c for consumer n can be formally expressed as: $P_{cn} = Prob\{U_{cn} \geq U_{jn} \text{ for all } j\}$. By making assumptions on the mathematical form of V and the probability distribution of ε_{nj} it is possibly to specify various families of discrete choice models. Note that the interpretation of this utility is associated with so-called conditional indirect utility from microeconomic theory. For analytical work, the form of V is often represented using the linear-in-parameters form $V(x_j, d_n; \beta) = f(x_j, d_n)'\beta =$ $z_{jn}\beta$, i.e., z_{jn} is a vector of explanatory variables based constructed as a function of attributes and/or consumer characteristics. The most widely used discrete choice model is multinomial logit (MNL), which yields choice probabilities using a simple, closed form expression. More complex model forms can be used to capture unobserved heterogeneity of consumer preferences. These include Nested Logit, and Mixed Multinomial Logit. An accessible reference on discrete choice models is [1].

There are many technical issues to be considered when developing discrete choice models for household vehicles that are beyond the scope of this paper. For example, the above description has clearly been provided to be consistent with our focus on new vehicle purchase models. However, a very real question is whether the scope of decision making should be expanded to address the entire household fleet (how many vehicles to own, and which ones), whether the emphasis should be on vehicle holdings or individual transactions, how to address the decision of whether to purchase new or used vehicles, and the role of vehicle utilization. For a more detailed treatment of how discrete choice models are applied to automobile demand and type choice, see [2].

4.2 Vehicle Market Definition

Based on section 4.1, the first requirement is to establish an attribute-based definition of the new vehicle offerings available for purchase during the relevant time period. As discussed previously, we have elected to pursue a rigorous approach using a high level of detail based on a number of considerations. In our current approach, individual vehicle offerings (denoted "core vehicles" [CVs]) have been identified that are clearly distinguishable from one another in terms of their physical functionality (e.g., vehicle type, size, body style) and operating characteristics (e.g., fuel efficiency and performance) as well as the make/model "families" to which they belong. In terms of specifics, our CVs generally correspond to the vehicle offerings itemized in the fueleconomy.gov database (but with some modifications). At this level of detail, it is possible to think of any particular CV as being available in multiple versions, due to additional "trim levels" or "option packages" that do not materially affect the core vehicle characteristics. However, because vehicle price is a critical determinant of consumer choice, completely ignoring the effect of trim levels in favour of "basic configurations" is ill advised. In the current approach, prices are averaged over trim levels, and the number of trim levels is recorded as an additional attribute. (Other approaches could be pursued in the future, but a full discussion is outside the current scope.) Constructing CVs required extensive processing and merging of data from multiple sources: fueleconomy.gov, Edmunds, the NHTSA VIN decoder, and also some information contained within the NVES itself. This process also incorporated aspects of the FARS system when developing vehicle definitions, because of the need to create a consistent and stable framework suitable for addressing multi-year longitudinal data.

One important practical issue is that, in real-world markets, vehicle offerings from multiple model years frequently available for sale as "new vehicles" within the same time period. Although most researchers have typically assumed away this problem, we have chosen to address it directly in our work. A detailed discussion is beyond the scope of this paper. Our current approach defines the vehicle market using offerings from multiple MYs based on our own detailed analysis (counts will be reviewed below), but this is a subject that will definitely require re-visiting. Note that expanding the modelling framework to span multiple quarters will require re-defining the vehicle market as a function of time.

To summarize, each CV is generally defined on the basis of the following: ModelYear, FARSMakeModel, FuelTechnologyType, EngineCharacteristics, DrivetrainType (AWD/4WD or not), TransmissionType (Automatic or Manual), VehicleClass and BodyStyle (discussed below). Generic attributes for each CV currently include Combined MPG, and Horsepower-to-Weight ratio (for traditional technologies). At this level of detail, MPG and HPtoWt are often (but not always) uniquely determined by the above definition (otherwise they are averages). A critically important attribute is PurchasePrice, which is always a challenge for vehicle choice modelling. Our current approach incorporates averages of purchase prices reported by the NVES respondents within each quarter. (A complete discussion, including the problem of missing values, is beyond the scope of this paper.)

Term VehicleClass used here represents a specific segmentation scheme based on a combination of vehicle type and size, and whether the vehicle is a "prestige brand." Vehicle type and size are based on the EPA's classification system, with some collapsing of size classes for passenger cars. The final scheme used here includes 15 classes. With regard to BodyStyle, cars within a segment can be further characterized using four levels: convertible, hatchback, sedan/coupe, and wagon. [The current system includes three types of cabs for pickups trucks (Regular, Crew, and Extended) although they are not yet being used.]

In all, the number of CVs in the 2015-Q2 market definition is 1,210, i.e., this is the size of the choice set (*J*) from which each household is assumed to make a purchase. The distribution by MY is: 258 (2014), 853 (2015), and 99 (2016). One might consider limiting the options to MY2015 offerings so that $J = 853$. However, this could create gaps in the coverage of offerings, and would also distort what is really going on in the market. Note also that most researchers have often limited their market definitions to "representative make/model" so that a typical value for *J* might fall in the 200-300 range.

4.3 Household Segmentation

Households are characterized using a combination of demographic and locational factors. For the current models, five annual income levels are used: < \$40K, \$40-80K, \$80-200K, > \$200K, and "missing" (to reduce loss of observations). Additional variables capture other demographic factors known to affect vehicle choice: household size, number of adults, presence of children, work/retirement status, and level of education (college grad). For locational effects, the 6-level regional variable discussed previously is used, as well as four levels of location type: Urban, Suburban, Rural, and missing. The segmentation approach relies heavily on the use of dummy variables, to allow for creation of specific "household types" to which multiple respondents can be assigned. In many cases this can lead to computational efficiencies (although see the discussion in the results section).

4.4 Methodological and Estimation Considerations

The requirements for estimating the models considered here are substantial, involving both econometric and numerical computational considerations that are highly technical and beyond the scope of this paper. One specific econometric issue is that the sampling scheme employed by NVES falls into the realm of endogenous/choice-based sampling, so that straightforward application of standard estimators developed under the assumption of random samples cannot be used. For a reference that serves as a starting point for this topic, see [3] and references cited therein. The estimation results presented here are based on a version of weighted exogenous sample maximum likelihood (WESML) developed specifically for these data. Additional work (not reported here) uses an alternative conditional maximum likelihood (CML) estimator. With regard to numerical computation, even a single quarter of NVES data includes close to 70,000 choice observations. Combined with the very large choice set size (see section 4.2) the estimation task is much larger than can be done using standard, commercially available software. Results were obtained using special-purpose software programmed by the authors using a combination of MATLAB and FORTRAN. Finally, the results presented here use the MNL model, although the intention is that this will be extended to Nested Logit to capture additional factors related to similarity and substitutability associated with known vehicle market structure.

4.5 Estimation Results

Producing reasonable estimation results for data of this type requires a relatively large number of parameters in order to capture a wide range of effects that, if ignored, would lead to substantial multicollinearity and bias for coefficients for key effects of interest. Specifically, the model specifications we are using include in the neighbourhood of 100 parameters. Many of these can viewed a "fixed effects" (in econometric terminology). We currently use a full set of dummy variables associated with 15 Vehicle Classes and 35 Brands (for a total of 48 coefficients). (Although the Brand coefficients could of interest, they are omitted here because their contribution to the discussion does not justify the space required.)

To begin, Table 3 reports coefficient estimates for key main effects (plus some interactions). Specifically, the first column includes estimation of coefficients for VehicleClass dummy variables. On one hand, these results appear to have implications that one might expect. For example, with the exception Standard Two-Seater vehicles, these estimates indicate that all other segments are preferred to the base class (Prestige Large Cars). Moreover, the relative coefficient sizes are consistent with recently demonstrated preferences by consumers for all types of SUVs as well as Standard Pickup trucks. On the other hand, due to the complexity of how other effects might interact with these, they should be interpreted with caution. On a related note, estimates at the bottom of column 2 indicate that, relative to standard sedans/coupes, the other car body styles (convertible, hatchback, wagon) are less preferred. One potentially problematic area for future consideration is the ambiguity surrounding how vehicles are assigned with respect to the heading "SUV/crossover." In the current vehicle definitions, many vehicles that might otherwise be denoted wagons/hatchbacks actually belong to the SUV category. The remainder are very clearly identifiable in the most traditional sense as "wagons" (or hatchbacks).

Table 3: Coefficient Estimates for Vehicle Classes and Primary Attributes

Variable	Coeff Est	Variable	Coeff Est
Large_Car_Prestige [Base]	0.000	Price [000's, Base = 40to80K]	-0.046
Large_Car_Standard	0.440	Price*LT40K	-0.050
Midsize_Car_Prestige	0.564	Price*80to200K	0.038
Midsize_Car_Standard	0.716	Price*GT200K	0.070
Minivan	0.270	Price*Missing	0.040
Small_Car_Prestige	0.495		
Small_Car_Standard	0.113	FuelCost (cents per mile)	-0.284
Small_Pickup	0.531	FuelCost*College	-0.051
Small_SUV_Prestige	1.318	LogNumTrims	1.192
Small_SUV_Standard	0.998	HPtoWeight	9.686
Standard_Pickup	1.404	AWD/4WD	0.430
Standard_SUV	1.540		
Standard_SUV_Prestige	2.278	Convertible	-0.570
Two_Seater_Prestige	0.103	Hatchback	-0.689
Two_Seater_Standard	-1.372	Wagon	-0.404

Of potentially greater interest are the coefficients for PurchasePrice and FuelCost. Similar to many model specifications appearing elsewhere, we have interacted price with dummy variables for various income levels. The baseline price coefficient corresponds to the $40-80K annual income group. These results indicate that a consumer in this income group (without a college education) would on average be "willing to pay" an additional $6.2 thousand in vehicle purchase price for a one-cent-per-mile improvement in fuel operating cost. (The amount increases by $1,000 for those with a college education.) For this to be "economically rational" would require a consumer that, e.g., drives 31,000 miles per year for 20 years, and values fuel savings with no discounting. (However, this also assumes that the consumer would expect fuel prices to stay constant at current levels). Although this estimate might seem high, it is remarkably close to estimates we have obtained with completely different data sets using similar techniques.

Another issue highlighted in Table 3 is the general challenge in estimating purchase price coefficients. In these results, the coefficient for households in the lowest income category is roughly double that for the base category, and the coefficient for the $80-200K group is getting rather close to zero. The coefficient for the high-income group is actually *positive* in this current version of the model! Although it would have been possible to find some way to bury this issue, we have elected to report these current results as-is. Much of our own recent work, as well as results by other researchers, demonstrate that estimation of price coefficients is still a major challenge, and one that we hope to address more aggressively as our program develops further. For example, expanding the model estimation to include data from additional quarters and years is potentially promising. We would note that coefficients on HPtoWeight and AWD are also remarkably similar to those we have obtained using other data sets.

However, one major difference from our previous work is the availability of substantial levels of plug-in vehicle sales. Note that the "base level" for fuel technology type in this analysis is Gasoline (as one might expect). Relative to the coefficients in Table 3, additional interaction effects have been estimated to capture differences from baseline utilities related to both fuel technology type and also other demographic effects. Table 4 includes interactions that have been specified for the entire sample (i.e., regardless of geographic region).

Table 4: Coefficients for Demographic and Fuel Technology Interactions

Variable	Coeff Est	Variable	Coeff Est	Variable	Coeff Est
Truck*Rural	0.579	HybridSUV	-1.453	BEV*Urban	0.013
Minivan*NumAdultsEQ1	-0.797	Hybrid*College	0.226	PHEV*Urban	0.206
Minivan*HasKids	1.248	BEV*College	0.018	BEV*Rural	-0.620
SUV*HasKids	0.625	PHEV*College	0.465	PHEV*Rural	0.018
		BEV*NumAdultsEq1	-0.454		

See Column 1 of Table 4 includes additional demographic interaction effects, including those that are important for a model that is limited exclusively to new vehicle purchase. The first effect (Truck*Rural) makes intuitive sense, in that Trucks would generally have greater utility in rural areas. However, the more general concern is that the model should take into account other household-level effects, including those that could occur due the presence of other vehicles in the household. For example, there are many households that hold multiple vehicles, with a mix of standard cars and light trucks. The number of vehicles is frequently correlated with the number of adults in the household. Table 4 reports a sizeable negative interaction between Minivan and One Adult. Moreover, there are positive interactions for households with kids and Minivans and/or SUVs.

Columns 2 and 3 of Table 4 report interactions related to fuel technology type. There is a negative interaction between Hybrid and SUV, relative to the only other category of Hybrids (passenger cars), i.e., there are no hybrid pickups or minivans. There are positive interactions between College and three alternative fuel technology types (Hybrid, BEV, and PHEV), which is consistent with a narrative related to awareness, knowledge, and preference for new technologies, and possibly attitudes toward environmental friendliness (however, see later discussion). Finally, there is a negative interaction between BEV and being a single-adult household. This captures the expected effect that BEVs are more likely to be purchased in multiple-vehicle households. Column 3 includes interactions between BEVs/PHEVs and location type. BEVs have a strong negative coefficient in rural areas, where at least one possible factor could be the requirement for longer driving ranges and a relative lack of public recharging infrastructure. PHEVs have a sizeable coefficient in urban areas versus rural areas. Although PHEVs don't have a problem with range limitation (and so no negative rural interaction), their relative attractiveness could increase with the increased availability of public charging infrastructure in urban areas.

The final set of coefficients focuses on differences in preference for alternative fuel technologies by region. See Table 5. The baseline fuel technology type is Gasoline, which is assumed to have a coefficient of zero for all regions. Results in Table 5 were ordered from left to right based on the BEV coefficient estimate (least negative on for EastCoast, and most negative for MidWest). Note that there is, of course, much room for statistical variation here, as well as the influence of other household-level interaction effects, the distribution of which will vary from region to region. Also, before proceeding we note for completeness a specific artefact of the current model specification: generic attributes on MPG and HPtoWeight were omitted for all BEVs, so preferences related to fuel efficiency and performance are necessarily absorbed into these coefficients. Having said that, objective measurements on fuel efficiency and performance would likely be higher for BEVs on average, meaning that these estimates are likely to be even more negative if these generic effects are specifically included.

Table 5: Coefficients for Alternative Fuel Technologies by Region

	EastCoast	California	WestNoCal	RockyMtn	GulfCoast	MidWest
BEV	-2.033	-2.134	-2.230	-2.274	-2.843	-3.765
PHEV	-2.873	-1.281	-1.417	-4.036	-2.834	-2.725
Hybrid	-1.329	-0.574	-1.104	-1.568	-1.463	-1.506
Diesel	-0.888	-1.010	-0.261	-0.544	-0.779	-1.058

For four of the six regions (including for California), the most negative coefficients are for BEV. However, in the case of California, even though the BEV coefficient is slightly more negative than for EastCoast, the coefficients for both PHEV and Hybrid in California are substantially smaller than in all other regions (and it also has the second-most-negative coefficient on diesel), meaning that these technologies are necessarily more competitive with BEVs. Conversely, EastCoast coefficients on PHEV and Hybrid track the non-Western regions more closely.

Although it would be possible to develop additional narratives that would be consistent with these findings based on, e.g., regional differences in policy and other factors, at this stage of development our concerns are more methodological. Specifically, these initial results are a starting point for future work that involves (i) expanding the scope of modelling to include data from multiple years that would allow tracking the evolution of these effects *over time*, and (ii) identifying structural mechanisms associated with how these effects might evolve. For example, introducing additional data related to region- or location-specific policies, details on the evolution of recharging infrastructure, etc., can be pursued within this modelling framework. One specific dimension that could contribute to these efforts involves expanding the framework to include the effect of attitudes and perceptions. We conclude by reviewing preliminary results related to this effort.

4.6 Incorporating Consumer Attitudes

The results in the previous section were developed as a logical first step for extending and adapting previously established methodologies to include RP data on plug-in purchases. As described previously, the NVES data also open up possibilities for applying newly methods for incorporating measurements on consumer attitudes. A longer-term goal (and the theoretically preferred approach) is to develop so-called Hybrid Choice Models within a generalized modelling framework that extends the standard RUM approach. See, e.g., [4]. However, in the short term we provide some preliminary results based on directly introducing attitudinal measurements as explanatory variables. (Although less than desirable, researchers have frequently used this approach in the past for practical reasons.)

Section 2 illustrated how averages of importance ratings on Technical Innovation can evolve over time. We use this rating scale, as well as two additional ones (importance of "Environmental Friendliness" and "Fuel Economy") to construct additional segmentation variables to use as explanatory variables. Currently, we define three dummy variable indicators for those consumers that chose a '5' on each importance scale (versus all other choices, including 'Missing'). These dummy variables were interacted with BEV and PHEV, yielding six additional coefficients. The model specification yielding the results in section 4.5 was augmented with these six variables, with no other changes. Coefficient estimates are provided in Table 6.

Table 6: Coefficient Estimates for Six Importance Rating Dummy Variable Interactions

Variable	Coeff Est
BEV*TechInnovImp_High	0.191
PHEV*TechInnovImp_High	-0.292
BEV*EnvFriendlyImp_High	0.424
PHEV*EnvFriendlyImp_High	0.780
BEV*FuelEconImp_High	0.773
PHEV*FuelEconImp_High	0.990

Only one of the six coefficients is negative in this exercise. To begin, it is useful to recall from Figure 1 that the average scores on the importance of Technical Innovation dropped dramatically between 2011 and 2015. (And, these coefficient estimates are based on data from 2015, Q2.) It is important to remember that these represent individual-level consumer effects that are being modelled simultaneously along with many other such effects. Having said this, examining the other coefficients reveals that the vast majority are changed very little from the values reported in section 4.5. The inevitable notable exceptions are the BEV and PHEV coefficients in Table 5. As might be expected, these coefficients are all systematically shifted in a *negative* direction, because the existence of the above coefficients (largely positive) will systematically increase the utility for BEVs and PHEVs for consumers that place a high importance on one or more of these dimensions.

One feature of Table 6 is that both Environmental Friendliness and Fuel Economy importance would seem to have bigger effects on BEV and PHEV utility that Technological Innovation. In terms of context, general public concern about climate change issues were probably increasing during this period. Moreover, a check of the data reveals that, after a run of gasoline price declines from 2014-Q2 through 2015-Q1, average gasoline prices increased between 2015-Q1 and 2015-Q2 by about 42 cents per gallon.

To illustrate the existence of multicollinearity of these measures, we also estimated a version of the model that excluded the variables on Fuel Economy. The resulting estimates are in Table 7. Note that the coefficients for these four coefficients have shifted in positive direction, compensating for the omission of the Fuel Economy variables. (These are in fact the types of effects that can be more appropriately addressed in a more theoretically correct framework.)

Table 7: Coefficient Estimates for Four Importance Rating Dummy Variable Interactions

Variable	Coeff Est
BEV*TechInnovImp_High	0.376
PHEV*TechInnovImp_High	-0.097
BEV*EnvFriendlyImp_High	0.773
PHEV*EnvFriendlyImp_High	1.246

5 Summary and Future Work

The material presented in this paper can be viewed as something of a progress report focusing on initial results from a current research program. On their own, the discrete choice modelling results presented here demonstrate a major first step in extending a relatively demand modelling framework to include actual RP data on plug-in vehicle sales. Although we did not focus on this aspect of the work, the results obtained thus far comprise a useable behavioural model that could be applied to evaluate a variety of alternative market and policy scenarios that would be applicable in the near term, but within limits. The main areas where the work will be extended include expanding the data sets and estimation to cover multiple quarters and years within a single framework. The first immediate benefit would be to greatly improve the identification of parameter estimates, hopefully in two specific ways. First, the ability to estimate purchase price coefficients should be greatly enhanced. Second, it should be possible to estimate more complex structural parameters using nested logit forms. Finally, as discussed in the paper, the models could be expanded to a hybrid choice modelling framework to incorporate rating scale data that capture consumer attitudes.

Acknowledgments

The authors wish to acknowledge the invaluable help and support of Alexander Edwards of Strategic Vision.

References

[1] Kenneth E. Train, *Discrete Choice Methods with Simulation*, Second Edition, Cambridge University Press, 2009.

[2] David S. Bunch and Belinda Chen, *Automobile Demand and Type Choice*, Handbook of Transport I: Transport Modeling (Second Edition), David A. Hensher and Kenneth J. Button, editors, Pergamon (2008), 463-479.

[3] Stephen R. Cosslett, Estimation from Endogenously Stratified Samples, Handbook of Statistics Vol. 11, G. S. Maddala, C. R. Rao and H. D. Vinod, editors, Elsevier Science Publishers B.V. (1993), 1-43.

[4] Joan Walker and Moshe Ben-Akiva, Generalized random utility model, Mathematical Social Sciences, 43 (2002), 303-343.

Authors

David S. Bunch

David is a professor in the Graduate School of Management and the Institute for Transportation Studies at the University of California, Davis. He has consulted on transportation policy issues for state and federal agencies, public utilities and the airline industry. He holds a Ph.D in Mathematical Sciences from Rice University, Texas. The work in this paper was performed while David was a visiting researcher at the King Abdullah Petroleum Science and Research Center (KAPSARC).

Rubal Dua

Rubal is a senior research associate at KAPSARC leading vehicle regulatory policy and shared mobility research using a consumer perspective. He holds a Ph.D from KAUST, KSA, a M.S. degree from the University of Pennsylvania, U.S. and a B.Tech degree from IIT Roorkee, India.