

1 **A Maximum Log-likelihood Based Data Fusion Model for Estimating Household's Vehicle**  
2 **Purchase Decision**  
3  
4

**Md Istiak Jahan\***

Graduate Student  
Department of Civil, Environmental &  
Construction Engineering  
University of Central Florida  
Email: md.istiak.jahan@ucf.edu  
ORCID number: 0000-0002-4056-7816

**Sachraa G. Borjigin**

Maryland Transportation Institute  
Department of Civil and Environmental  
Engineering  
University of Maryland, College Park  
1173 Glenn L. Martin Hall  
4298 Campus Dr, College Park, MD 20742  
College Park, MD 20742  
Email: sachraa@outlook.com

**Nneoma M. Ugwu**

Graduate Student  
Department of Civil and Environmental  
Engineering  
University of Maryland, College Park  
1173 Glenn L. Martin Hall  
4298 Campus Dr, College Park, MD 20742  
Email: nugwu@umd.edu

**Tanmoy Bhowmik**

Assistant Professor  
Department of Civil, and Environmental  
Engineering  
Portland State University  
Tel: 407-927-6574  
Email: tbhowmik@pdx.edu  
ORCID number: 0000-0002-0258-1692

**Jiehong Lou**

Assistant Research Professor  
Center for Global Sustainability  
School of Public Policy  
University of Maryland, College Park  
7805 Regents Dr.  
College Park, MD 20742  
Email: jlou@umd.edu

**Deb A. Niemeier**

Clark Distinguished Chair Professor  
Director, Center for Disaster Resilience  
Maryland Transportation Institute  
Department of Civil and Environmental  
Engineering  
University of Maryland, College Park  
1173 Glenn L. Martin Hall  
4298 Campus Dr, College Park, MD 20742  
College Park, MD 20742  
Email: niemeier@umd.edu

**Naveen Eluru**

Professor  
Department of Civil, Environmental & Construction Engineering  
University of Central Florida  
Tel: 407-823-4815  
Email: naveen.eluru@ucf.edu  
ORCID number: 0000-0003-1221-4113

5  
6  
7  
8

---

\* Corresponding Author

1 **Abstract**

2 The growing adoption of electric vehicles offers a potential opportunity to reduce transportation  
3 sector carbon footprint. In our research, we studied vehicle purchase behavior with emphasis on  
4 alternative fuel vehicles using the vehicle purchase dataset “MaritzCX New Vehicle Customer  
5 Study”. This study consisted of a two-level modeling approach. In the first level, purchasing of a  
6 new car was estimated based on consumers socio-economic characteristics. In the second level,  
7 the vehicle purchase decision was examined with a two-dimensional dependent variable – vehicle  
8 type and fuel type. We employed an innovative data fusion approach that probabilistically links  
9 records from MaritzCX with records from National Household Travel Survey with the objective  
10 of identifying new independent variables affecting the decision process while maximizing data fit.  
11 The final model included a host of independent variables from four different categories: vehicle-,  
12 economic-, demographic-, and spatial characteristics. Finally, the model results were employed to  
13 conduct an elasticity analysis.

14

15 **Keywords:** Vehicle purchase decision, fuel type, vehicle type, fusion model, MaritzCX data

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

## 1 BACKGROUND

2 In the United States, the over-reliance on personal automobile for mobility needs is well  
3 recognized. The number of households reported to own at least one vehicle has increased from  
4 86.5% in 1983 to 91.1% in 2017 (Mcguckin et al., 2018). The over-reliance on automobiles has  
5 economic and environmental implications. On the economic front, over reliance contributes to  
6 increase in household expenditure in response to gas prices, congestion related losses (time and  
7 money), suburban sprawl and increased infrastructure construction and maintenance costs (Okeke  
8 et al., 2020; Kanyepe et al. , 2021). On the environmental front, increased reliance on automobile  
9 alternative results in significant noise and air pollution contributing to increased carbon emissions  
10 and worsening public health (Lowe et al., 2022; Morency et al., 2020). In fact, a typical passenger  
11 vehicle driven for 11,500 miles at a fuel efficiency of 22 miles per gallon is expected to emit 4.6  
12 metric tons of CO<sub>2</sub> annually (USEPA, 2022). The growing adoption of alternative fuel vehicles  
13 (AFVs) including Hybrid vehicles, electric vehicles (EVs) and plug-in hybrid electric vehicles  
14 (PHEVs) offers a mechanism for reducing the impact of automobile dependency on the  
15 environment. The carbon emissions from these alternative fuel vehicles are reported to be 17-30%  
16 lower than the emissions from conventional gasoline or diesel vehicles (Blink, 2023).

17 In the US, the sales of various AFVs have increased over time. For instance, light duty  
18 plug-in EVs in 2021 were 608,000; about twice the sales in 2020 (Minos, 2022). The growing  
19 adoption of these alternative fuel vehicles across the country raises an important question for  
20 transportation planners, energy companies and vehicle manufacturers – *What are factors that*  
21 *encourage or dissuade the purchase of AFVs?* The objective of the current research is to study  
22 vehicle purchase behavior with emphasis on AFVs using a scarcely used vehicle purchase dataset  
23 MaritzCX New Vehicle Customer Study (NVCS). MaritzCX data is a long-serving vehicle  
24 purchase behavior consumer survey that provides details of the various vehicles purchased or  
25 leased across US and Canada. The dataset is the most comprehensive survey of vehicle purchases  
26 with about 12.5 million survey responses in the last 50 years. The data is employed by vehicle  
27 manufacturers to draw on insights on customer behavior including evolving trends, changes to  
28 behavior and brand value (MaritzCX, 2019). While the MaritzCX data provides the most  
29 comprehensive vehicle purchase data, it is not employed in transportation modeling to understand  
30 population vehicle purchase behavior for two reasons. First, MaritzCX records represent a self-  
31 selected sample of vehicle purchasers (and not the general population). Second, MaritzCX data  
32 provides very limited data on the household characteristics of the vehicle purchases. Hence, the  
33 rich set of vehicle purchase observations are not useful in understanding vehicle fleet evolution at  
34 the household resolution.

35 In the current research effort, we propose solutions to address these limitations and develop  
36 a framework that will allow us to tap into the rich information available in MartizCX data. The  
37 *self-selection of vehicle purchasers* is addressed by developing a vehicle purchase decision model  
38 using data from the 2017 National Household Travel Survey (NHTS, 2017). From the NHTS data  
39 we create a binary dependent variable – new vehicle purchase decision based on vehicle acquisition  
40 year reported in the survey. The households that acquired a new vehicle are assigned a value of  
41 one and rest of the households are assigned a value of zero. The dependent variable created is  
42 analyzed to identify the factors influencing the decision to purchase a vehicle. The binary model  
43 thus allows us to create the sample of households that are in the market for purchasing vehicles –  
44 the sample represented in MaritzCX data.

45 To improve the *independent variables available* in the MaritzCX data, we employ an  
46 innovative data fusion approach that probabilistically links records from MartizCX with records

1 from NHTS. The fusion process, developed recently by the study (Bhowmik et al., 2024a;  
2 Bhowmik et al., 2024b), is guided by a maximum likelihood approach that allows for fusing two  
3 datasets without any common identifier (mathematical details are provided in Methodology  
4 section) with the objective of improving the model fit of the choice variable of interest. In our  
5 research, the choice variable of interest is the combination of vehicle type (defined as Utility  
6 vehicles, Sport sedans, Sedans, and Pickup trucks) and fuel type (defined as Gasoline and non-  
7 Gasoline<sup>1</sup>). The fused dataset is employed to analyze vehicle purchase decisions with eight  
8 alternatives using a multinomial logit model. The model estimation process clearly illustrates how  
9 the model developed with fused records (i.e., considering the newly added independent variables  
10 from NHTS) offer significant improvement in data fit. The model exercise is supplemented with  
11 elasticity analysis to illustrate the applicability of the proposed model.

12 The remainder of the paper is organized as follows: The earlier literature section provides  
13 a brief overview of vehicle ownership models. The next section outlines the methodological details  
14 while the Data Section summarizes the datasets. The Model results section summarizes the results  
15 of the models estimated. The applicability of the proposed framework is illustrated in the elasticity  
16 analysis section. The final section concludes the paper.

## 17 2 EARLIER WORK AND STUDY CONTRIBUTIONS

### 18 2.1 Earlier Research

19  
20 Vehicle ownership models have received significant attention in transportation literature (see  
21 (Anowar et al., 2014; Ma and Ye, 2019) for review efforts). In our study, we provide a brief review  
22 of relevant literature. In vehicle ownership studies, the dimensions of interest include: (a) vehicle  
23 type choice such as sedan, and sports utility vehicle (SUV) (Spissu et al., 2009; Baltas and  
24 Saridakis, 2013; Mabit, 2014; Gillingham et al., 2015; Bubeck et al., 2016; Cirillo et al., 2016;  
25 Jian et al., 2017; Nazari et al., 2019), (b) comparative evaluation of various fuel types considering  
26 costs (Ahn et al., 2008; Bolduc et al., 2008; Fang, 2008; Bhat et al., 2009; Sheldon and Dua, 2018,  
27 2021; Guo et al., 2019; Ryu et al., 2020), and (c) adoption of alternative fuel vehicles including  
28 EVs (Qian and Soopramanien, 2011; Eppstein et al., 2011; Hoen and Koetse, 2014; Chen et al.,  
29 2015; Bubeck et al., 2016; Hackbarth and Madlener, 2016; Daina et al., 2017; Mulholland et al.,  
30 2018; Liao et al., 2019; Kumar and Chakrabarty, 2020; Lee and Brown, 2021; Caggiani, Prencipe  
31 and Ottomanelli, 2021; Fevang et al., 2021; Ackaah et al., 2022; Ayetor et al., 2023). The data  
32 sources employed for analysis vary by the vehicle ownership model resolution. For disaggregate  
33 analysis at the household level revealed preference data such as National Household Travel  
34 Survey, urban region-specific data (such as origin destination data for Montreal) and MaritzCX  
35 data are employed (Anowar et al., 2016). Several studies, especially for understanding the  
36 preferences for emerging vehicle types have employed stated preference surveys (Cirillo et al.,  
37 2016; Menon et al., 2019). The spectrum of modeling approaches employed for vehicle ownership  
38 include multinomial logit model (Zhao et al., 2018; Fevang et al., 2021; Sabouri et al., 2021; Kiran  
39 and Shanmugam, 2017; Hara and Asahi, 2020), nested logit model (Stinson et al., 2020; Zhou et  
40 al., 2020), latent class logit model (Khan & Habib, 2021), ordered-logit model (Sabouri et al.,  
41 2021), Poisson regression model (Sabouri et al., 2021), and multiple discrete-continuous extreme  
42

---

<sup>1</sup> The share of EVs in the 2017-18 MaritzCX data sample are small. Hence, we employed a non-gasoline characterization of our dependent variable. While the non-gasoline characterization includes multiple fuel types, we can employ the framework developed here to consider that non-gasoline market will primarily evolve into an EV market.

1 value model (Ahn et al., 2008; Bhat et al., 2009; Jian et al., 2017). These studies employed a  
 2 diverse range of exogenous variables including – household’s demographics, socio-economic  
 3 attributes, commuting pattern characteristics, transportation network characteristics, land use  
 4 characteristics, built environment characteristics, vehicle characteristics, and policy related  
 5 attributes. Traditional vehicle choice studies that employed vehicle sales data usually limit  
 6 themselves to vehicle characteristics (such as vehicle price, internal area, allowed load, vehicle  
 7 age, engine power, cylinder volume, wheel type, fuel type, body type and vehicle brand) (see  
 8 (Anderhofstadt & Spinler, 2019; Dhanabalan et al., 2018; Harahap et al., 2019; Joshi & Bhatt,  
 9 2018; Kim & Kim, 2014; Knez et al., 2014; Ma & Mayburov, 2021; Østli et al., 2017; Raza &  
 10 Masmoudi, 2020)). On the other hand, vehicle choice studies that employed survey datasets (such  
 11 as California Household Travel Survey data and datasets from SMS or email invitations) are found  
 12 to emphasize several household socioeconomic characteristics (such as age, gender, household  
 13 income, ethnicity, number of household vehicles, employment, educational qualification, and  
 14 annual mileage) (see (Arokiaraj & M.Banumathi, 2014; Bauer, 2018; Biswas et al., 2014; Chang  
 15 & Hsiao, 2011; Eluru et al., 2010; Liu et al., 2016; Nerurkar et al., 2023; Pierce & Connolly, 2023;  
 16 van Huyssteen & Rudansky-Kloppers, 2024)). These studies are found to employ a very limited  
 17 number of vehicle characteristics (such as fuel type, body type, vehicle brand and vehicle age).  
 18 The reasoning for these data limitation is two-fold. *First*, vehicle characteristics and array of  
 19 sociodemographic information are not readily available in a single data source. *Second*, for  
 20 estimating vehicle ownership models employing different categories of exogenous variables, it  
 21 might be necessary to merge several datasets. The merging process is straight forward in datasets  
 22 with a clear matching identifier variable. However, the vehicle sales datasets and survey datasets  
 23 do not have any common identifier variable that can be utilized for merging those datasets. To the  
 24 authors’ knowledge, no research has undertaken a merging exercise for two datasets without a  
 25 matched variable for developing vehicle ownership models. In this study we estimated households’  
 26 vehicle purchase decision, as a combination of vehicle type and fuel type, by integrating an  
 27 extensive array of vehicle characteristics (from MaritzCX vehicle purchase data) alongside  
 28 household’s economic and demographic characteristics (from 2017 NHTS data) through an  
 29 innovative fusion of merging two distinct datasets without a matched identifier variable.  
 30

## 31 **2.2 Current Study in Context**

32 The current study contributes to vehicle choice and purchase literature with the following  
 33 three objectives. *First*, the study utilizes a large-scale vehicle purchase dataset scarcely employed  
 34 in transportation literature. To be sure, multiple studies employed MaritzCX data to study vehicle  
 35 ownership dimensions such as the role of fuel economy (Leard et al., 2020; Ankney and Leard,  
 36 2021), adoption of EVs (Holland et al., 2016, 2019; Jabbari et al., 2017; Xing et al., 2021), and  
 37 purchasing new or used passenger cars (Leard, 2022). However, the models developed with  
 38 MaritzCX data do not consider several important household variables influencing vehicle purchase  
 39 decisions such as household age distribution, gender distribution, current vehicle ownership,  
 40 vehicle availability, number of adults and children in the household, employment characteristics,  
 41 and educational attainment of the household members. The current study is motivated by the need  
 42 to improve our understanding of vehicle purchase decisions by incorporating additional  
 43 independent variables drawn from NHTS and incorporated along with the rich details of vehicle  
 44 fuel and type choice from MaritzCX data. *Second*, the dependent variable of interest in our analysis  
 45 is combination of fuel (gasoline and non-gasoline) and vehicle type (utility vehicle, sport sedan,  
 46 sedan, and pickup truck) alternatives. Fuel type is an important aspect to consider while purchasing

1 vehicle, because of its effect not only on the maintenance and driving cost, but also on the  
2 environmental footprint of the vehicle. On the other hand, vehicle type represents consumers'  
3 lifestyle and commuting requirements. Therefore, considering both fuel and vehicle type allows  
4 us to form a more inclusive decision. *Finally*, a novel data fusion approach has been utilized for  
5 merging two datasets with no matched identifier variable. The proposed data fusion approach is  
6 geared towards identifying new independent variables affecting the dependent variable while  
7 maximizing the fit for the dependent variable of interest (see Figure 1). The reader will note that  
8 the records in NHTS and MaritzCX do not correspond to the same individuals or households. The  
9 proposed data fusion is based on common attributes across the two datasets. Four identical  
10 variables – age category, income category, household's state, and household's location classified  
11 as urban or rural – are present in both MaritzCX and NHTS dataset. The fusion algorithm  
12 hypothesizes that households with matching attributes are likely to have an increased likelihood  
13 for making similar choices. The approach tests two weight mechanisms to recognize that multiple  
14 matches from NHTS dataset should contribute to only one record from MaritzCX (see (Bhowmik  
15 et al., 2024a; Bhowmik et al., 2024b) for examples of similar methods in different contexts). The  
16 data fusion process and model selection procedures are summarized in Figure 1 and 2.

17 In summary, the current study is conducted in two parts. In the first part, a binary logit  
18 model was estimated using the 2017 NHTS data to understand the impact of different household  
19 economic and demographic characteristics on household decision to purchase a new vehicle<sup>2</sup>. This  
20 model allows us to identify the subsample of NHTS data that represent the households purchasing  
21 a new vehicle. In the second part, the NHTS records of this subsample were fused with the  
22 MaritzCX data employing the proposed fusion approach.

---

<sup>2</sup> The reader would note that the NHTS data already identifies the households who purchased new vehicle(s). Thus, we do not strictly need the binary logit component for our analysis. However, when the model is applied to other datasets without an identifier for new purchases, having a purchase decision model will be useful as a two-level analysis. Hence, we included this in our paper.

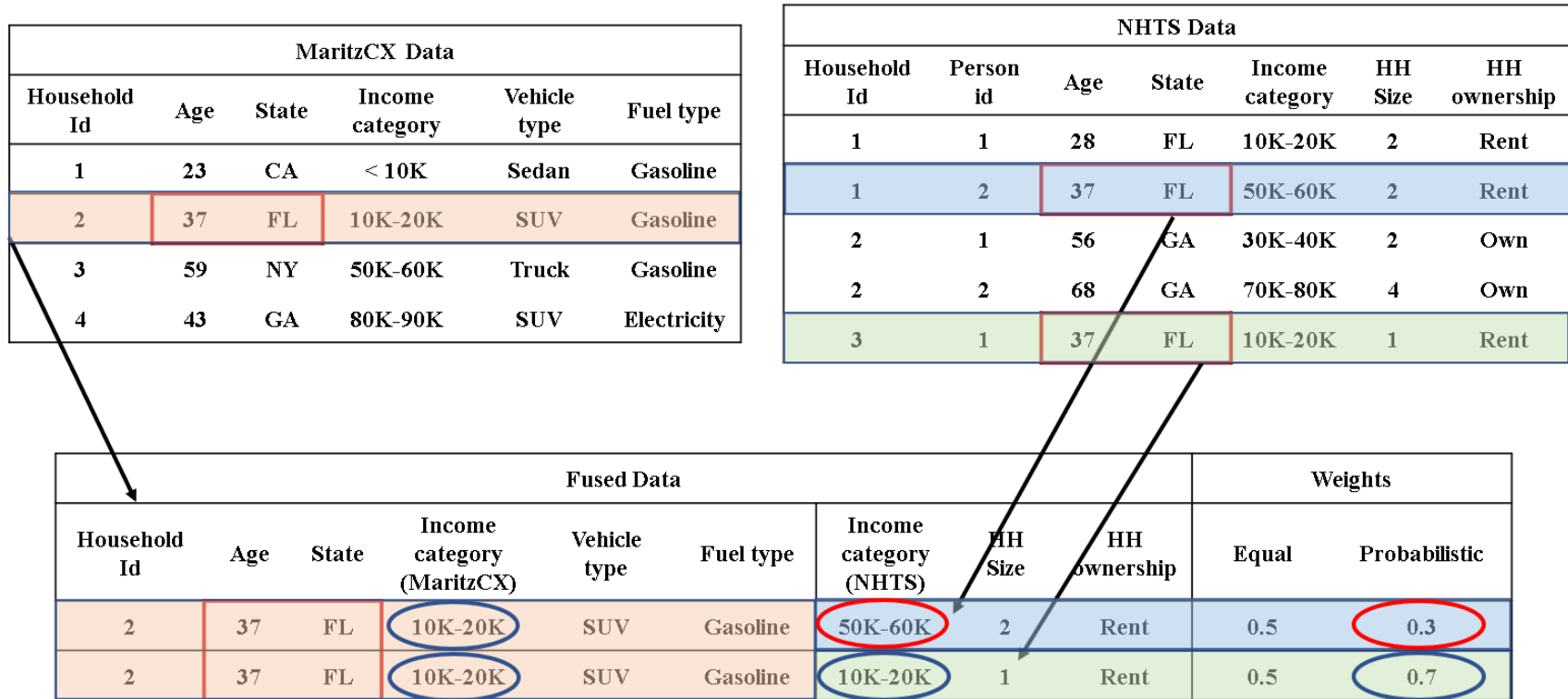


Figure 1: MaritzCX and NHTS Data Fusion Illustration

1  
2  
3

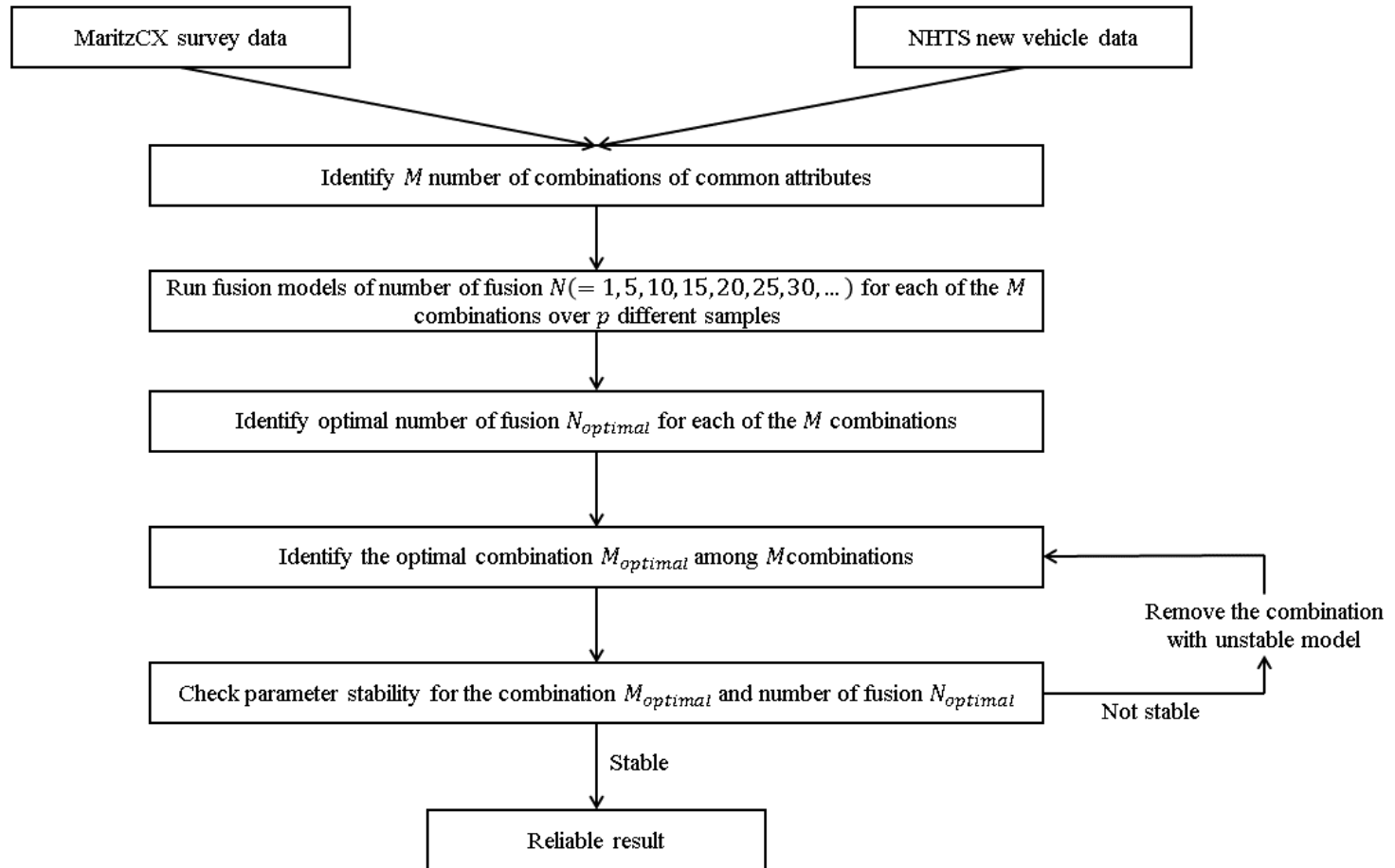


Figure 2: Flow Chart of Fusion Algorithm

1  
2



1 **3 METHODOLOGY**

2 In this section, we present the methodological framework adopted in the study for analyzing  
 3 household’s vehicle purchase decision. The model structure consists of two models: 1) a binary  
 4 logit (BL) model to identify the households buying a new car, and 2) a data fusion driven model  
 5 to analyze individual’s vehicle choice by fuel types and vehicle types. The methodologies of the  
 6 proposed frameworks are described below.

7  
 8 **3.1 Binary Logit Model**

9 Let us assume  $i(i = 1,2,3, \dots, N)$  as the index for household in the NHTS data. Now for the binary  
 10 logit framework, the probability expression is as follows:

11  
 12 
$$\Lambda[y_i] = \begin{cases} \pi_i & y_i = 1 \\ 1 - \pi_i & y_i = 0 \end{cases} \dots \dots \dots (1)$$

13 where  $\Lambda[y_{ij}]$  represents the probability that household  $i$  will buy a new car or not (yes/no) and it  
 14 will be determined based on  $\pi_i$ . Here  $y_i = 1$ , if household  $i$  buys a new car, and 0, otherwise. With  
 15 this notation, the equation for  $\pi_i$  is as follows:

16  
 17  
 18 
$$\pi_i = \frac{\exp(\gamma\eta_i)}{1 + \exp(\gamma\eta_i)} \dots \dots \dots (2)$$

19 where,  $\eta_i$  is a vector of attributes (including the constant) and  $\gamma$  is a conformable parameter vector  
 20 of the binary logit model to be estimated. This  $\pi_i$  will help us determining the household that will  
 21 be fused with MaritzCX data.

22  
 23  
 24 **3.2 Fusion Model**

25 The fusion model structure has a decision component and a weight component. In the decision  
 26 model component, a multinomial logit formulation for the vehicle purchase variable is considered  
 27 as we have eight alternatives in the choice dimension. Let’s, assume  $j$  and  $k$  are two index  
 28 representing fuel type and vehicle type respectively. In our study, the values of the  $j$  were assigned  
 29 as follows: gasoline ( $j = 1$ ), non-gasoline ( $j = 2$ ); and the values of the  $k$  were assigned as  
 30 follows: utility vehicle ( $k = 1$ ), sport sedan ( $k = 2$ ), sedan ( $k = 3$ ), and pickup trucks ( $k = 4$ ).  
 31 There are  $n (= 1,2,3, \dots N)$  individuals in the MaritzCX data and  $m$  possible matches from the  
 32 NHTS dataset. With this notation, the vehicle purchase propensity takes the following form:

33  
 34 
$$v_{nm,jk}^* = \alpha X_{n,jk} + \beta' S_{nm,jk} + \varepsilon_{nm,jk}, v_{nm,jk} = 1 \text{ if } v_{nm,jk}^* > 0; v_{nm,jk} = 0 \text{ otherwise } \dots (3)$$

35 where  $v_{nm,jk}^*$  represents the propensity of the individual  $n$  for the  $m^{th}$  fused record to purchase  
 36 fuel type  $j$  and vehicle type  $k$ .  $v_{nm,jk}$  is the observed vehicle choice, that is 1 if the person  $n$  in  
 37 MaritzCX for the  $m^{th}$  fused record from NHTS had purchased fuel type  $j$  and vehicle type  $k$ ; and  
 38 0 otherwise.  $X_{n,jk}$  is a vector of attributes from the MaritzCX dataset that influence the vehicle  
 39 purchase choice and  $\alpha$  is the corresponding coefficients to be estimated (including a scalar  
 40 constant).  $S_{nm,jk}$  is the vector of attributes from the NHTS dataset that affect the purchase decision  
 41 and  $\beta'$  is the corresponding vector of coefficients to be estimated.  $\varepsilon_{nm,jk}$  is an idiosyncratic error  
 42

1 term assumed to be identically and independently Type I Gumbel distributed. Based on this, the  
 2 probability for person  $n$  for the  $m^{th}$  fused records to purchase fuel type  $j$  and vehicle type  $k$  given  
 3 by:  
 4

$$5 \quad P_{nm,jk} = \frac{\exp(v_{nm,jk}^*)}{\sum_{k=1}^K \sum_{j=1}^J \exp(v_{nm,jk}^*)} \dots \dots \dots (4)$$

6  
 7 The weight component also takes the form of a latent multinomial logit structure allocating  
 8 the probability for each NHTS user being paired with the MaritzCX individual. The matched  
 9 weightage propensity is determined based on a latent probability value estimated using a  
 10 multinomial logit model as follows:  
 11

$$12 \quad w_{nm} = \frac{\exp(\Psi Z_{nm})}{\sum_{m=1}^M \exp(\Psi Z_{nm})} \dots \dots \dots (5)$$

13  
 14 where  $Z_{nm}$  is a column vector of attributes for individual  $n$  and fused record  $m$  that influences the  
 15 propensity of matching the NHTS data with the MaritzCX data.  $Z_{nm}$  represents the variables that  
 16 are present in both datasets but not used for fusion.  $\Psi$  is the corresponding vector coefficients to  
 17 be estimated.  
 18

19 **3.3 Model Estimation**

20 Based on above notations, the overall weighted probability for each individual in the MaritzCX  
 21 data can be written as:  
 22

$$23 \quad L_n = \sum_{m=1}^M w_{nm} * P_{nm,jk} \dots \dots \dots (6)$$

24  
 25 where  $L_n$  is the weighted probability the person  $n$  in the MaritzCX dataset has the corresponding  
 26 vehicle choice  $P_{jk}$ . Employing equation 6, several additional variables from the NHTS dataset will  
 27 be employed to develop an improved vehicle purchase choice model. Finally, the log-likelihood  
 28 function for the fused dataset is defined as:  
 29

$$30 \quad LL = \sum_{n=1}^N \log(L_n) \dots \dots \dots (7)$$

31  
 32 The proposed matching algorithm has been estimated using a maximum likelihood based  
 33 econometric model. We have used the GAUSS Matrix Programming software for estimating the  
 34 models.  
 35

36 **4 DATA DESCRIPTION**

37 The MaritzCX vehicle purchase data and the National Household Travel Survey (NHTS) 2017  
 38 data were used in our analysis. MaritzCX data, providing the household vehicle ownership records  
 39 from the year 2010 to 2018 of 1,813,674 households across the entire United States, was used for

1 vehicle purchase related information. It contains information about vehicle purchase year, vehicle  
2 model year, fuel type, body type, vehicle make, model, cost, and engine characteristics. In addition,  
3 it provides information about consumers' age, income, and housing locations (states and  
4 urban/rural). The MaritzCX data encompasses both household level (such as household income,  
5 location, and state) and person level (such as age of the customer who bought the car) attributes  
6 representing only one member from each household. On the other hand, the NHTS 2017 survey  
7 data was utilized in this study for detailed socio-economic and spatial characteristics including,  
8 household size, number of working members, number of adults, household vehicle availability,  
9 and educational attainment of the household members. This dataset contains the records of 118,100  
10 households representing the characteristics of 103,091,506 households across the entire United  
11 States.

#### 12 13 **4.1 Dependent Variable**

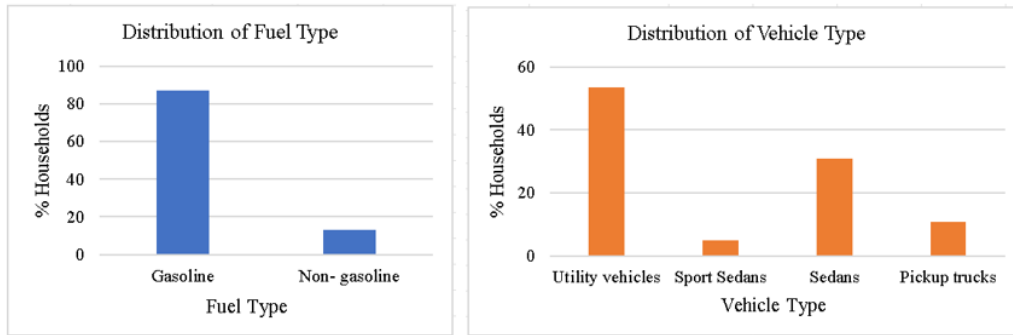
14 This study consisted of a two-level modeling approach. In the first level, purchasing of a new car  
15 was estimated based on consumers current socio-economic characteristics. For this model,  
16 households that acquired a new vehicle are assigned a value of one and the rest of the households  
17 are assigned a value of zero. In our analysis, vehicles of model year 2016 and 2017, and acquired  
18 in 2017 in NHTS 2017 data were considered as new vehicles. To ensure that model overfitting is  
19 not an issue, 25,000 out of 118,100 household records were randomly chosen for modeling  
20 purposes. The share of new vehicles in the estimation dataset was found to be 12.3%.

21 In the second level, the vehicle purchase decision was estimated for a two-dimensional  
22 dependent variable – vehicle type and fuel type. In our study, the variable of interest is the  
23 combination of vehicle types and fuel types, and the dependent variable is categorized into 8  
24 categories (4 vehicle types x 2 fuel types). However, presence of alternatives with a very small  
25 share can affect the estimation of parameters for the smaller share alternatives. To circumvent this,  
26 we employ a dimension specific estimation approach for independent variable impacts. To  
27 elaborate, we estimate the impact of independent variables for vehicle type and fuel type  
28 separately. Then, after estimation, the impact of any alternative can be computed as the sum of the  
29 parameters for each dimension. This approach, sometimes referred to as portfolio modeling, is  
30 commonly employed for joint choice modeling in transportation (such as mode and departure time  
31 choice, mode and activity type choice; see (Anowar et al., 2015; Imani et al., 2014) for examples  
32 of such approaches). Since the estimation was conducted for only the new vehicles, vehicles of  
33 model year 2016 and 2017, and bought in 2017 in MaritzCX data were considered for this analysis.  
34 Based on this criterion, among 1,813,674 customer records 187,092 customers were found to buy  
35 a new car. After eliminating the cases with missing values, 103,385 customer records were retained  
36 in the study dataset. For the data fusion and modeling purpose 5,000 customers with a new car  
37 were randomly selected from the MaritzCX data<sup>3</sup>. Further, the distributions of various fuel types  
38 and vehicle types are shown in Figure 3. It is found that gasoline operated vehicles hold the highest

---

<sup>3</sup> The reader would note that full datasets were not used in our analysis for two reasons. *First*, estimating models with large samples can potentially inflate the t-statistics resulting in too many variables being significant leading to overfitting. So, using sampled datasets allows us to arrive at specifications that are less prone to overfitting. *Second*, the computational burden of running models with full datasets is significant in our research exercise. As the reader is aware, we run models after fusing different numbers of records – 1, 5, 10, 15, 20, 25, 30 and 35. So, when we estimate a model with a 5000-sample size for MaritzCX data with 15 fusion records from NHTS dataset we are estimating models with 5000\*15 records. Given the number of comparisons we do across variable combinations and fusion sizes, the sampling approach offers a reasonable compromise.

1 share of 86.76% in fuel type distribution, while utility vehicles hold the higher share of 53.40% in  
2 vehicle type distribution. The description of the dependent variables is presented in Table 1.  
3



4  
5 **Figure 3: Distribution of Dependent Variables**  
6

#### 7 **4.2 Independent Variable**

8 A host of independent variables from four different categories: 1) vehicle characteristics and 2)  
9 household economic characteristics, 3) household demographic characteristics, and 4) spatial  
10 characteristics are utilized in this study. The vehicle characteristics data were drawn from  
11 MaritzCX data. All other variables are drawn from NHTS data and considered through the fusion  
12 process. In our study, the fusion model is developed by fusing the NHTS data to the MaritzCX  
13 data. Among 118,100 households in the NHTS data, 14,274 records are selected for fusion purpose  
14 based on the new vehicle criterion. Considering the data resolution, a dataset of 28,995 household  
15 members from the 14,274 households in NHTS data is employed to fuse with the MaritzCX data.  
16 The resampling procedure of both databases is presented in Figure 4. The descriptive statistics of  
17 the independent variables, used in the model specification, were presented in Table 1. The common  
18 variables across MaritzCX and NHTS data were identified in Figure 5 (variables with results from  
19 both datasets). In the model estimation process several functional forms of the variables were  
20 tested and the final specification was based on statistical significance at the 90% level.

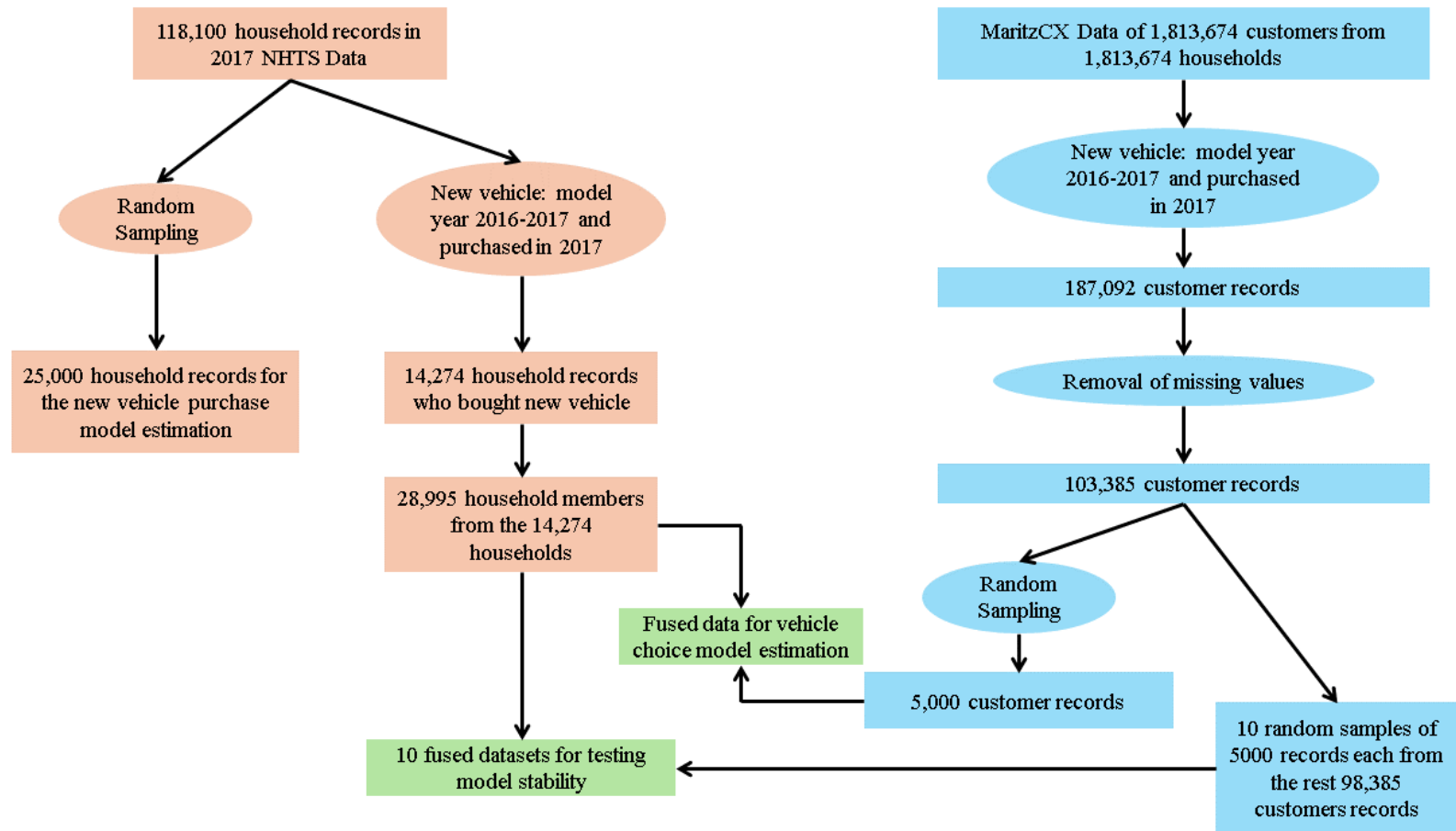
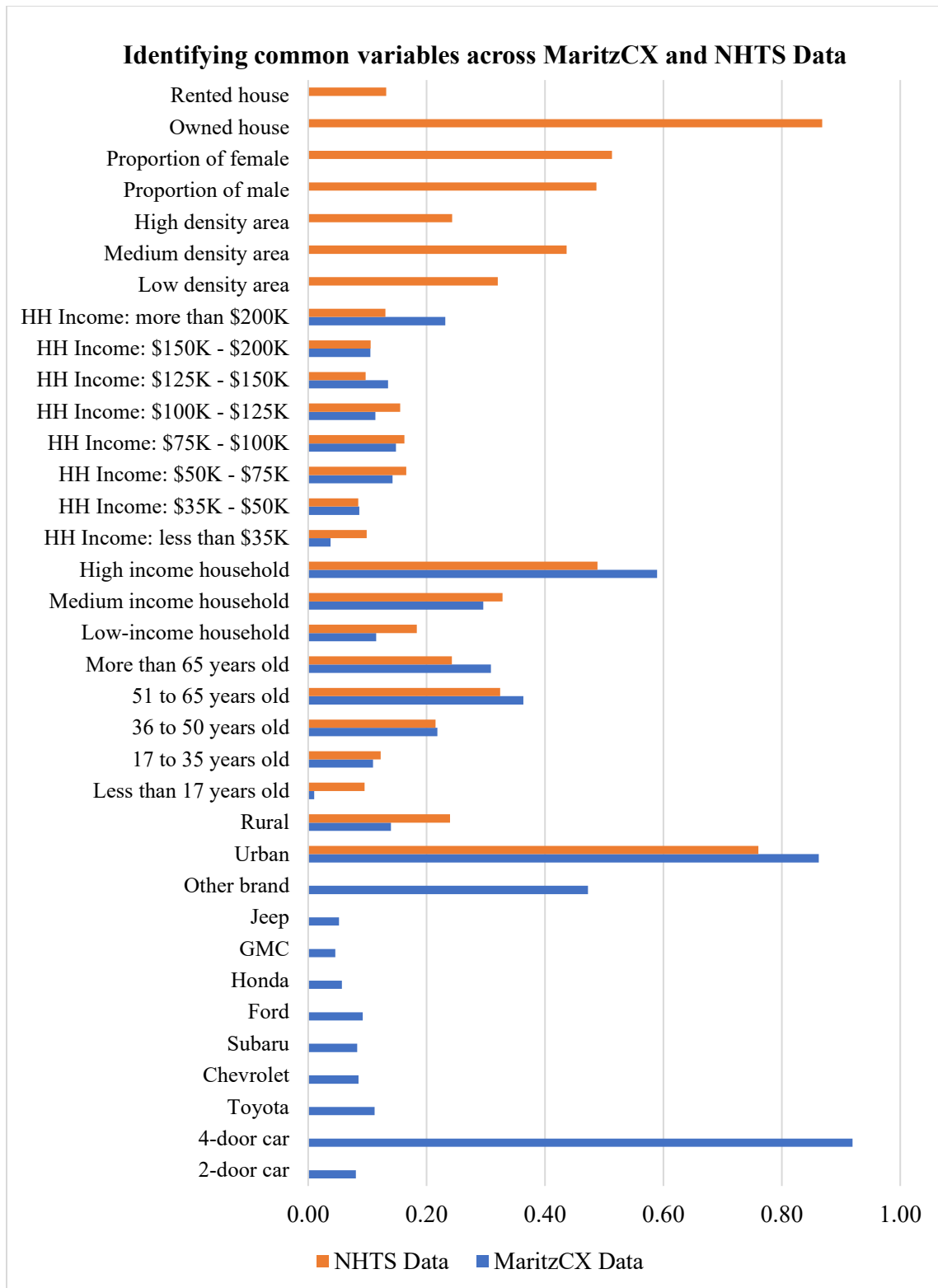


Figure 4: Resampling procedure of the MaritzCX and NHTS datasets

1  
2  
3



**Figure 5: Mean of variables in MaritzCX and NHTS data**

1  
2  
3  
4  
5

1 **Table 1: Descriptive Statistics of the Model Variables**

Variable name	Description	Minimum	Maximum	Mean
<i>MaritzCX Data (N = 5,000)</i>				
Fuel efficiency (mpg)	Mileage driven by per gallon gas (equivalent to 33.7 kWh of electricity)	13.35	77.20	25.79
Vehicle price (in thousands)	Vehicle price in USD	16.88	81.14	35.74
Engine size (liters)	The volume of fuel and air that can be pushed through a car's cylinders	1.40	6.40	2.87
2-door car	Number of doors: 2	0.00	1.00	0.08
4-door car	Number of doors: 4	0.00	1.00	0.92
Chevrolet	Car brand: Chevrolet	0.00	1.00	0.09
Ford	Car brand: Ford	0.00	1.00	0.09
Honda	Car brand: Honda	0.00	1.00	0.06
Toyota	Car brand: Toyota	0.00	1.00	0.11
Subaru	Car brand: Subaru	0.00	1.00	0.08
Jeep	Car brand: Jeep	0.00	1.00	0.05
GMC	Car brand: GMC	0.00	1.00	0.05
Other brands	Other than the above-mentioned brands	0.00	1.00	0.47
Low-income household	Household income less than 50K	0.00	1.00	0.12
Medium income household	Household income in between 50K and 100K	0.00	1.00	0.30
High income household	Household income more than 100K	0.00	1.00	0.58
Less than 17 years old	Age category: less than 17 years	0.00	1.00	0.01
17 to 35 years old	Age category: less than 17 to 35 years	0.00	1.00	0.11
36 to 50 years old	Age category: less than 36 to 50 years	0.00	1.00	0.22
51 to 65 years old	Age category: less than 51 to 65 years	0.00	1.00	0.36
More than 65 years old	Age category: more than 65 years	0.00	1.00	0.31
Location: Urban	Urban = 1/ Rural = 0	0.00	1.00	0.86
Location: Rural	Urban = 0/ Rural = 1	0.00	1.00	0.14
Northeast	Included states: Connecticut, Massachusetts, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont	0.00	1.00	0.20
Midwest	Included states: Iowa, Illinois, Indiana, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota, and Wisconsin	0.00	1.00	0.24
South	Included states: Alabama, Arkansas, District of Columbia, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Oklahoma, South Carolina, North Carolina, Tennessee, Texas, Virginia, and West Virginia	0.00	1.00	0.34

West	Included states: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, New Mexico, Nevada, Oregon, Utah, Washington, and Wyoming	0.00	1.00	0.22
<b><i>NHTS Data (N = 25,000; household level data for vehicle purchase model)</i></b>				
Low density area	Population density: 0-499 persons per square mile	0.00	1.00	0.32
Medium density area	Population density: 500-3999 persons per square mile	0.00	1.00	0.42
High density area	Population density: 4000 or more persons per square mile	0.00	1.00	0.26
Household size	Count of household members	1.00	11.00	2.17
Number of workers	Count of household workers	0.00	7.00	1.04
Number of adults	Count of household members of at least 18 years old	1.00	8.00	1.81
Number of children	Count of household members of more than 18 years old	0.00	8.00	0.36
Owned house	Household ownership: Own a house	0.00	1.00	0.79
Rented house	Household ownership: Do not own house	0.00	1.00	0.21
Proportion of male	Count of male members/ Household size	0.00	1.00	0.45
Proportion of female	Count of female members/ Household size	0.00	1.00	0.55
Proportion of Caucasian-American	Count of Caucasian-American members/ Household size	0.00	1.00	0.84
Proportion of African American	Count of African American members/ Household size	0.00	1.00	0.07
Proportion of Asian American	Count of Asian-American members/ Household size	0.00	1.00	0.04
Proportion of other races	Count of other races/ Household size	0.00	1.00	0.05
Proportion of people aged less than 17 years old	Count of members aged less than 17 years/ Household size	0.00	1.00	0.01
Proportion of people aged 17 to 35 years	Count of members aged 17 to 35 years/ Household size	0.00	1.00	0.18
Proportion of people aged 35 to 50 years	Count of members aged 35 to 50 years/ Household size	0.00	1.00	0.19
Proportion of people aged 51 to 65 years	Count of members aged 51 to 65 years/ Household size	0.00	1.00	0.32
Proportion of people aged more than 65 years old	Count of members aged more than 65 years/ Household size	0.00	1.00	0.31
Urban area	Household located in urban area	0.00	1.00	0.77
Rural area	Household located in rural area	0.00	1.00	0.23
<b><i>NHTS Data (N = 28,995; person level data for fusion model- considering new vehicles only)</i></b>				
Owned house	Household ownership: Own a house	0.00	1.00	0.87



Rented house	Household ownership: Do not own house	0.00	1.00	0.13
Low density area	Population density: 0-499 persons per square mile	0.00	1.00	0.32
Medium density area	Population density: 500-3999 persons per square mile	0.00	1.00	0.44
High density area	Population density: 4000 or more persons per square mile	0.00	1.00	0.24
Proportion of male	Count of male members/ Household size	0.00	1.00	0.49
Proportion of female	Count of female members/ Household size	0.00	1.00	0.51
Low-income household	Household income less than 50K	0.00	1.00	0.18
Medium income household	Household income in between 50K and 100K	0.00	1.00	0.33
High income household	Household income more than 100K	0.00	1.00	0.49
Less than 17 years old	Age category: less than 17 years	0.00	1.00	0.10
17 to 35 years old	Age category: less than 17 to 35 years	0.00	1.00	0.12
36 to 50 years old	Age category: less than 36 to 50 years	0.00	1.00	0.22
51 to 65 years old	Age category: less than 51 to 65 years	0.00	1.00	0.32
More than 65 years old	Age category: more than 65 years	0.00	1.00	0.24
Location: Urban	Urban = 1/ Rural = 0	0.00	1.00	0.76
Location: Rural	Urban = 0/ Rural = 1	0.00	1.00	0.24
<b><i>Dependent Variable</i></b>				
Gasoline	Fuel type: Gasoline	0.00	1.00	0.87
Non-gasoline	Fuel type: Plug in hybrid, hybrid, diesel, natural gas, and electric	0.00	1.00	0.13
Utility vehicles	Vehicle type: Passenger vans, sport utility, and station wagon	0.00	1.00	0.53
Sport sedans	Vehicle type: Convertible and coupe	0.00	1.00	0.05
Sedans	Vehicle type: Hatchback and sedan	0.00	1.00	0.31
Pick-up trucks	Vehicle type: Pick-up and truck wagon	0.00	1.00	0.11

1  
2 **5 SELECTION OF MODEL WITH FUSED DATA**

3 The proposed data fusion is based on common attributes present in MaritzCX and NHTS datasets.  
4 Four identical variables – age category, income category, household’s state, and household’s  
5 location classified as urban or rural – are present in both datasets. The reader will note that as the  
6 number of matching variables increases the number of potential matches from NHTS will reduce.  
7 In fact, as the number of matched variables increase, we end up 0 matches. Hence, we do not  
8 consider all variables for matching. For instance, in our context, we consider matching two or three  
9 variables to ensure adequate matching records exist.

10 The matching process recognizes that for any set of matching records, multiple matches  
11 are likely to exist. Among these matches, there is no way to identify an “ideal match”. Hence, we  
12 consider the fusion process with multiple candidates randomly from the pool of matched records.  
13 We systematically test the model fit of the dependent variable for different number of fusion

1 records to identify the “optimal” number of fusion records. While increasing the size of fusion  
 2 records improves the model, the increase in computation time needs to be recognized. Another  
 3 aspect to consider is that as we fuse multiple NHTS records for each MaritzCX record, the number  
 4 of vehicle purchase decisions will repeat with each fused record. We include a weight variable that  
 5 ensures each MaritzCX record accounts for only one record. For example, if the number of  
 6 matched records is 10, we ensure that across the 10 records the newly added weight adds to 1. The  
 7 actual weight can be evaluated in two approaches: deterministic and probabilistic. In the  
 8 deterministic approach an equal weight is assumed (so 1/10 in the example). In the probabilistic  
 9 approach, we let the model allocate the weight for each of the fused records. In this process, the  
 10 records that offer the largest improvement in prediction will have higher weights. The weight  
 11 function is scored based on the similarity/dissimilarity of the common attributes that were not used  
 12 for fusion. The weight score is expected to be higher for records with higher similarity. Given the  
 13 inherent random nature of the fusion process, we repeat the fusion process for a fixed number of  
 14 fused records multiple times to ensure that the results are reliable. The model parameters across  
 15 these samples are compared using a modified Wald t-test to ensure parameter stability. After  
 16 establishing parameter stability, the fused dataset and the model is finalized.

17 The data fusion and the selection of optimal model are conducted following the algorithm  
 18 presented in Figure 2. *In the first step*, 10 variable combinations: 1) age and income, 2) age and  
 19 location, 3) age and state, 4) age, income and location, 5) age, income and state, 6) age, location  
 20 and state, 7) income and location, 8) state and location, 9) state and income, and 10) state, income  
 21 and location were considered for data fusion. *In the second step*, for each combination 1-, 5-, 10-,  
 22 15-, 20-, 25-, and 30 NHTS records were randomly fused to each MaritzCX record, and an eight  
 23 alternative MNL model was developed with the fused data. Given the inherent random nature of  
 24 the fusion process, we repeat the fusion process for every number of fused records multiple times  
 25 to ensure that the results are reliable. *In the third step*, the improvement of the model LL (relative  
 26 to the model with only MaritzCX dataset) at each combination was compared. Based on the LL  
 27 improvement, the optimum number of fusion records for each variable combination was selected.  
 28 The model LL improvements ( $= LL_{Fusion\ model} - LL_{MaritzCX\ only\ model}$ ) for different variable  
 29 combinations and different number of fusion records are presented in Table 2. The reader will note  
 30 that, with the increase of the number of fusion records, the sample size of the dataset also increases.  
 31 However, the inclusion of the weight variable ensures that, the total number of cases of the fused  
 32 dataset is equivalent to the total number of cases of the MaritzCX only data. *In the fourth step*, the  
 33 selected models for all variable combinations were compared, and the optimum combination was  
 34 selected. Figure 6 shows the LL improvement for different variable combinations. It is noticeable  
 35 that for the combination of income and location the LL improvement is the highest. Therefore, this  
 36 combination was selected as the optimum final combination. Figure 6 also shows the optimum  
 37 number of fusion records for the combination income and location. It is noticeable that, after fusion  
 38 size = 15, the LL value does not change significantly. Further, increasing the number of fusion  
 39 records requires increased computational time. Considering the insignificant improvement in LL  
 40 and computational time, the optimum number of fusion records was finalized as 15. Therefore, the  
 41 proposed model framework was estimated fusing 15 NHTS records to each MaritzCX record based  
 42 on income, and location combination. The reader would note that the development of the fused  
 43 dataset does not impose any sample size restrictions on the model estimation. Further, the fused  
 44 dataset will be selected for adoption only if the fused variables (newly added variables) improve  
 45 the model fit for the dependent variable.

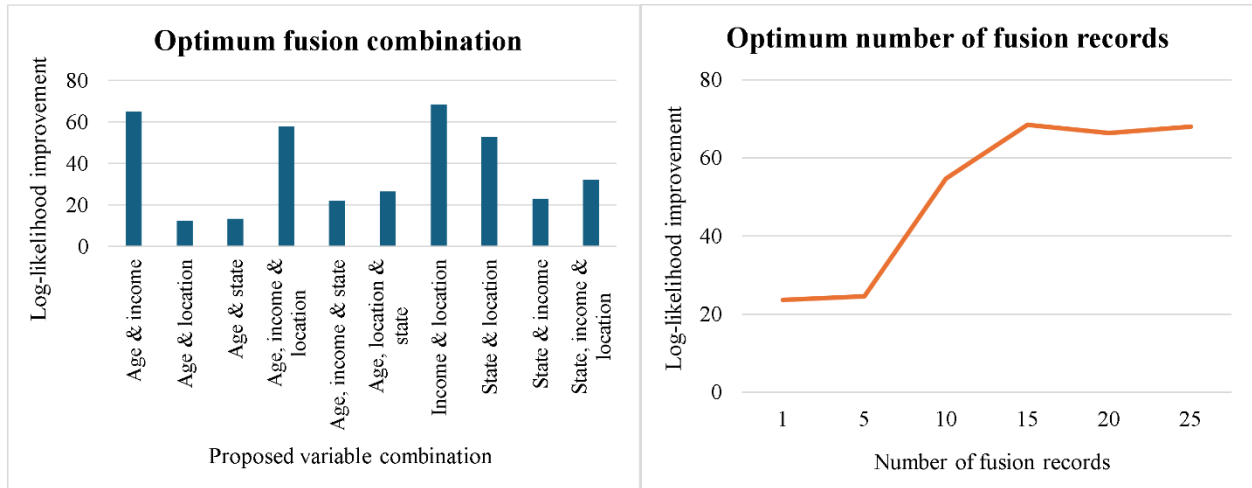
46

1 **Table 2: Log-likelihood improvements of different models**

Variable combination	Number of fusion records							
	1	5	10	15	20	25	30	35
Age and income	6.25	19.77	32.60	55.77	58.69	59.59	<b>64.92</b>	65.80
Age and location	1.81	3.89	7.01	7.60	10.95	<b>12.56</b>	12.09	---
Age and state	<b>13.36</b>	13.11	13.34	13.56	13.89	---	---	---
Age, income and location	25.20	25.27	26.00	25.87	52.24	53.70	<b>57.98</b>	---
Age, income and state	14.44	21.26	<b>22.06</b>	21.96	22.10	---	---	---
Age, location and state	25.09	25.61	<b>26.45</b>	25.82	26.53	---	---	---
Income and location	23.70	24.60	54.70	<b>68.52</b>	66.39	68.02	---	---
State and location	15.37	16.02	29.63	45.34	49.84	<b>52.92</b>	---	---
State and income	20.75	22.84	<b>23.12</b>	23.05	20.94	---	---	---
State, income and location	27.69	28.90	29.23	31.66	<b>32.25</b>	31.35	---	---

*N.B.: Bold number indicates the optimum number of fusion records*

2



3 **Figure 6: Improvement of LL across Various Model Estimations**

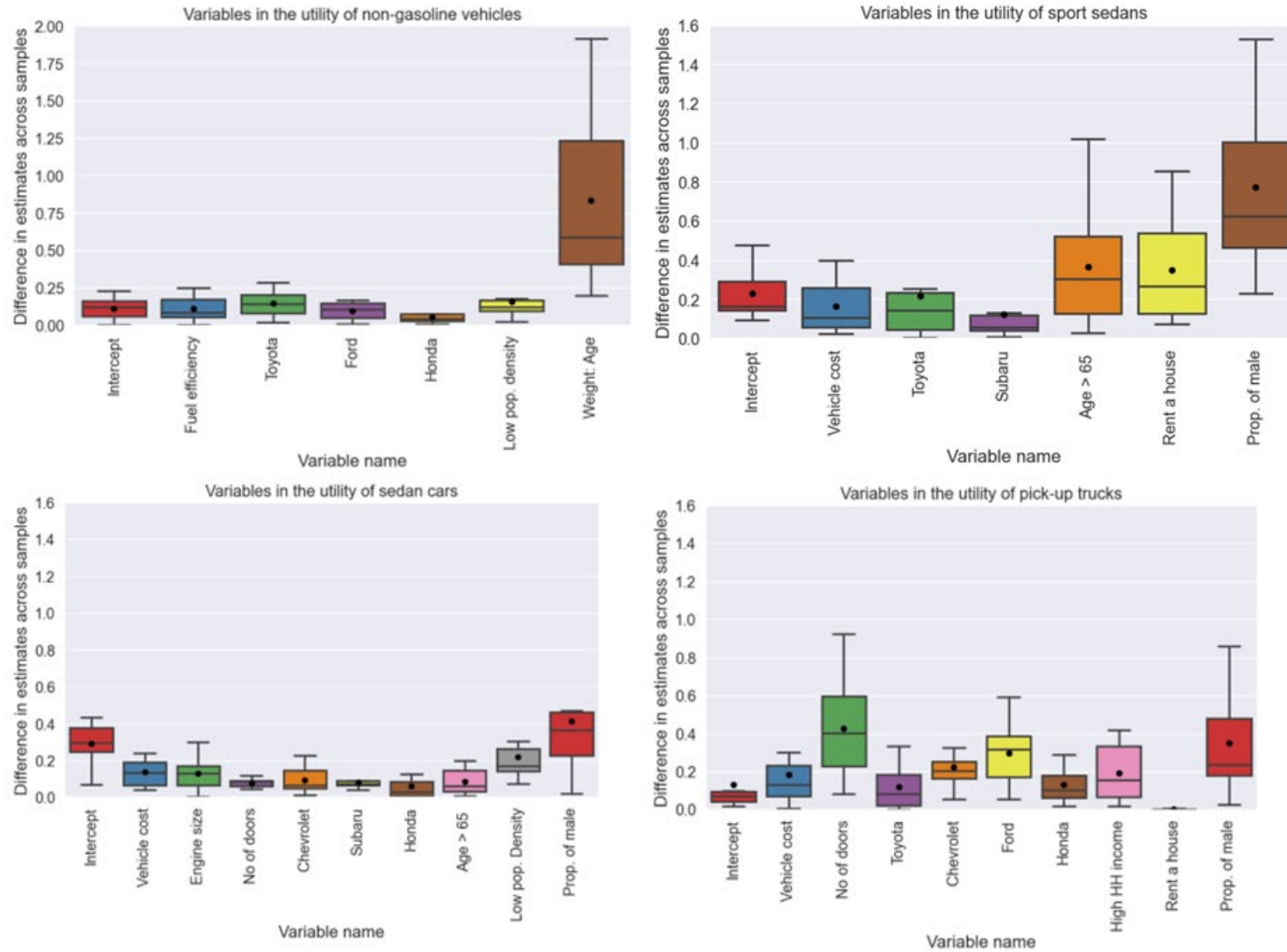
4

5 *In the final step*, the stability of the fusion model is tested. Since the 15 NHTS records were  
 6 selected randomly, it is important to ensure the random sampling does not affect the stability of  
 7 estimates. For this reason, we conducted model estimation by employing multiple random samples  
 8 following the same process used for the estimation sample. For all of these samples fusion model  
 9 specifications described were estimated. The reader would note that across the samples, it is not  
 10 likely that the parameters estimate remain identical. However, the focus is on examining if the  
 11 parameters across these multiple samples exhibit statistically significant variability. For this  
 12 purpose, we consider the mean of the parameters across the samples as the population estimate.  
 13 Subsequently, a revised Wald t-test statistic is generated for each sample parameter relative to  
 14 population mean parameter as follows:  
 15

1 Parameter test statistic =  $abs \left[ \frac{\text{Sample parameter} - \text{Population benchmark}}{\sqrt{SE_{sample}^2 + SE_{population}^2}} \right] \dots \dots \dots (8)$

2  
3  
4  
5  
6  
7  
8  
9  
10

It was found that the parameter test statistic is less than critical t statistic value (1.65) for 90% confidence interval for all variables except for the weight variable – age. For the age variable, a small number of samples (3) violate the 1.65 test statistics. But as noted in the box plot, the mean and distribution are well under the 1.65 level. The reader would note that, out of 340 (= 34 variables x 10 models) test statistics generated only 3 values are larger than 1.65. This implies that sample randomness does not affect the parameter stability across the samples (see distribution of t-statistic across all variables in Figure 7).



1  
2

**Figure 7: Test Statistics for Parameter Estimates Across Samples**

1 **6 EMPIRICAL RESULTS**

2 This study involves estimation of two model structures: 1) a binary logit model; and 2) a  
 3 multinomial logit model. The model estimation process involved the following steps. First, the  
 4 binary logit model was developed to estimate the propensity for purchasing of a new car using  
 5 NHTS dataset. Second, the new car records in NHTS dataset were fused with MaritzCX dataset  
 6 following the procedure described above, and a multinomial logit model with probabilistic  
 7 weighting was developed to estimate the choice of a new car. The estimation results of the models  
 8 with statistically significant coefficients (at 90% confidence level) are described here.

9  
 10 **6.1 Binary Logit Model**

11 The estimation results of the binary logit model are shown in Table 3.

12  
 13 **6.1.1 Household Economic Characteristics**

14 Several characteristics of the household were tested in our model. Among them household income,  
 15 household ownership, and number of workers were found to offer significant impact on purchasing  
 16 a new car. It is noticeable that medium and high-income households are more likely to purchase a  
 17 new car compared to low-income households. Households not owning a house are less likely to  
 18 buy a new car. Finally, the positive impact of the number of workers indicates that an increase in  
 19 the number of workers in a household increases the likelihood of purchasing a car.

20  
 21 **6.1.2 Household Demographic Characteristics**

22 In terms of demographic characteristics, several variables were found to significantly affect car  
 23 purchasing decisions. The number of children was found to have a negative impact indicating a  
 24 lower likelihood of purchasing a car with an increased number of children. With respect to the  
 25 proportion of female members in a household, proportion of male members offers a negative  
 26 impact on car purchasing. Households with a higher proportion of African American individuals  
 27 are found to have lower likelihood of buying a new car. Across different age categories individuals  
 28 in the age between 51 and 65 years old are found to have a lower probability of buying a new car,  
 29 while individuals of age more than 65 years old are more inclined toward purchasing a car. Finally,  
 30 people residing in rural areas are less likely to buy a car relative to people residing in urban areas.

31  
 32 **6.1.3 Spatial Characteristics**

33 Several spatial factors were found to offer significant influence on household’s vehicle purchase  
 34 decision. These variables represent vehicle purchasing fixed effects for various parts of the  
 35 country.

36  
 37 **Table 3: Estimation Result of Binary Logit Model**

Variables	Estimates	T-stat
Intercept	-2.48	-859.91
<b><i>Economic Characteristics</i></b>		
Household income (Base: Low income)		
High income household	0.89	469.98
Medium income household	0.49	270.86
Household ownership (Base: Own a house)		
Do not own a house	-0.15	-91.09

Number of workers	0.12	141.79
<b><i>Demographic Characteristics</i></b>		
Number of children	-0.01	-18.19
Gender (Base: Proportion of females)		
Proportion of males	-0.04	-19.02
Ethnicity (base: Other categories)		
Proportion of African – American individuals	-0.07	-28.08
Age (Base: Other categories)		
Proportion of individuals aged between 51 and 65 years	-0.10	-48.49
Proportion of individuals of age more than 65 years	0.06	25.56
Location (Base: Urban areas)		
Rural areas	-0.11	-62.75
<b><i>Spatial Characteristics (Base: South)</i></b>		
Northeast	-0.05	-29.50
Midwest	0.02	11.80
West	-0.17	-92.34

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

**6.2 Multinomial Logit Model**

The multinomial logit model was estimated with a two-dimensional dependent variable. The first dimension includes fuel type, and the second dimension includes vehicle type. The model estimation process employs a dimension specific approach for variable consideration as this will result in a more parsimonious model structure. The estimation results of the model are shown in Table 4. The reader would note that the variables considered in the model are drawn from MaritzCX and NHTS fused database. We discuss the results by grouping the variables into their source.

**6.2.1 MaritzCX Data**

**6.2.1.1 Vehicle Characteristics**

Several vehicle related variables were found to have a significant impact on our model. Fuel efficiency is found to offer a positive correlation with non-gasoline operated vehicles. In general, gasoline operated vehicles convert less than 40% of their energy to the vehicle usable power, leaving the rest as waste heat. In contrast, the conversion rate is 85% or higher for the electric motors in EV or fuel cells in hydrogen vehicles (Bolor et al., 2019). Further, it has been found to have similar impacts across all categories of vehicles. Vehicle price offered a positive correlation with sport sedans and pickup trucks; however, it is negatively correlated with sedans. The result is indicative of customers opting for sport sedans and pickup trucks in spite of their price. It is also interesting to observe that the impact of price is similar across both fuel types (Noel et al., 2020). A negative impact of engine size on sedans indicates that preferences for larger engine size decreases the likelihood of buying sedans. In comparison to 4-door cars, 2-door cars provide a negative preference for sedans; however, it is positively associated with pickup trucks. Finally, several vehicle brands were found to have a significant impact on our model representing brand specific fixed effects, which are not interpretable after the addition of several vehicle related attributes in the model.

1 **6.2.1.2 Household Economic Characteristics**

2 In terms of household economic characteristics, high income category households present a  
3 negative association with pick-up trucks compared with low- and medium-income households.  
4 Surprisingly, household income was found to have no significant impact on fuel choice decision  
5 (see (Mohamed et al., 2018; Singh et al., 2020) for similar findings in earlier research). It is  
6 possible that fuel choice decisions are more likely to be affected by attitudinal preferences and are  
7 not solely affected by household income.

8  
9 **6.2.1.3 Household Demographic Characteristics**

10 Only one demographic characteristic – age – was present in the MaritzCX data, and it has been  
11 found to offer significant impact in the vehicle choice model. It is noticeable that individuals aged  
12 more than 65 years old are more likely to purchase sedans but less likely to buy sport sedans.

13  
14 **6.2.1.4 Spatial Characteristics**

15 Several spatial factors were tried in our model, but they were not found to offer any significant  
16 effects on fuel type and vehicle type.

17  
18 **6.2.2 NHTS Data**

19  
20 **6.2.2.1 Household Economic Characteristics**

21 Various economic characteristics were tried in our model. However, only household ownership is  
22 found to offer significant impact on vehicle type. Households, not owning a house, are more likely  
23 to buy sport sedans and less likely to buy pickup trucks. On the other hand, they do not present  
24 any differential impacts across fuel types.

25  
26 **6.2.2.2 Household Demographic Characteristics**

27 Several demographic attributes were found to have significant influence on our model. The  
28 findings indicate that households in low dense areas are more likely to buy non-gasoline operated  
29 vehicles, and less likely to buy sedans. This is an interesting finding and could be a reflection of  
30 the availability of space for installing solar panels or other renewable energy sources fueling their  
31 non-gasoline fuel operated vehicles (such as battery EV, plug-in EV, and plug-in-hybrid EV).  
32 Further, from our two-dimensional modeling framework, the combined impact of low dense areas  
33 on non-gasoline sedans can be computed as sum of the impact of non-gasoline fuel type and sedan  
34 vehicle type as (0.41-0.30). Finally, households with higher proportion of males are less likely to  
35 buy sport sedan and sedan vehicle types, and more likely to buy pickup trucks. However, they do  
36 not show any differential impacts across fuel types.

37  
38 **6.2.2.3 Spatial Characteristics**

39 As discussed earlier, NHTS data was fused with MaritzCX data for enhancing the estimation of  
40 vehicle choice. In this context, for any common variable(s) across both datasets, MaritzCX data,  
41 being the host data, will be preferred for model estimation. Hence, these variables from NHTS  
42 were not considered in the multinomial logit model.

43  
44 **6.2.3 Weight Variable**

45 There were 4 common variables in MartizCX and NHTS data. Of these 4 variables, the fusion  
46 employed a 2-variable combination (income, and location). Hence, two common variables (state,



1 and age) existed in the fused data to parameterize the weight. So, two indicator variables were  
 2 created as one if the state (age) variable matches in both datasets and zero otherwise. Between  
 3 these two variables age is found to offer significant impact on our model as a weight variable. The  
 4 positive and significant coefficient of the weight variable implies that the contribution of records  
 5 is significantly higher when the age of MaritzCX data matches with the age of NHTS data. Thus,  
 6 we are able to differentiate between contribution of same age records to the overall model fit.  
 7

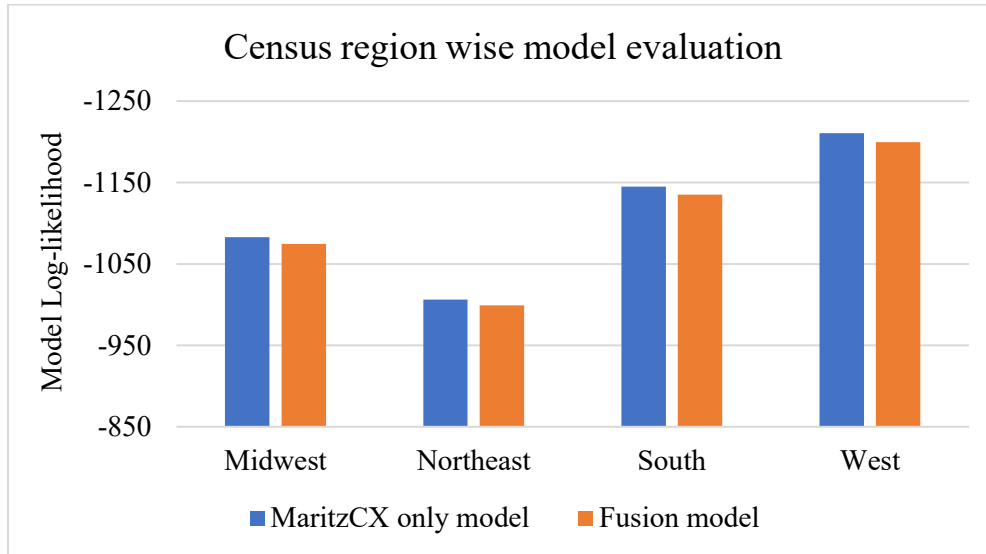
8 **Table 4: Estimation Result of Multinomial Logit Model**

Parameters	Base: Gasoline	Base: Utility vehicles		
	Non-gasoline	Sport sedans	Sedans	Pickup trucks
	Coefficient (T-value)	Coefficient (T-value)	Coefficient (T-value)	Coefficient (T-value)
Intercept	-6.19 (-22.11)	-8.15 (-12.49)	8.78 (25.22)	-8.55 (-9.94)
<b>MaritzCX Data</b>				
<i>Vehicle Characteristics</i>				
Fuel efficiency (mpg)	0.15 (14.76)	--	--	--
Vehicle price (in thousands)	--	0.16 (10.86)	-0.03 (-3.98)	0.14 (9.21)
Engine size (liters)	--	--	-2.87 (-26.25)	--
Number of doors (base: 4-door car)				
2-door car	--	--	-1.73 (-6.28)	3.16 (12.30)
Vehicle brands (Base: Other brands)				
Toyota	0.98 (7.67)	-1.39 (-3.25)	--	2.96 (16.07)
Chevrolet	--	--	-0.66 (-3.42)	2.45 (12.44)
Subaru	--	-1.42 (-3.35)	-0.64 (-4.54)	
Ford	-1.04 (-4.31)	--	--	2.23 (11.65)
Honda	-1.65 (-4.00)	--	-0.51 (-2.65)	1.34 (4.85)
<i>Economic Characteristics</i>				
Household income (base: Low and medium income)				
High income	--	--	--	-0.51 (-4.12)
<i>Demographic Characteristics</i>				
Age (Base: Less than 65 years old)				
Age more than 65 years old	--	-0.53 (-2.74)	0.34 (3.95)	--
<b>NHTS Data</b>				
<i>Economic Characteristics</i>				
Household ownership (Base: Own a house)				
Rent a house	--	2.32 (8.89)	--	-8.48 (-2.42)
<i>Demographic Characteristics</i>				
Population density (Base: High and medium density)				
Low population density	0.41 (2.60)	--	-0.30 (-2.14)	--
Gender (Base: Proportion of female)				
Proportion of male	--	-2.05 (-2.97)	-0.62 (-1.67)	1.76 (1.81)
<b>Weight Variable</b>				
Age	0.60 (2.25)			

9

1 **7 MODEL VALIDATION**

2 To evaluate the performance of the proposed fusion model, a prediction exercise was conducted  
 3 at different census regions using a validation dataset of 4,000 records (1,000 records from each  
 4 region). In the validation exercise, we evaluated the model performance of the MaritzCX only  
 5 model (that includes only MaritzCX variables) and the fusion model (that includes both MaritzCX  
 6 and NHTS variables). The predicted LL (BIC) values for the MaritzCX only model and the fusion  
 7 model are -4,445 (9,106) and -4,409 (9,100) respectively. Thus, the test statistic of the log-  
 8 likelihood ratio test is  $(2(-4409+4445))$  or 72, with an additional eight parameters in the fusion  
 9 model. The test statistic obtained is greater than the corresponding  $\chi^2$  value at any level of  
 10 significance highlighting the superiority of the proposed model. Clearly, the fused model offered  
 11 significant improvement highlighting the value of the new independent variables form NHTS  
 12 dataset. In addition, we computed the prediction accuracies of the two modeling frameworks on  
 13 the validation data. For this purpose, we computed the number of records for which the chosen  
 14 alternative has the highest predicted probability. The results indicate that the fusion model  
 15 (prediction accuracy = 71.4%) outperforms the MaritzCX only model (prediction accuracy =  
 16 69.5%). Further, we compared the performance of the two modeling frameworks across the four  
 17 census regions. The results of the comparison are shown in Figure 8. From the figure it can be  
 18 observed that, for every census region our proposed fusion model outperforms the MartizCX only  
 19 model. The result supports our hypothesis that building a fused dataset using NHTS dataset  
 20 significantly improves model predictive power across different data segments. The finding is quite  
 21 encouraging and bodes well for future applications of the proposed model.  
 22



23 **Figure 8: LL comparison between proposed fusion model and MaritzCX only model**

24  
25  
26  
27  
28  
29  
30  
31  
32

**8 ELASTICITY EFFECTS**

The coefficients of the independent variables in Table 4 do not directly provide the exact magnitude of the impact of variables on household’s vehicle purchase decision. The effects of the variables might change across different combinations of fuel type and body type. To evaluate this variability, we computed the elasticity of the variables. More specifically, we estimated the percentage change in the probability of buying a vehicle of specific fuel and vehicle type in response to any change in the explanatory variables. For continuous independent variables in the

1 model, the change was conducted in an increment of 10% (see (Bhowmik et al., 2022) for similar  
2 analysis). For indicator variables, the change was obtained by changing the value of the variable  
3 to one for the subsample of observations for which the variable takes a value of zero and vice-  
4 versa (see (Kabli et al., 2020) for similar analysis). The variables fuel efficiency, vehicle price and  
5 engine size are alternative specific. In the elasticity analysis, we examined how the changes in any  
6 of these variables for each alternative affect the probability of all the alternatives (self and cross-  
7 elasticity effects). The elasticity effects are shown in Table 5. The underlined values in the table  
8 indicate the changes in the probability of any alternative due to the changes in exploratory variables  
9 associated with the respective alternative (self-elasticity).

10 Several observations can be made from the results in Table 5. First, as expected, we find  
11 fuel efficiency as one of the important factors influencing vehicle purchase decisions, particularly  
12 for non-gasoline vehicles. The results illustrate that increasing the fuel efficiency by 10 % will  
13 result in an increased likelihood of purchasing non-gasoline vehicles irrespective of the vehicle  
14 types. Second, we find an interesting trend for the cost/price variable. With respect to sports sedans  
15 and pickup trucks, the effect is positive irrespective of the fuel type suggesting people considering  
16 these vehicles are not deterred by price. On the other hand, people are more sensitive to price when  
17 considering the purchase of a sedan. The effect is more pronounced (higher reduction) for gasoline  
18 sedans as indicated by the higher negative magnitude in the Table. Third, engine size of the  
19 vehicles also plays a significant role in vehicle purchase decision. The results suggest people's  
20 lower preference of buying sedans with large engine size, perhaps indicative of the increased fuel  
21 consumption. Fourth, the elasticity effects of vehicle brands indicate that while purchasing non-  
22 gasoline pickup trucks, customers show higher affinity for Toyota brand. Finally, changes in  
23 several economic and demographic characteristics employed in the model (through fusion from  
24 NHTS), are found to offer notable impact on vehicle type and fuel type selections.  
25

1 **Table 5: Elasticity Effect Analysis for All Fuel Type and Vehicle Type Combinations**

Parameter		Gasoline utility vehicles	Gasoline sport sedans	Gasoline sedans	Gasoline pick-up trucks	Non-gasoline utility vehicle	Non-gasoline sport sedans	Non-gasoline sedans	Non-gasoline pick-up trucks
		% changes in the probability							
Fuel efficiency*	Non-gasoline sport sedans	-0.20	-0.63	-0.12	-0.04	-0.20	<u>46.05</u>	-0.08	-0.04
	Non-gasoline sedans	-1.17	-0.89	-0.87	-0.87	-1.36	-1.16	<u>16.85</u>	-1.35
	Non-gasoline pick-up trucks	-0.45	-0.19	-0.26	-1.93	-0.77	-0.14	-0.29	<u>41.61</u>
	Non-gasoline utility vehicle	-2.91	-2.11	-1.79	-2.15	<u>40.11</u>	-2.89	-1.34	-3.46
Vehicle price*	Gasoline sport sedans	-3.25	<u>54.47</u>	-1.79	-1.72	-2.52	-10.39	-0.99	-1.35
	Gasoline sedans	1.75	1.47	<u>-4.56</u>	0.87	1.67	1.75	0.79	0.96
	Gasoline pick-up trucks	-4.66	-2.33	-2.22	<u>32.87</u>	-4.72	-1.36	-1.80	-17.27
	Non-gasoline sport sedans	-0.33	-1.06	-0.20	-0.07	-0.32	<u>75.35</u>	-0.12	-0.07
	Non-gasoline sedans	0.25	0.18	0.18	0.17	0.29	0.25	<u>-3.49</u>	0.27
	Non-gasoline pick-up trucks	-0.62	-0.27	-0.34	-2.69	-1.04	-0.20	-0.38	<u>57.09</u>
Engine size*	Gasoline sedans	11.33	9.46	-29.53	5.68	10.79	11.24	5.07	6.36
	Non-gasoline sedans	1.46	1.10	1.08	1.06	1.73	1.48	-20.96	1.72
Number of doors:2		-24.27	10.95	-65.38	267.07	-24.69	8.86	-49.99	269.35
Toyota		-36.23	-65.30	-22.05	168.65	31.70	-30.52	11.27	408.80
Chevrolet		-18.53	-0.43	-35.82	172.20	-19.51	1.92	-26.21	170.90
Subaru		13.67	-57.29	-18.01	7.52	12.46	-58.25	-13.09	8.51
Ford		-17.46	-4.65	-9.63	152.62	-62.02	-55.24	-27.04	8.63
Honda		1.41	7.03	-20.13	111.59	-73.08	-71.76	-43.42	-51.84
Age more than 65 years old		-1.86	-15.17	6.70	-1.16	-2.05	-17.34	3.28	-1.60
High income household		-1.01	-1.11	-0.38	6.02	-1.03	-0.27	-0.50	5.88
Household ownership: No		-1.01	-1.11	-0.38	6.02	-1.03	-0.27	-0.50	5.88
Low density area		0.76	0.73	-9.59	-0.59	26.35	41.86	2.96	18.47
Proportion of male		0.06	-5.54	-1.51	5.78	-0.11	-5.64	-1.08	5.99
<i>*Alternative specific variable</i>									

2

## 9 CONCLUSION

The growing adoption of alternative fuel vehicles (AFVs) across the country has resulted in increased attention on modeling AFV adoption in the US. Several research attempts have developed models on AFV adoption primarily relying on stated preference data. There is limited research employing revealed preference data for AFV adoption. The objective of the current research is to study vehicle purchase behavior with emphasis on AFVs using a vehicle purchase dataset MaritzCX New Vehicle Customer Study (NVCS). MaritzCX data provides the most comprehensive vehicle purchase data in the US and Canada. However, it is not employed in transportation modeling to understand population vehicle purchase behavior for two reasons. First, MaritzCX records represent a self-selected sample of vehicle purchasers (and not the general population). Second, MaritzCX data provides very limited data on the household characteristics of the vehicle purchasers. For estimating vehicle ownership models employing MaritzCX data along with different categories of exogenous variables, it might be necessary to merge several datasets. However, the MaritzCX datasets and different survey datasets that represent household characteristics do not have any common identifier variable that can be utilized for merging those datasets. Hence, the rich set of vehicle purchase observations are not useful in understanding vehicle fleet evolution at the household resolution.

In the current research effort, we propose solutions to address these limitations and develop a framework that will allow us to tap into the rich information available in MaritzCX data. The self-selection of vehicle purchasers is addressed by developing a vehicle purchase decision model using data from the 2017 National Household Travel Survey (NHTS, 2017). To improve the independent variables available in the MaritzCX data, we employ an innovative data fusion approach that probabilistically links records from MaritzCX with records from NHTS. The proposed data fusion is based on common attributes across the two datasets. Four identical variables – age category, income category, household’s state, and household’s location classified as urban or rural – are present in both MaritzCX and NHTS dataset. Therefore, in the current study, 10 variable combinations: 1) age and income, 2) age and location, 3) age and state, 4) age, income and location, 5) age, income and state, 6) age, location and state, 7) income and location, 8) state and location, 9) state and income, and 10) state, income and location were tested for data fusion. The proposed model framework was estimated fusing 15 NHTS records with each MaritzCX record based on income and location combination. The reader would note that the development of the fused dataset does not impose any sample size restrictions on the model estimation. Further, the fused dataset will be selected for adoption only if the fused variables (newly added variables) improve the model fit for the dependent variable. The model fit measures and independent variables significant clearly illustrate how the model framework has been significantly improved based on the fused data (relative to MaritzCX only data). The model results also highlight various interesting patterns of customers’ AFV purchase decision. Among vehicle characteristics, fuel efficiency is found to offer a positive correlation with AFVs. The results illustrate that increasing the fuel efficiency by 10 % will result in an increased likelihood of purchasing AFVs irrespective of the vehicle types. It is also interesting to observe that the impact of vehicle price is similar across both fuel types. Further, among socioeconomic variables, household economic characteristics (such as household income and household ownership) do not provide any significant impact on fuel choice decision. On the other hand, the findings indicate that households in low dense areas are more likely to buy AFVs.

The coefficients of the independent variables of the proposed model do not directly provide the exact magnitude of the impact of variables on household’s vehicle purchase decision. The

1 effects of the variables might change across different combinations of fuel type and body type. To  
2 evaluate this variability, we generated elasticity effects in response to changes to independent  
3 variables. The elasticity results illustrate how the model system can be applied to identify  
4 important variables.

5 The proposed methodology can be employed to examine the effect of incentives on  
6 customer's vehicle purchase behavior. The incentives can be incorporated as an increase in  
7 household income and employed to examine the impact on fuel type and body type choices. In  
8 addition to estimating customers' vehicle choice behavior, the proposed methodology can also be  
9 applied to address diverse planning challenges, including the utilization of – a) social media data  
10 or location-based smartphone data, fused with household's socio-demographic-, economic-,  
11 vehicle ownership- and commuting characteristics for travel behavior analysis, disaster evacuation  
12 and public experience assessment regarding various transportation modes (see (Bhowmik et al.,  
13 2024b) for example application for evacuation behavior analysis).

14 To be sure, this study is not without limitations. As mentioned earlier, we had to consider  
15 EVs in our modeling framework within the non-gasoline fuel type category. With more years of  
16 recent data, this assumption can be relaxed to focus directly on EVs. The approach proposed in  
17 our study will need a minor update to account for this change.

## 18 **ACKNOWLEDGMENTS**

19 The authors would like to thank Dr. Joshua Linn (Associate Professor, Department of Agricultural  
20 and Resource, University of Maryland, Email: linn@umd.edu), Dr. James Archsmith (Assistant  
21 Professor, Department of Agricultural and Resource Economics, University of Maryland, Email:  
22 archsmit@umd.edu) and the Resources For the Future (RFF) for providing the data utilized in this  
23 study.  
24

## 25 **AUTHORS' CONTRIBUTION**

26 **Md Istiak Jahan:** Software, Validation, Formal analysis, Investigation, Data Curation, Writing -  
27 Original Draft, Visualization. **Tanmoy Bhowmik:** Conceptualization, Methodology, Software,  
28 Formal analysis, Investigation, Visualization. **Sachraa G. Borjigin:** Investigation, Writing -  
29 Original Draft. **Jiehong Lou:** Writing - Original Draft, Visualization. **Nneoma M. Ugwu:** Writing  
30 - Original Draft. **Deb A. Niemeier:** Writing - Original Draft. **Naveen Eluru:** Conceptualization,  
31 Methodology, Investigation, Writing - Review & Editing, Supervision.  
32

## 33 **DISCLOSURE STATEMENT**

34 The authors declare that they have no known competing financial interests or personal  
35 relationships that could have appeared to influence the work reported in this paper.  
36

## 37 **REFERENCES**

- 38 Ackaah, W., Kanton, A. T., & Osei, K. K. (2022). Factors influencing consumers' intentions to  
39 purchase electric vehicles in Ghana. *Transportation Letters*, 14(9), 1031–1042.  
40 <https://doi.org/10.1080/19427867.2021.1990828>  
41 Ahn, J., Jeong, G., & Kim, Y. (2008). A forecast of household ownership and use of alternative  
42 fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics*, 30(5),  
43 2091–2104. <https://doi.org/10.1016/j.eneco.2007.10.003>  
44 Anderhofstadt, B., & Spinler, S. (2019). Factors affecting the purchasing decision and operation  
45 of alternative fuel-powered heavy-duty trucks in Germany – A Delphi study. *Transportation*  
46 *Research Part D: Transport and Environment*, 73, 87–107.  
47

- 1 <https://doi.org/10.1016/j.trd.2019.06.003>
- 2 Ankney, K., & Leard, B. (2021). *How Much Do Consumers Value Fuel Cost Savings? Evidence*  
3 *from the Passenger Vehicle Leasing Market* (Issue August).
- 4 Anowar, S., Eluru, N., & Miranda-Moreno, L. F. (2016). Analysis of vehicle ownership evolution  
5 in Montreal, Canada using pseudo panel analysis. *Transportation*, 43, 531–548.  
6 <https://doi.org/10.1007/s11116-015-9588-z>
- 7 Anowar, S., Eluru, N., Miranda-Moreno, L. F., & Lee-Gosselin, M. (2015). Joint econometric  
8 analysis of temporal and spatial flexibility of activities, vehicle type choice, and primary  
9 driver selection. *Transportation Research Record*, 2525, 32–41.  
10 <https://doi.org/10.3141/2495-04>
- 11 Anowar, S., Eluru, N., Miranda-Moreno, L. F., & Miranda-moreno, L. F. (2014). Alternative  
12 Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive  
13 Review. *Transport Reviews*, 34(4), 441–473. <https://doi.org/10.1080/01441647.2014.915440>
- 14 Arokiaraj, D., & M.Banumathi. (2014). Factors Influencing the Purchase Decision of Passenger  
15 Cars in Puduchery. *IJEMR*, 4(4), 1–10.
- 16 Ayetor, G. K., Nyarko, F. K., & Andoh, P. Y. (2023). Achieving Cost Parity for Battery Electric  
17 Vehicles in Africa: a case study of Ghana. *Transportation Letters*, 00(00), 1–13.  
18 <https://doi.org/10.1080/19427867.2023.2168348>
- 19 Baltas, G., & Saridakis, C. (2013). An empirical investigation of the impact of behavioural and  
20 psychographic consumer characteristics on car preferences: An integrated model of car type  
21 choice. *Transportation Research Part A: Policy and Practice*, 54, 92–110.  
22 <https://doi.org/10.1016/j.tra.2013.07.007>
- 23 Bauer, G. (2018). The impact of battery electric vehicles on vehicle purchase and driving behavior  
24 in Norway. *Transportation Research Part D: Transport and Environment*, 58(December  
25 2017), 239–258. <https://doi.org/10.1016/j.trd.2017.12.011>
- 26 Bhat, C. R., Sen, S., & Eluru, N. (2009). The impact of demographics, built environment attributes,  
27 vehicle characteristics, and gasoline prices on household vehicle holdings and use.  
28 *Transportation Research Part B: Methodological*, 43(1), 1–18.  
29 <https://doi.org/10.1016/j.trb.2008.06.009>
- 30 Bhowmik, T., Eluru, N., Hasan, S., Culotta, A., & Chandra, K. (2024). Predicting hurricane  
31 evacuation behavior synthesizing data from travel surveys and social media. *Transportation*  
32 *Research Part C*, 165(July), 104753. <https://doi.org/10.1016/j.trc.2024.104753>
- 33 Bhowmik, T., Iraganaboina, N. C., & Eluru, N. (2024). A Novel Maximum Likelihood Based  
34 Probabilistic Behavioral Data Fusion Algorithm for Modeling Residential Energy  
35 Consumption. *PLoS ONE*, Under revi.
- 36 Bhowmik, T., Yasmin, S., & Eluru, N. (2022). Accommodating for systematic and unobserved  
37 heterogeneity in panel data: Application to macro-level crash modeling. *Analytic Methods in*  
38 *Accident Research*, 33, 100202. <https://doi.org/10.1016/j.amar.2021.100202>
- 39 Biswas, A., Mukherjee, A., & Roy, M. (2014). Leveraging Factors for Consumers' Car Purchase  
40 Decisions- A Study in an Emerging Economy. *Journal of Management Policies and*  
41 *Practices*, 2(2), 99–111. <https://doi.org/10.4324/9781315576657-19>
- 42 Blink. (2023). *How EVs Are Reducing Carbon (CO2) Emissions*.  
43 <https://doi.org/https://blinkcharging.com/blog/how-evs-are-reducing-carbon-co2-emissions>
- 44 Bolduc, D., Boucher, N., & Alvarez-Daziano, R. (2008). Hybrid Choice Modeling of New  
45 Technologies for Car Choice in Canada. *Transportation Research Record: Journal of the*  
46 *Transportation Research*, 2082, 63–71. <https://doi.org/10.3141/2082-08>

- 1 Boloor, M., Valderrama, P., Statler, A., & Garcia, S. (2019). *Electric Vehicle Basics*. Natural  
2 Resources Defence Council. [https://www.nrdc.org/bio/madhur-boloor/electric-vehicle-  
4 basics](https://www.nrdc.org/bio/madhur-boloor/electric-vehicle-<br/>3 basics)
- 4 Bubeck, S., Tomaschek, J., & Fahl, U. (2016). Perspectives of electric mobility: Total cost of  
5 ownership of electric vehicles in Germany. *Transport Policy*, 50, 63–77.  
6 <https://doi.org/10.1016/J.TRANPOL.2016.05.012>
- 7 Caggiani, L., Prencipe, L. P., & Ottomanelli, M. (2021). A static relocation strategy for electric  
8 car-sharing systems in a vehicle-to-grid framework. *Transportation Letters*, 13(3), 219–228.  
9 <https://doi.org/10.1080/19427867.2020.1861501>
- 10 Chang, T., & Hsiao, W. (2011). Consumers' Automotive Purchase Decisions: The Significance of  
11 Vehicle-based Infotainment Systems. *African Journal of Business Management*, 5(11), 4152–  
12 4163. <https://doi.org/10.5897/AJBM10.480>
- 13 Chen, T. D., Wang, Y., & Kockelman, K. M. (2015). Where are the electric vehicles? A spatial  
14 model for vehicle-choice count data. *Journal of Transport Geography*, 43, 181–188.  
15 <https://doi.org/10.1016/J.JTRANGE.2015.02.005>
- 16 Cirillo, C., Xu, R., & Bastin, F. (2016). A Dynamic Formulation for Car Ownership Modeling.  
17 *Transportation Science*, 50(1), 322–335. <https://doi.org/10.1287/TRSC.2015.0597>
- 18 Daina, N., Sivakumar, A., & Polak, J. W. (2017). Modelling electric vehicles use: a survey on the  
19 methods. *Renewable and Sustainable Energy Reviews*, 68, 447–460.  
20 <https://doi.org/10.1016/J.RSER.2016.10.005>
- 21 Dhanabalan, T., Subha, K., Shanthi, R., & Sathish, A. (2018). Factors influencing consumers' car  
22 purchasing decision in indian automobile industry. *International Journal of Mechanical  
23 Engineering and Technology*, 9(10), 53–63.
- 24 Eluru, N., Bhat, C. R., Pendyala, R. M., & Konduri, K. C. (2010). A joint flexible econometric  
25 model system of household residential location and vehicle fleet composition/usage choices.  
26 *Transportation*, 37(4), 603–626. <https://doi.org/10.1007/s11116-010-9271-3>
- 27 Eppstein, M. J., Grover, D. K., Marshall, J. S., & Rizzo, D. M. (2011). An agent-based model to  
28 study market penetration of plug-in hybrid electric vehicles. *Energy Policy*, 39(6), 3789–  
29 3802. <https://doi.org/10.1016/j.enpol.2011.04.007>
- 30 Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with  
31 an application to the effects of residential density. *Transportation Research Part B:  
32 Methodological*, 42(9), 736–758. <https://doi.org/10.1016/J.TRB.2008.01.004>
- 33 Fevang, E., Figenbaum, E., Fridstrøm, L., Halse, A. H., Hauge, K. E., Johansen, B. G., & Raaum,  
34 O. (2021). Who goes electric? The anatomy of electric car ownership in Norway.  
35 *Transportation Research Part D: Transport and Environment*, 92, 102727.  
36 <https://doi.org/10.1016/J.TRD.2021.102727>
- 37 Gillingham, K., Iskhakov, F., Munk-Nielsen, A., Rust, J., & Schjerning Bertel. (2015). *A Dynamic  
38 Model of Vehicle Ownership, Type Choice, and Usage*.
- 39 Guo, D., Wang, J., Zhao, J. B., Sun, F., Gao, S., Li, C. D., Li, M. H., & Li, C. C. (2019). A vehicle  
40 path planning method based on a dynamic traffic network that considers fuel consumption  
41 and emissions. *Science of The Total Environment*, 663, 935–943.  
42 <https://doi.org/10.1016/J.SCITOTENV.2019.01.222>
- 43 Hackbarth, A., & Madlener, R. (2016). Willingness-to-pay for alternative fuel vehicle  
44 characteristics: A stated choice study for Germany. *Transportation Research Part A: Policy  
45 and Practice*, 85, 89–111. <https://doi.org/10.1016/J.TRA.2015.12.005>
- 46 Hara, K., & Asahi, Y. (2020). Feature analysis of customers purchasing cars in Japan. *Lecture*



- 1        *Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and*  
2        *Lecture Notes in Bioinformatics)*, 12184 LNCS, 201–211. [https://doi.org/10.1007/978-3-030-](https://doi.org/10.1007/978-3-030-50020-7_13/FIGURES/8)  
3        [50020-7\\_13/FIGURES/8](https://doi.org/10.1007/978-3-030-50020-7_13/FIGURES/8)
- 4 Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B.  
5        (2019). Implementation of Naïve Bayes Classification Method for Predicting Purchase. *2018*  
6        *6th International Conference on Cyber and IT Service Management, CITSM 2018, Citsm*, 1–  
7        5. <https://doi.org/10.1109/CITSM.2018.8674324>
- 8 Hoen, A., & Koetse, M. J. (2014). A choice experiment on alternative fuel vehicle preferences of  
9        private car owners in the Netherlands. *Transportation Research Part A: Policy and Practice*,  
10        *61*, 199–215. <https://doi.org/10.1016/J.TRA.2014.01.008>
- 11 Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2016). Are There Environmental  
12        Benefits from Driving Electric Vehicles? The Importance of Local Factors †. *American*  
13        *Economic Review*, *106*(12), 3700–3729. <https://doi.org/10.1257/aer.20150897>
- 14 Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2019). Distributional effects of air  
15        pollution from electric vehicle adoption. *Journal of the Association of Environmental and*  
16        *Resource Economists*, *6*(S1), S65–S94.  
17        [https://doi.org/10.1086/701188/SUPPL\\_FILE/2016184APPENDIX.PDF](https://doi.org/10.1086/701188/SUPPL_FILE/2016184APPENDIX.PDF)
- 18 Imani, A. F., Ghafghazi, G., Eluru, N., & Pinjari, A. R. (2014). A multiple-discrete approach for  
19        examining vehicle type use for daily activity participation decisions. *Transportation Letters*,  
20        *6*(1), 1–13. <https://doi.org/10.1179/1942787513Y.0000000001>
- 21 Jabbari, P., Chernicoff, W., & Mackenzie, D. (2017). Analysis of electric vehicle purchaser  
22        satisfaction and rejection reasons. *Transportation Research Record*, *2628*(1), 110–119.  
23        <https://doi.org/10.3141/2628-12>
- 24 Jian, S., Rashidi, T. H., & Dixit, V. (2017). An analysis of carsharing vehicle choice and utilization  
25        patterns using multiple discrete-continuous extreme value (MDCEV) models. *Transportation*  
26        *Research Part A: Policy and Practice*, *103*, 362–376.  
27        <https://doi.org/10.1016/J.TRA.2017.06.012>
- 28 Joshi, D., & Bhatt, V. (2018). A Study on Factors Influencing Consumer’S Preference While  
29        Making Purchase Decision of First Own Car in Ahmedabad City. *Roots International Journal*  
30        *of Multidisciplinary Researches*, *2*(2), 67–76.
- 31 Kabli, A., Bhowmik, T., & Eluru, N. (2020). A multivariate approach for modeling driver injury  
32        severity by body region. *Analytic Methods in Accident Research*, *28*, 100129.  
33        <https://doi.org/10.1016/J.AMAR.2020.100129>
- 34 Kanyepe, J., Tukuta, M., & Chirisa, I. (2021). Urban Land-use and Traffic Congestion: Mapping  
35        the Interaction. *Journal of Contemporary Urban Affairs*, *5*(1), 77–84.  
36        <https://doi.org/10.25034/ijcua.2021.v5n1-6>
- 37 Khan, N. A., & Habib, M. A. (2021). Development of households’ vehicle choice models for travel  
38        accompanying arrangements. *Transportation Engineering*, *4*(November 2020), 100071.  
39        <https://doi.org/10.1016/j.treng.2021.100071>
- 40 Kim, J., & Kim, M. S. (2014). Analysis of automobile repeat-purchase behaviour on CRM.  
41        *Industrial Management and Data Systems*, *114*(7), 994–1006. [https://doi.org/10.1108/IMDS-](https://doi.org/10.1108/IMDS-01-2014-0031)  
42        [01-2014-0031](https://doi.org/10.1108/IMDS-01-2014-0031)
- 43 Kiran, P., & Shanmugam, V. (2017). A Logistic regression model to identify the key attributes  
44        considered by consumers for purchasing a car. *International Journal of Economic Research*.
- 45 Knez, M., Jereb, B., & Obrecht, M. (2014). Factors influencing the purchasing decisions of low  
46        emission cars: A study of Slovenia. *Transportation Research Part D: Transport and*

- 1        *Environment*, 30(2014), 53–61. <https://doi.org/10.1016/j.trd.2014.05.007>
- 2 Kumar, P., & Chakrabarty, S. (2020). *Total Cost of Ownership Analysis of the Impact of Vehicle*  
3 *Usage on the Economic Viability of Electric Vehicles in India*. 2674(11), 563–572.  
4 <https://doi.org/10.1177/0361198120947089>
- 5 Leard, B. (2022). Estimating consumer substitution between new and used passenger vehicles.  
6 *Journal of the Association of Environmental and Resource Economists*, 9(1), 27–49.  
7 <https://doi.org/10.1086/715814>
- 8 Leard, B., Linn, J., & Springel, K. (2020). *Have US Fuel Economy and Greenhouse Gas Emissions*  
9 *Standards Improved Social Welfare?*
- 10 Lee, R., & Brown, S. (2021). Evaluating the role of behavior and social class in electric vehicle  
11 adoption and charging demands. *ISCIENCE*, 24, 102914.  
12 <https://doi.org/10.1016/j.isci.2021.102914>
- 13 Liao, F., Molin, E., Timmermans, H., & van Wee, B. (2019). Consumer preferences for business  
14 models in electric vehicle adoption. *Transport Policy*, 73(October 2018), 12–24.  
15 <https://doi.org/10.1016/j.tranpol.2018.10.006>
- 16 Liu, J., Wang, X., & Khattak, A. (2016). Customizing driving cycles to support vehicle purchase  
17 and use decisions: Fuel economy estimation for alternative fuel vehicle users. *Transportation*  
18 *Research Part C: Emerging Technologies*, 67, 280–298.  
19 <https://doi.org/10.1016/j.trc.2016.02.016>
- 20 Lowe, M., Adlakha, D., Sallis, J. F., Salvo, D., Cerin, E., Moudon, A. V., Higgs, C., Hinckson, E.,  
21 Arundel, J., Boeing, G., Liu, S., Mansour, P., Gebel, K., Puig-Ribera, A., Mishra, P. B.,  
22 Bozovic, T., Carson, J., Dygrýn, J., Florindo, A. A., ... Giles-Corti, B. (2022). City planning  
23 policies to support health and sustainability: an international comparison of policy indicators  
24 for 25 cities. *The Lancet Global Health*, 10(6), e882–e894. [https://doi.org/10.1016/S2214-](https://doi.org/10.1016/S2214-109X(22)00069-9)  
25 [109X\(22\)00069-9](https://doi.org/10.1016/S2214-109X(22)00069-9)
- 26 Ma, J., & Mayburov, I. (2021). Using DID Model to Analyse the Impact of Vehicle Purchase Tax  
27 on the Sales of Alternative Fuel Vehicle. *IOP Conference Series: Earth and Environmental*  
28 *Science*, 849(1). <https://doi.org/10.1088/1755-1315/849/1/012006>
- 29 Mabit, S. L. (2014). Vehicle type choice under the influence of a tax reform and rising fuel prices.  
30 *Transportation Research Part A: Policy and Practice*, 64, 32–42.  
31 <https://doi.org/10.1016/j.tra.2014.03.004>
- 32 MaritzCX. (2019). *Vehicle Manufacturers' Secret Sauce, The MaritzCX New Vehicle Customer*  
33 *Survey, Marks 50-Year Anniversary*. InMoment. [https://inmoment.com/news/vehicle-](https://inmoment.com/news/vehicle-manufacturers-secret-sauce-the-maritzcx-new-vehicle-customer-survey-marks-50-year-anniversary/)  
34 [manufacturers-secret-sauce-the-maritzcx-new-vehicle-customer-survey-marks-50-year-](https://inmoment.com/news/vehicle-manufacturers-secret-sauce-the-maritzcx-new-vehicle-customer-survey-marks-50-year-anniversary/)  
35 [anniversary/](https://inmoment.com/news/vehicle-manufacturers-secret-sauce-the-maritzcx-new-vehicle-customer-survey-marks-50-year-anniversary/)
- 36 McGuckin, N., Fucci, A., & Jenkins, D. E. (2018). *Summary of Travel Trends: 2017 National*  
37 *Household Travel Survey*.
- 38 Menon, N., Barbour, N., Zhang, Y., Pinjari, A. R., & Mannering, F. (2019). Shared autonomous  
39 vehicles and their potential impacts on household vehicle ownership: An exploratory  
40 empirical assessment. *International Journal of Sustainable Transportation*, 13(2), 111–122.  
41 <https://doi.org/10.1080/15568318.2018.1443178>
- 42 Minos, S. (2022). *New Plug-in Electric Vehicle Sales in the United States Nearly Doubled from*  
43 *2020 to 2021*. US Department of Energy. [https://www.energy.gov/energysaver/articles/new-](https://www.energy.gov/energysaver/articles/new-plug-electric-vehicle-sales-united-states-nearly-doubled-2020-2021)  
44 [plug-electric-vehicle-sales-united-states-nearly-doubled-2020-2021](https://www.energy.gov/energysaver/articles/new-plug-electric-vehicle-sales-united-states-nearly-doubled-2020-2021)
- 45 Mohamed, M., Higgins, C. D., Ferguson, M., & Réquia, W. J. (2018). The influence of vehicle  
46 body type in shaping behavioural intention to acquire electric vehicles: A multi-group

- 1 structural equation approach. *Transportation Research Part A: Policy and Practice*,  
2 116(May), 54–72. <https://doi.org/10.1016/j.tra.2018.05.011>
- 3 Morency, P., Plante, C., Dubé, A. S., Goudreau, S., Morency, C., Bourbonnais, P. L., Eluru, N.,  
4 Tétreault, L. F., Hatzopoulou, M., Iraganaboina, N. C., Bhowmik, T., & Smargiassi, A.  
5 (2020). The potential impacts of urban and transit planning scenarios for 2031 on car use and  
6 active transportation in a metropolitan area. *International Journal of Environmental Research  
7 and Public Health*, 17(14), 1–15. <https://doi.org/10.3390/ijerph17145061>
- 8 Mulholland, E., Tattini, J., Ramea, K., Yang, C., & Ó Gallachóir, B. P. (2018). The cost of  
9 electrifying private transport – Evidence from an empirical consumer choice model of Ireland  
10 and Denmark. *Transportation Research Part D: Transport and Environment*, 62, 584–603.  
11 <https://doi.org/10.1016/J.TRD.2018.04.010>
- 12 Nazari, F., Mohammadian, A. (Kouros), & Stephens, T. (2019). Modeling electric vehicle adoption  
13 considering a latent travel pattern construct and charging infrastructure. *Transportation  
14 Research Part D: Transport and Environment*, 72, 65–82.  
15 <https://doi.org/10.1016/J.TRD.2019.04.010>
- 16 Nerurkar, V., Barge, P., Bhatt, V., Rastogi, S., & Agarwal, B. (2023). Factors Influencing  
17 Consumer Decision to Purchase a Car. *Marketing and Management of Innovations*, 14(1),  
18 24–40. <https://doi.org/10.21272/mmi.2023.1-03>
- 19 Noel, L., Zarazua de Rubens, G., Kester, J., & Sovacool, B. K. (2020). Understanding the socio-  
20 technical nexus of Nordic electric vehicle (EV) barriers: A qualitative discussion of range,  
21 price, charging and knowledge. *Energy Policy*, 138(January).  
22 <https://doi.org/10.1016/j.enpol.2020.111292>
- 23 Okeke, F. O., Okosun, A. E., Udeh, C. A., & Okekeogbu, C. J. (2020). Cities for people: The  
24 dependency & impact of automobile in the life of city dwellers. *European Journal of  
25 Sustainable Development*, 9(3), 157–178. <https://doi.org/10.14207/ejsd.2020.v9n3p157>
- 26 Østli, V., Fridstrøm, L., Johansen, K. W., & Tseng, Y. Y. (2017). A generic discrete choice model  
27 of automobile purchase. *European Transport Research Review*, 9(2).  
28 <https://doi.org/10.1007/s12544-017-0232-1>
- 29 Pierce, G., & Connolly, R. (2023). Disparities in the “who” and “where” of the vehicle purchase  
30 decision-making process for lower-income households. *Travel Behaviour and Society*,  
31 31(December 2021), 363–373. <https://doi.org/10.1016/j.tbs.2023.02.003>
- 32 Qian, L., & Soopramanien, D. (2011). Heterogeneous consumer preferences for alternative fuel  
33 cars in China. *Transportation Research Part D: Transport and Environment*, 16(8), 607–613.  
34 <https://doi.org/10.1016/J.TRD.2011.08.005>
- 35 Raza, S., & Masmoudi, M. (2020). Consumer Vehicle Purchase Decision-making during COVID-  
36 19. *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*,  
37 692–696. <https://doi.org/10.1109/DASA51403.2020.9317187>
- 38 Ryu, D., Kim, C., & Lee, S. (2020). Development of GHG Emission Factor for Passenger Vehicles  
39 by Fuel Type and Size Class Using Statistical Analysis Method. *International Journal of  
40 Automotive Technology* 2020 21:5, 21(5), 1293–1302. [https://doi.org/10.1007/S12239-018-  
0122-8](https://doi.org/10.1007/S12239-018-<br/>41 0122-8)
- 42 Sabouri, S., Tian, G., Ewing, R., Park, K., & Greene, W. (2021). The built environment and vehicle  
43 ownership modeling: Evidence from 32 diverse regions in the U.S. *Journal of Transport  
44 Geography*, 93, 103073. <https://doi.org/10.1016/J.JTRANGEO.2021.103073>
- 45 Sheldon, T. L., & Dua, R. (2018). Gasoline savings from clean vehicle adoption. *Energy Policy*,  
46 120, 418–424. <https://doi.org/10.1016/J.ENPOL.2018.05.057>

- 1 Sheldon, T. L., & Dua, R. (2021). How responsive is Saudi new vehicle fleet fuel economy to fuel-  
2 and vehicle-price policy levers? *Energy Economics*, 97.  
3 <https://doi.org/10.1016/J.ENERCO.2020.105026>
- 4 Singh, V., Singh, V., & Vaibhav, S. (2020). A review and simple meta-analysis of factors  
5 influencing adoption of electric vehicles. *Transportation Research Part D: Transport and*  
6 *Environment*, 86(August). <https://doi.org/10.1016/j.trd.2020.102436>
- 7 Spissu, E., Pinjari, A. R., Pendyala, R. M., & Bhat, C. R. (2009). A copula-based joint multinomial  
8 discrete-continuous model of vehicle type choice and miles of travel. *Transportation*, 36(4),  
9 403–422. <https://doi.org/10.1007/S11116-009-9208-X/TABLES/2>
- 10 Stinson, M., Auld, J., & Mohammadian, A. K. (2020). Light duty vehicle choice models examining  
11 alternative fuel technology preferences among commercial fleet owners. *Transportation*  
12 *Research Procedia*, 46, 309–316. <https://doi.org/10.1016/j.trpro.2020.03.195>
- 13 USEPA. (2022). *Greenhouse Gas Emissions from a Typical Passenger Vehicle*.  
14 [https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-](https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle#typical-passenger)  
15 [vehicle#typical-passenger](https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle#typical-passenger)
- 16 van Huyssteen, N., & Rudansky-Kloppers, S. (2024). Factors influencing consumers' purchase  
17 decisions regarding personal motor vehicle insurance in South Africa. *Cogent Business and*  
18 *Management*, 11(1), 1–21. <https://doi.org/10.1080/23311975.2023.2293488>
- 19 Xing, J., Leard, B., & Li, S. (2021). What does an electric vehicle replace? *Journal of*  
20 *Environmental Economics and Management*, 107, 102432.  
21 <https://doi.org/10.1016/j.jeem.2021.102432>
- 22 Ye, J. M. and X. (2019). Modeling Household Vehicle Ownership in Emerging Economies.  
23 *Journal of the Indian Institution of Science*, 94(4), 647–671.
- 24 Zhao, C., Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., & Fertner, C. (2018). Cycling  
25 environmental perception in Beijing – A study of residents' attitudes towards future cycling  
26 and car purchasing. *Transport Policy*, 66, 96–106.  
27 <https://doi.org/10.1016/J.TRANPOL.2018.02.004>
- 28 Zhou, F., Zheng, Z., Whitehead, J., Perrons, R. K., Washington, S., & Page, L. (2020). Examining  
29 the impact of car-sharing on private vehicle ownership. *Transportation Research Part A:*  
30 *Policy and Practice*, 138, 322–341. <https://doi.org/10.1016/J.TRA.2020.06.003>
- 31