

1 **DEVELOPMENT OF A VEHICLE FLEET COMPOSITION MODEL SYSTEM:**  
2 **RESULTS FROM AN OPERATIONAL PROTOTYPE**

3  
4 **Venu M. Garikapati** (*corresponding author*)

5 Georgia Institute of Technology  
6 School of Civil and Environmental Engineering  
7 Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355.  
8 Tel: (480) 522-8067; Email: [venu.garikapati@gatech.edu](mailto:venu.garikapati@gatech.edu)

9  
10 **Daehyun You**

11 Georgia Institute of Technology  
12 School of Civil and Environmental Engineering  
13 Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355.  
14 Tel: (480) 626-3716; Email: [daehyun.you@gatech.edu](mailto:daehyun.you@gatech.edu)

15  
16 **Ram M. Pendyala**

17 Georgia Institute of Technology  
18 School of Civil and Environmental Engineering  
19 Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355  
20 Tel: 404-894-2201; Fax: 404-894-2278; Email: [ram.pendyala@ce.gatech.edu](mailto:ram.pendyala@ce.gatech.edu)

21  
22 **Kyunghwi Jeon**

23 Maricopa Association of Governments  
24 302 N First Avenue, Suite 300, Phoenix, AZ 85003.  
25 Tel: (602) 452-5079; Fax: (602) 254-6490; Email: [kjeon@azmag.gov](mailto:kjeon@azmag.gov)

26  
27 **Vladimir Livshits**

28 Maricopa Association of Governments  
29 302 N First Avenue, Suite 300, Phoenix, AZ 85003.  
30 Tel: (602) 452-5079; Fax: (602) 254-6490; Email: [vlivshits@azmag.gov](mailto:vlivshits@azmag.gov)

31  
32 **Chandra R. Bhat**

33 The University of Texas at Austin  
34 Department of Civil, Architectural and Environmental Engineering  
35 301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172  
36 Tel: (512) 471-4535, Fax: (512) 475-8744; Email: [bhat@mail.utexas.edu](mailto:bhat@mail.utexas.edu)

37  
38 *Submitted for Presentation and Publication*

39  
40 Word count: 6,706 text + 6 tables/figures x 250 = 8,206 words  
41 95<sup>th</sup> Annual Meeting of the Transportation Research Board

42  
43 Committee on Transportation Planning Applications (ADB50)

44  
45 August 2015

46

1 **ABSTRACT**

2 This paper presents the estimation and validation results of a vehicle fleet composition simulator  
3 that can be integrated with a larger activity-based microsimulation model system. The motivation  
4 behind the development of this fleet composition simulator is two-fold. First, it is desirable to  
5 predict the vehicle fleet mix to accurately quantify the emission profile in a region as vehicle  
6 technologies and fuel types evolve. This will provide planners the ability to evaluate the potential  
7 impacts of a host of emission control strategies. Second, knowledge of household vehicle fleet  
8 mix will enable modeling the ‘type’ of vehicle at the trip/tour level in existing activity based  
9 models (ABMs). This will not only add to the behavioral representation of travel in ABMs but  
10 also facilitate an accurate assessment of emission hotspots, and emissions along specific travel  
11 sheds. A heuristic algorithm is applied together with other model components to accurately  
12 predict the fleet mix of individual households where vehicle types are defined by body type and  
13 age. The model system performs well in replicating the base year fleet mix patterns for the  
14 Greater Phoenix metropolitan region, for which the model was developed.

15  
16 **Keywords:** *vehicle fleet composition simulator, vehicle fleet mix, vehicle ownership modeling,*  
17 *travel demand forecasting, activity-based modeling*

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

## 1. INTRODUCTION

Household vehicle fleet composition and its impact on personal travel has been an area of significant interest in the field of travel demand modeling. In the recent past, there have been significant advances in the activity based modeling arena in generating various attributes of a synthetic population and depicting how these characteristics (at both household/individual levels) impact activity-travel patterns. Increasing recognition is being paid to the role that household vehicle fleet composition plays that controls and constrains the travel behavior of a household. The vehicle fleet composition (mix of vehicle types owned by a household) influences and is influenced by household activity-travel patterns. In the absence of information about vehicle fleet composition, it is impossible to predict the ‘type’ of vehicle that an individual will choose for a particular tour/trip. For this reason, vehicle fleet composition modeling is gaining attention. Knowledge of the type of vehicle used for personal travel will help in the accurate estimation of energy consumption and vehicular emissions at the trip/tour level, and in turn at the sub-regional and regional level. A vehicle fleet composition model system, when integrated with an activity-based model, would provide planning authorities with an effective means to estimate the potential impacts of emission-reduction and alternative-fuel-promotion strategies.

Considerable progress has been made in the modeling of vehicle fleet composition and utilization in the recent past. This progress is primarily motivated by the desire to curb emissions from personal travel. The United States accounts for 16% of all GHG emissions in the world (World Research Institute, 2014). In the US, transportation accounted for 28% of greenhouse gas emissions in 2011 (EPA, 2014) and 70% of all petroleum consumption (EIA, 2013). A variety of policies may be invoked to achieve emission targets. For example, policies that discourage the use of highly polluting vehicles and promote the use of fuel-efficient/clean-fuel vehicles are in place in several countries including the United States. The challenge, however, is that it is difficult to predict the outcomes of such strategies as most travel demand forecasting models do not incorporate vehicle fleet composition model components sensitive to policy interventions. This has motivated the development of comprehensive vehicle fleet composition simulators.

Interest in analyzing and modeling household auto ownership patterns is not new in travel behavior research. Several early studies examined auto ownership in terms of the number of vehicles owned (Lerman and Ben-Akiva, 1976; Kain and Fauth, 1978; Golob and Burns, 1978). These and other studies that modeled vehicle ownership used multinomial logit (Manski and Sherman, 1980; Mannering and Winston, 1985) and nested logit specifications (Hocherman et al, 1983; Berkovec and Rust, 1985) to predict vehicle ownership patterns. These were followed by studies that modeled vehicle holdings at the household level (Kitamura et al, 2000). Brownstone et al (2000) studied household’s preferences for alternative fuel vehicles using a mixed logit specification that utilized both stated preference and revealed preference data from a survey conducted in California. Choo and Mokhtarian (2004) studied the effect of an individual’s travel attitudes, lifestyle, and mobility on vehicle type choice. The vehicle fleet owned by a household is influenced by the travel desires as well as composition of the household.

One of the key issues associated with modeling vehicle fleet mix is that household vehicle fleet composition is not a single discrete choice phenomenon, but a multiple discrete choice problem as households may choose and own multiple vehicles - acquiring, replacing, or discarding vehicles over time. To estimate a single discrete choice model of vehicle fleet mix with  $n$  vehicle alternatives, a total of  $2^n - 1$  combinations of alternatives exist in the choice set. The size of the choice set explodes as number of elemental alternatives increases. Also, traditional random utility maximization models do not consider the effect of satiation (variety-seeking

1 behavior) in the context of vehicle type choice. This is an important factor to be considered in  
2 fleet composition modeling as household often own multiple vehicle types and use them to  
3 varying levels.

4 Bhat (2005) formulated the Multiple Discrete Continuous Extreme Value (MDCEV)  
5 model to overcome the limitations of single discrete choice models. Bhat and Sen (2006)  
6 proposed an MDCEV-MNL model to model vehicle holdings and usage (annual mileage for  
7 each vehicle type). Bhat, et al (2009) enhanced the joint MDCEV-MNL model framework to  
8 include random coefficients/error components. Paleti, et al (2011) developed a comprehensive  
9 vehicle fleet composition, utilization, and evolution framework for integration in activity-based  
10 microsimulation models of travel demand. The framework consists of a vehicle selection module  
11 in which a joint discrete-continuous copula-based model predicts vehicle fleet composition and  
12 utilization.

13 Musti and Kockelman (2011) estimated a vehicle choice (MNL) model using data from a  
14 stated preference survey conducted in Austin, Texas. They also developed a microsimulator of  
15 vehicle transactions to simulate change of vehicle fleet composition over time. The base year  
16 fleet characteristics are treated as exogenous or given, and the simulator then evolves the  
17 household's fleet based on transaction and vehicle choice models. Pendyala, et al (2012) applied  
18 a socio-economic model system for activity-based modeling to the region of Southern California.  
19 A component of the socio-economic model system is CEMSELTS, which includes a fleet  
20 composition module that simulates the vehicle fleet owned by a household and assigns a primary  
21 driver to each of the vehicles owned. Although these simulators constitute promising  
22 developments in the field, the need for an operational comprehensive vehicle fleet composition  
23 simulator remains.

24 The current research effort intends to add to the existing body of work by developing an  
25 operational prototype of a model system that predicts the vehicle fleet mix of households, where  
26 vehicles are classified by body type and age. The system also predicts the mileage allocated to  
27 each of the vehicles owned by the household. The system comprises several components that  
28 together predict fleet mix for a population of synthetic households generated within an activity-  
29 based travel model system. The modeling framework and preliminary results have been  
30 documented in an earlier paper (You et al, 2014). The previous paper focused on the multiple  
31 discrete continuous extreme value (MDCEV) model of vehicle fleet composition and presented a  
32 scenario analysis that portrayed the policy sensitivity of the model in predicting changes in fleet  
33 mix in response to varying zonal accessibility measures. This paper presents a detailed  
34 description of an operational prototype of the vehicle fleet composition model system. In  
35 particular, the paper describes a heuristic mileage reallocation (HMR) algorithm that is  
36 developed as part of the fleet mix model system. The model system is developed and  
37 demonstrated using the Greater Phoenix (Maricopa County) metropolitan region in the United  
38 States as a testbed.

39 The remainder of this paper is organized as follows. The next section discusses the data  
40 preparation and presents the descriptive statistics of data used for estimating various components  
41 of the model system. This is followed by a detailed account of components in the model system.  
42 The fourth section presents model estimation results and the performance of each of the model  
43 components in a sample replication setting. The final section presents conclusions and proposes  
44 avenues for future research.

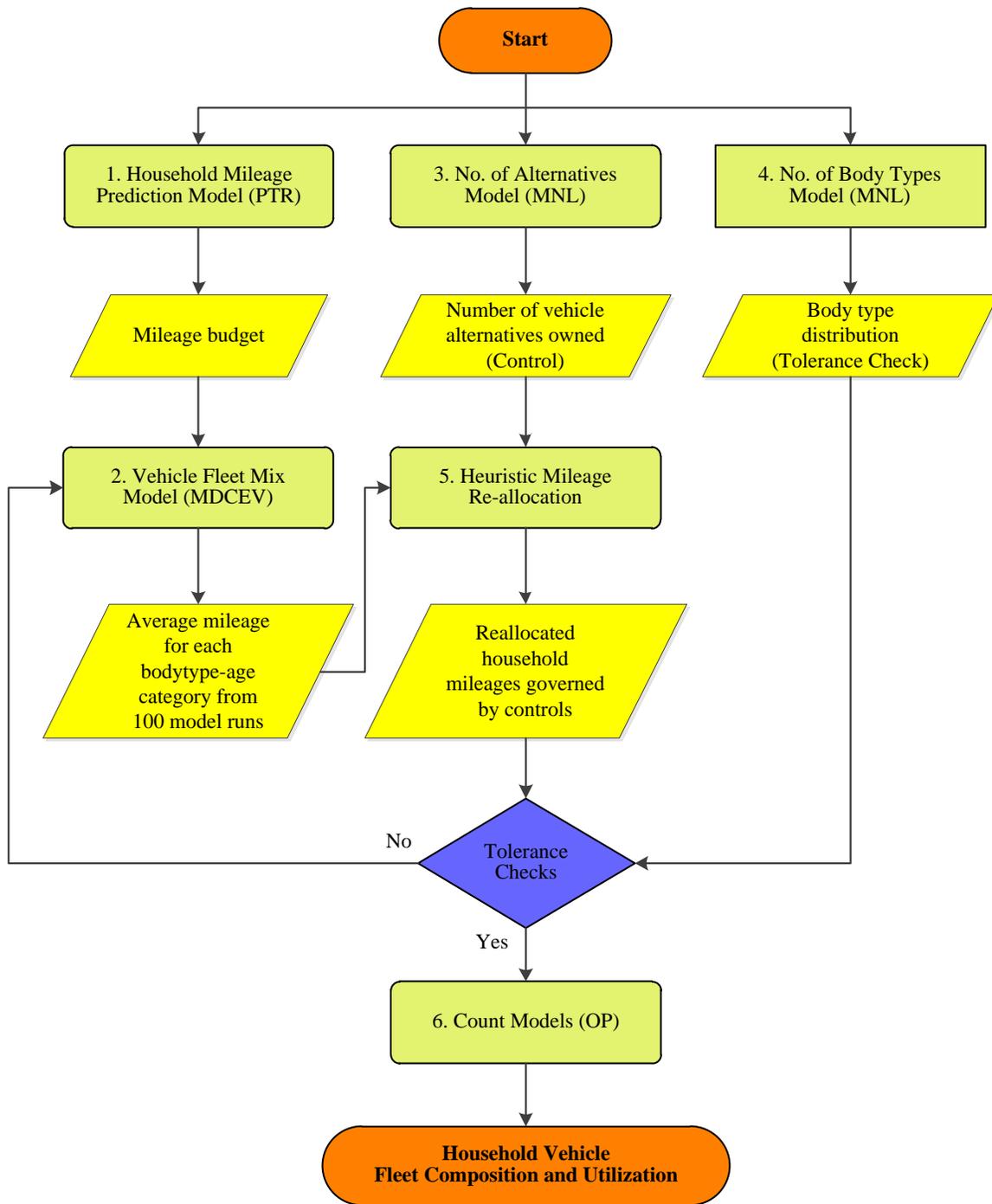
1 **2. DATA**

2 The data used in this model development effort is from the 2008-2009 National Household  
3 Travel Survey (NHTS). NHTS collects data on socio-economic, demographic, vehicle  
4 ownership, and travel characteristics of households in the nation. Detailed information pertaining  
5 to the vehicles owned by a household are available in the dataset. The Maricopa Association of  
6 Governments (MAG), the metropolitan planning organization (MPO) for the Greater Phoenix  
7 metropolitan region (essentially comprising of Maricopa County), purchased an add-on sample  
8 for the development and update of travel demand models for the region. The survey data set was  
9 cleaned and processed for the development of the vehicle fleet composition model components.  
10 The characteristics of the survey sample are provided in You, et al (2014) and not included here  
11 in the interest of brevity. The final estimation dataset consisted of 4,262 households owning a  
12 total of 7,785 vehicles. The vehicle characteristics of the data set are also presented in You, et al  
13 (2014).

14 Households own, on average, two vehicles, with the ratio of number of vehicles to  
15 number of drivers close to unity. For the current effort, a total of 14 vehicle alternatives are  
16 considered. Four body types: car, van, SUV, and pick-up truck, sub-classified by 3 vintage  
17 categories: 0-5 years, 6-11 years and 12 years or older, are considered. Motorbike is an  
18 additional alternative with no vintage classification. Finally, an alternative called the “non-  
19 motorized vehicle” is included; this alternative accounts for all of the walk/bike travel  
20 undertaken by a household and is always assumed to be chosen or consumed (at least to a  
21 minimal degree). The vehicle dataset of the NHTS is reorganized to create a vehicle profile for  
22 each household; the final dataset has information on the body type and age of all vehicles owned  
23 by the household. Household demographics, network accessibility measures, and zonal  
24 characteristics are then appended to this dataset. Accessibility measures such as percent of  
25 regional employment accessible within 10 minutes and 30 minutes by auto or transit are  
26 developed from network travel time data and appended to the dataset. The comprehensive dataset  
27 thus formed was used to estimate all of the model components in the vehicle fleet composition  
28 model system.

29  
30 **3. FRAMEWORK FOR VEHICLE FLEET COMPOSITION MODELING**

31 The methodological framework of the vehicle fleet composition model system is shown in  
32 Figure 1. The first element in the model system is a household mileage prediction model that  
33 predicts the annual motorized mileage consumption of households. This step is necessary as the  
34 subsequent component in the model system, the MDCEV model of vehicle fleet mix, requires a  
35 mileage budget that must be allocated across vehicle types chosen by a household. Motorized  
36 mileage is estimated using a non-linear regression (power form) model. Once the motorized  
37 mileage for each household is predicted, non-motorized mileage is computed using a preset  
38 formula ( $0.5 \text{ mile/person/day} \times \text{household size} \times 365 \text{ days/year}$ ) as every household will  
39 inevitably undertake some amount of non-zero mileage consumption in this category (say,  
40 walking to and from a parking location). The combined annual mileage is given as input to the  
41 MDCEV model, which will predict the vehicle fleet mix owned by the household and allocate  
42 this mileage budget to all of the vehicle types chosen by the household. Any vehicle type with a  
43 zero mileage allocation is not chosen into the fleet; only those alternatives with non-zero mileage  
44 constitute the vehicle fleet mix.



1

2

PTR: Power Transformed Regression

3

MNL: Multinomial Logit

4

MDCEV: Multiple Discrete Continuous Extreme Value

5

OP: Ordered Probit

6

7

**FIGURE 1 Vehicle fleet composition model framework**

8

1 The MDCEV model is estimated and applied in such a fashion that every household (in a  
2 synthetic population) will consume at least some non-motorized mileage. The MDCEV model,  
3 when applied in forecasting mode, gives a different output (prediction) each time a simulation is  
4 run. Which simulation result should be considered final? Will the output of the MDCEV model  
5 accurately depict the body type distribution in the population? To address these questions, a  
6 methodology was proposed previously (You et al, 2014) in which a model that separately  
7 predicts the distribution of the number of vehicle body types in the population is used. The idea  
8 behind the proposed methodology was that the MDCEV model would be run repeatedly until the  
9 output of the MDCEV model matches the frequency distribution of the number of vehicle body  
10 types predicted by the separate multinomial logit (MNL) model.

11 It was observed that using a single simulation of the MDCEV model was not enough to  
12 accurately predict the distribution of the number of body types predicted by the MNL model  
13 within set tolerance limits. While the MDCEV model predicts vehicle fleet mix quite well, the  
14 implied body type distribution from the MDCEV model result almost always over predicted the  
15 proportion of households owning a single vehicle body type and under predicted all of the other  
16 categories. To overcome this issue, the MDCEV model simulation may be run repeatedly and  
17 mileage consumptions from each simulation stored. After a number (say 100) of simulation  
18 applications of the MDCEV model, the average mileage consumption across all alternatives can  
19 be computed for each household. As the averaging process yields a non-zero mileage in many  
20 alternatives, a Heuristic Mileage Reallocation algorithm was developed and implemented in the  
21 vehicle fleet composition model system.

22 Because each simulation run of the MDCEV model gives a slightly different result, the  
23 average mileage consumption result from repeated MDCEV model runs is likely to indicate that  
24 a household owns (consumes) almost all of the vehicle categories. In reality, a household might  
25 own only a subset of the vehicle categories in the choice set. The HMR algorithm reallocates the  
26 mileage budget across alternatives for each household in such a fashion that, in the aggregate, the  
27 model replicates the distribution of the number of body types in the population predicted by the  
28 separate multinomial logit model. This multinomial logit model predicts the number of distinct  
29 body types owned by a household. For example, if a household owns a car 0-5 years old, a car 6-  
30 11 years old, and a van 6-11 years old, the number of distinct body types owned by the  
31 household is 2 (car and van). While the vehicle body type distribution for the dataset is known in  
32 the base year (from survey data), this distribution is unknown for any future year (or scenario).  
33 The MNL model of number of body types predicts this distribution for any future year scenario.  
34 The resulting distribution serves as a control that the HMR algorithm strives to match.

35 The mileage reallocation algorithm needs (as input) information on the number of distinct  
36 categories of vehicle alternatives owned by each household in the synthetic population. For this  
37 purpose, the vehicle fleet composition modeling framework includes an MNL model of the  
38 number of vehicle alternatives. If a household owns a car 0-5 years old, a car 6-11 years old, and  
39 a van 6-11 years old, the number of vehicle alternatives owned by this household is three. Thus,  
40 the HMR algorithm takes the outputs of the MDCEV model and the MNL model of number of  
41 vehicle alternatives as inputs. The logic followed by HMR algorithm is shown in Figure 2. The  
42 output of the MNL model of number of vehicle alternatives provides information on the number  
43 of vehicle types a household owns. Based on the output of the MDCEV model, a cumulative  
44 mileage distribution is computed for each household. Based on a Monte Carlo random draw, a  
45 vehicle alternative is selected from the cumulative distribution as 'owned' by the household. The  
46 chosen alternative is removed from the dataset, thereby eliminating the choice of the same

1 alternative multiple times. This process is carried out ‘ $k$ ’ times, where  $k$  is the number of vehicle  
2 alternatives predicted by the separate multinomial logit model. Figure 3 presents a schematic  
3 illustrating the application of the Heuristic Mileage Reallocation (HMR) algorithm. In the  
4 sample illustration, a household is predicted to own three vehicle alternatives. Therefore, three  
5 random draws (without replacement) are performed to identify the specific alternatives chosen  
6 by a household. The mileage budget is scaled appropriately to ensure that the final mileage total  
7 matches the original mileage budget predicted for the household.

8         Once the HMR algorithm reallocates the mileage across alternatives for all households,  
9 the implied aggregate vehicle body type distribution output by this step is compared against the  
10 body type distribution predicted by the MNL model of number of body types. The absolute  
11 percent difference between the distributions is computed and checked against a user defined  
12 tolerance limit (say 3%). If the HMR algorithm passes the tolerance check, the output from this  
13 step is accepted and used as input to a series of ordered probit count models. If not, the entire  
14 application is repeated after calibrating the model components as necessary. This process is  
15 carried out repeatedly until the percent difference between the two distributions satisfies the  
16 tolerance criterion.

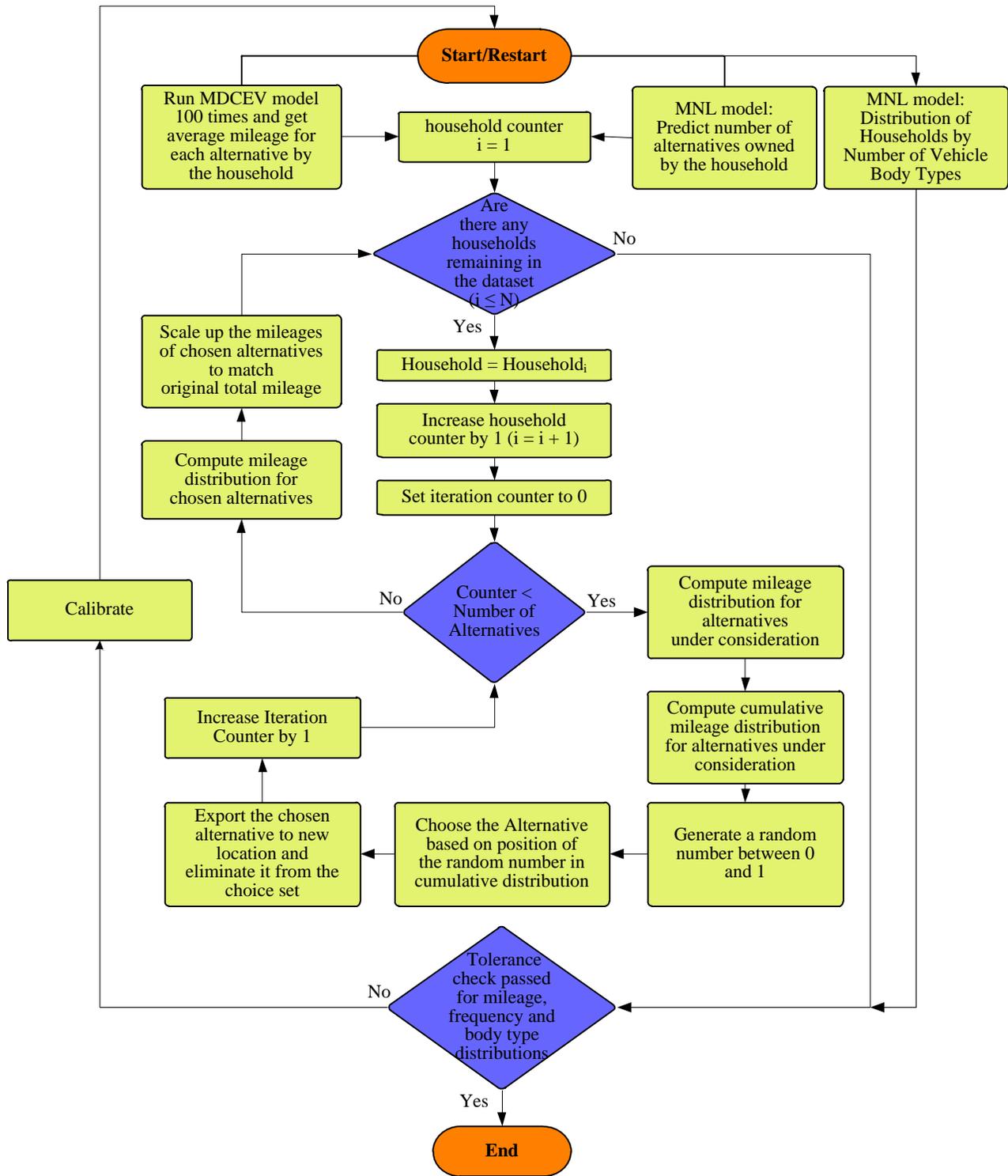
17         The ordered probit count models determine whether all of the mileage consumed by a  
18 household in a particular vehicle alternative belongs to one or multiple vehicles. The count  
19 models are necessary as the vintage classifications are aggregated into three categories for ease  
20 of estimation and application of the fleet composition model system. For example, suppose the  
21 output of the HMR algorithm indicates that a household uses the car 0-5 years old alternative to  
22 travel 25000 miles annually. The count model determines whether all of this mileage is  
23 consumed using just one car 0-5 years old, or if the household owns multiple cars in the 0-5 year  
24 category. Ideally, a count model should be estimated for each of the 13 different vehicle  
25 categories defined for the MDCEV model, but this will increase the number of individual  
26 components in the model system while also decreasing the data available to estimate each of the  
27 individual count models. So, it was felt prudent to estimate one count model for each of the body  
28 types, with vintage serving as an explanatory variable in the models. If the household has non-  
29 zero mileage consumption in any of the vintages of a particular body type, the count model of  
30 that body type is applied to find out the ‘actual’ count of vehicles of that body type-age  
31 combination in that household.

32         At the end of application of the entire model system, the fleet composition of the  
33 household including body type, age and count of vehicles of each body type-age category is  
34 known along with their respective annual mileage consumption. A software package has been  
35 developed in the open source coding language ‘R’, to both estimate and apply the vehicle fleet  
36 composition model system.

#### 38 **4. MODEL ESTIMATION AND APPLICATION RESULTS**

39 Each of the model components in the vehicle fleet composition model system was estimated and  
40 then tested in isolation to check prediction accuracy. Each individual component of the model  
41 system performed quite well in isolation. The model system was then applied sequentially on the  
42 entire estimation dataset to see how well it would be able to ‘replicate’ the observed vehicle fleet  
43 mix composition for the survey sample. It should be noted that this is not a strict validation  
44 exercise where a subset of the data is set aside for validation purpose, but rather a replication of  
45 observed patterns in the survey data. A separate holdout sample could not be accommodated due  
46 to the need to utilize the full sample (size) for model estimation.

1



2  
3  
4  
5

**FIGURE 2 Heuristic mileage reallocation algorithm**

Output from 100 MDCEV Simulation Runs

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van 6_11	Van ≥12	Motor	Total
5000	8000	400	200	500	100	1100	1000	300	2000	7000	550	50	26200

Input to HMR Algorithm

Compute Cumulative Mileage Distribution

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van 6_11	Van ≥12	Motor	Total
5000	8000	400	200	500	100	1100	1000	300	2000	7000	550	50	26200
19.1%	30.5%	1.5%	0.8%	1.9%	0.4%	4.2%	3.8%	1.1%	7.6%	26.7%	2.1%	0.2%	100.0%
19.1%	49.6%	51.1%	51.9%	53.8%	54.2%	58.4%	62.2%	63.4%	71.0%	97.7%	99.8%	100.0%	-

Random Draw Without Replacement

Number of Vehicle Alternatives: 3 (Predicted by the MNL Model)

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van 6_11	Van ≥12	Motor	Total
5000	8000	400	200	500	100	1100	1000	300	2000	7000	550	50	26200
19.1%	30.5%	1.5%	0.8%	1.9%	0.4%	4.2%	3.8%	1.1%	7.6%	26.7%	2.1%	0.2%	100.0%
19.1%	49.6%	51.1%	51.9%	53.8%	54.2%	58.4%	62.2%	63.4%	71.0%	97.7%	99.8%	100.0%	-

Update Cumulative Mileage Distribution

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van ≥12	Motor	Total
5000	8000	400	200	500	100	1100	1000	300	2000	550	50	19200
26.0%	41.7%	2.1%	1.0%	2.6%	0.5%	5.7%	5.2%	1.6%	10.4%	2.9%	0.3%	100.0%
26.0%	67.7%	69.8%	70.8%	73.4%	74.0%	79.7%	84.9%	86.5%	96.9%	99.7%	100.0%	-

Update Cumulative Mileage Distribution

Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van ≥12	Motor	Total
8000	400	200	500	100	1100	1000	300	2000	550	50	14200
56.3%	2.8%	1.4%	3.5%	0.7%	7.7%	7.0%	2.1%	14.1%	3.9%	0.4%	100.0%
56.3%	59.2%	60.6%	64.1%	64.8%	72.5%	79.6%	81.7%	95.8%	99.6%	100.0%	-

Final Vehicle Type Choice Selection

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van 6_11	Van ≥12	Motor	Total
5000	0	0	0	0	0	1100	0	0	0	7000	0	0	13100
38.2%	0.0%	0.0%	0.0%	0.0%	0.0%	8.4%	0.0%	0.0%	0.0%	53.4%	0.0%	0.0%	100%

Scale up the Mileage of Chosen Alternatives to Match the Original Total Mileage

Car 0_5	Car 6_11	Car ≥12	Pick 0_5	Pick 6_11	Pick ≥12	SUV 0_5	SUV 6_11	SUV ≥12	Van 0_5	Van 6_11	Van ≥12	Motor	Total
10000	0	0	0	0	0	2200	0	0	0	14000	0	0	26200

FIGURE 3 Illustration of heuristic mileage reallocation algorithm

1  
2  
3

1 It is not possible to furnish comprehensive estimation results within the scope of this  
2 paper. For the sake of brevity, only selected illustrative estimation results are furnished in the  
3 paper. However, the estimation and test application of each of the model components is briefly  
4 described in this section.

#### 5 6 **4.1 Mileage Prediction Model**

7 Several alternative model forms were explored to model the annual motorized mileage consumed  
8 by households. A non-linear (power) regression model fit the data best. The estimation results of  
9 the mileage prediction model are omitted in the interest of brevity. The model included a host of  
10 socio-economic, demographic, and residential location and accessibility variables as explanatory  
11 factors. From the model estimation results, it was observed that households in the highest  
12 income category show a higher motorized mileage consumption, while households in the lowest  
13 income category are likely to have lower consumptions. Households residing in traffic analysis  
14 zones (TAZs) with greater accessibility are likely to have lower motorized mileage consumption.  
15 Lund (2003) presents a similar result and found that increased accessibility enhances pedestrian  
16 travel in a neighborhood. Results of the replication exercise are shown in Panel A of Figure 4.  
17 The results shown are for a calibrated model, where a slight adjustment was made to the constant  
18 in the mileage prediction equation to match the observed distribution more closely. It can be seen  
19 that the mileage prediction model replicates the observed mileage distribution quite well.

#### 20 21 **4.2 MDCEV Model of Fleet Mix**

22 A detailed discussion of estimation/replication results of the MDCEV model along with results  
23 of exercises to test the sensitivity of the model to varying accessibility measures were presented  
24 in an earlier paper (You et al, 2014). The MDCEV model forecasting procedure was  
25 implemented using the code developed by Pinjari and Bhat (2011). The forecasting code was  
26 translated from GAUSS matrix programming language to open source statistical programming  
27 language 'R' for ease of integration into the fleet composition model system. The MDCEV  
28 model takes mileage budget as input and predicts the fleet mix of a household. It was observed  
29 that a single instance of the MDCEV model predicts the fleet composition of the dataset quite  
30 well, but it tended to over predict the ownership of a single vehicle body type and under predict  
31 ownership of multiple body types. While extensive calibration of the model would address this  
32 issue to a certain extent, it was felt prudent to search for alternative approaches to address this.  
33 Several alternative approaches were explored and the HMR algorithm presented in this paper  
34 provided the best predictions of fleet composition patterns observed in the dataset. The HMR  
35 algorithm requires three essential inputs: i) average mileage consumption allocation predicted by  
36 the MDCEV model, ii) number of vehicle alternatives owned by a household, and iii) the  
37 aggregate distribution of the number of vehicle body types in the population.

#### 38 39 **4.3 MNL Model of Number of Vehicle Alternatives**

40 The MDCEV fleet mix model has a total of 13 motorized alternatives (4 body types x 3 vintage  
41 categories + motorbike). The MNL model of number of vehicle alternatives predicts the number  
42 of distinct vehicle categories owned by a household. Ideally the model should allow for a total of  
43 13 alternatives, but observations from data revealed that the maximum number of distinct  
44 alternatives owned by any household in the dataset is five. Therefore, an MNL model is  
45 estimated with 6 categories (0, 1, 2, 3, 4, and  $\geq 5$  vehicle alternatives) and the final category ( $\geq 5$   
46 vehicles) is considered as the base alternative. The purpose of this model is to provide a control

1 total to the HMR algorithm in order to reallocate the mileage consumption distribution obtained  
2 by averaging the results from 100 MDCEV model simulations. Model estimation results for this  
3 component are presented in Panel A of Table 1.

4 Estimation results show that lowest income households are likely to own fewer vehicle  
5 alternatives, while medium and high income households tend to own multiple vehicle  
6 alternatives, as expected. Increasing income and presence of higher number of workers in the  
7 household positively influence ownership of multiple vehicle alternatives. Presence of children is  
8 found to positively influence owning multiple vehicle alternatives. Households with children  
9 may choose to own multiple vehicle types to accommodate disparate travel needs and patterns.  
10 The sample replication results of the MNL model of number of vehicle alternatives is presented  
11 in Panel B of Figure 4. The results presented are for the uncalibrated version of the model and it  
12 can be seen that the predicted distributions match the observed patterns quite well. This suggests  
13 that the predictions from this model constitute an appropriate input to reallocate the average  
14 mileage distribution predicted by the MDCEV model using the HMR algorithm.

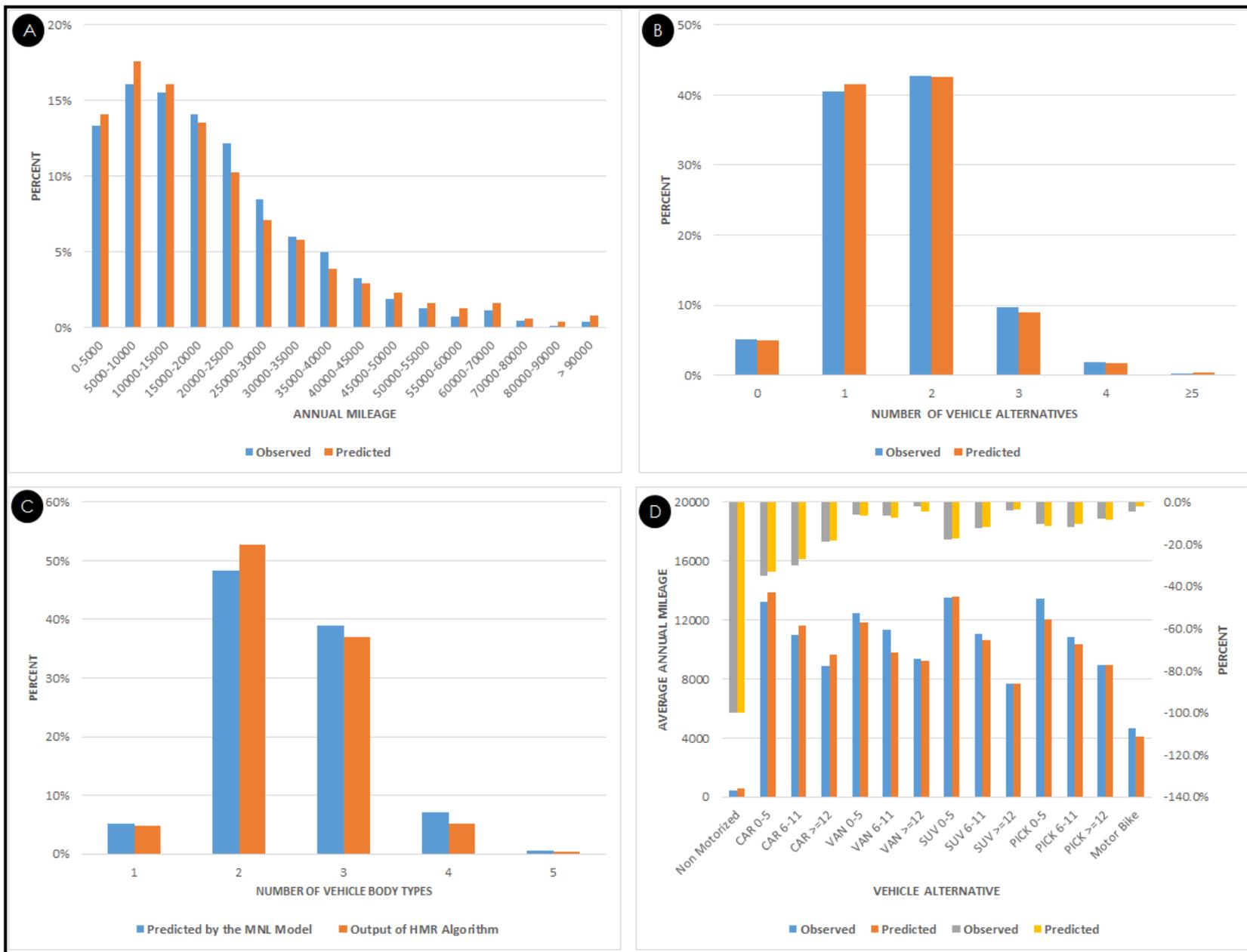
#### 15 16 **4.4 MNL Model of Number of Body Types**

17 This model predicts the number of distinct body types owned by a household. Including  
18 motorbike, there are a total of five distinct body types (car, van, sport utility vehicle, pickup  
19 truck, and motorbike). Only one household in the data set was observed to own vehicles of all  
20 five body types. As this observation may be treated as an outlier, the MNL model specification  
21 considered a maximum of four vehicle body types. Model estimation results for this component  
22 are presented in Panel B of Table 1. Model estimation results indicate that lowest and low  
23 income households are most likely to own zero or one vehicle body type. This finding is  
24 consistent with the results of the MNL model of number of alternatives. Households in this  
25 income range appeared in the zero and one vehicle alternative category, which automatically  
26 places them in the equivalent category in the vehicle body type model. Households with more  
27 adults and drivers are likely to own multiple vehicle body types. This speaks to the variety  
28 seeking nature of different individuals in such households. The presence of children is found to  
29 negatively influence owning a single vehicle category exclusively. For such households, the need  
30 to chauffeur children is likely to necessitate owning a bigger vehicle (such as a van or SUV) in  
31 addition to owning another smaller vehicle body type for commuting travel. This is further  
32 corroborated by the significance of the same variable in the three vehicle body type category.  
33 Results of the standalone replication of this model are not presented in the interest of brevity.  
34 The uncalibrated model performed exceedingly well in predicting the distribution of the number  
35 of body types in the survey sample. An enhancement of the structure (currently under  
36 development) combines the estimation of number of body types with the number of alternatives  
37 using a nested logit model structure, to reflect the significant correlation that likely exists  
38 between these two dimensions of car ownership.

#### 39 40 **4.5 Heuristic Mileage Reallocation (HMR) Algorithm**

41 This component of the model system reallocates average mileage predicted by the MDCEV  
42 algorithm across choice alternatives using a heuristic approach. The logical sequence of HMR  
43 algorithm is depicted in Figure 2 and a sample illustration of the process is depicted in Figure 3.  
44 After the HMR algorithm is applied for a particular household, the model system predicts a  
45 selection of ‘*n*’ vehicle alternatives.





1  
2  
3  
**FIGURE 4 Model system replication results - part I: (A) mileage prediction model, (B) MNL model of number of alternatives, (C) MNL model of number of body types, and (D) MDCEV model of vehicle fleet mix**

1           Aggregate mileage from these alternatives will only account to a portion of the total  
2 annual motorized mileage consumption of the household. In order to match the total motorized  
3 mileage for the household, mileages of the selected alternatives are scaled up proportionally.  
4 Mileage reallocation is done for every household in the dataset using this algorithm. In the end,  
5 the following distributions from the output of the HMR algorithm are checked against observed  
6 distributions for the base year.

7  
8 *Body type distribution:* The predicted body type distribution is derived from the output of the  
9 HMR algorithm and compared against the distribution predicted by the MNL model of number  
10 of body types. If both of the distributions match within a set tolerance criterion, the HMR  
11 algorithm stops and the output of the algorithm serves as input to subsequent components in the  
12 model system. If the distributions do not match within the tolerance limit, the HMR algorithm is  
13 run again and this process is repeated a number of times (user defined) to see if the HMR  
14 algorithm outputs the expected distribution as predicted by the MNL model of the number of  
15 body types. If it does not, the MDCEV model is re-calibrated and the process is repeated. The  
16 calibration exercise is done with due caution to avoid any unintended consequences that such  
17 changes might bring. In the current empirical context, the body type distribution derived from the  
18 output of the HMR algorithm matches quite well with the distribution of the number of body  
19 types predicted by the corresponding MNL model (shown in Panel C, Figure 4).

20  
21 *Average annual mileage distribution:* Predicted and observed annual mileage distributions are  
22 compared. Only households owning a particular vehicle alternative are considered in the  
23 computation of average annual mileage values for that alternative. The output of the HMR  
24 algorithm is found to match the observed patterns of annual average mileage distribution quite  
25 well. Results presented (bottom axis, Panel D, Figure 4) are for a slightly calibrated version of  
26 the MDCEV model presented in an earlier paper (You et al, 2014).

27  
28 *Vehicle frequency distribution:* The percentage ownership for each vehicle alternative is  
29 predicted and compared against the observed ownership pattern (top axis, Panel D, Figure 4). It  
30 can be seen that the model system predicts the vehicle frequency distribution quite well.

31  
32           The output of the HMR algorithm that matches the control distributions on all three measures  
33 discussed above serves as input to the count models.

#### 34 35 **4.6 Count Models**

36 Count models are necessary due to the aggregate categorical treatment of vehicle age (0-5 years,  
37 6-12 years, 12 years or more) in the MDCEV model. If a household owns multiple vehicles of  
38 the same body type-vintage combination, the MDCEV model is not able to return the number of  
39 vehicles owned by a household within that alternative. The MDCEV model is only able to  
40 indicate that the household owns that vehicle type without providing any indication of the count  
41 of vehicles owned within the specific vehicle type chosen. Count models are applied for all  
42 vehicle alternatives that have non-zero mileage consumptions for a household. Ordered probit  
43 count models are estimated for car, van, SUV and pick-up truck body types. The maximum  
44 number of cars that a household may own is set at three in the current model specification due to  
45 data limitations. Similarly, the maximum number of vans, SUVs, and pick-up trucks that a  
46 household may own is set at two (again, due to data limitations). These values are set based on

1 distributions of vehicle counts of different types in the estimation dataset. Motorbike category  
2 does not have a corresponding count model as the number of households owing multiple  
3 motorbikes is found to be very low.

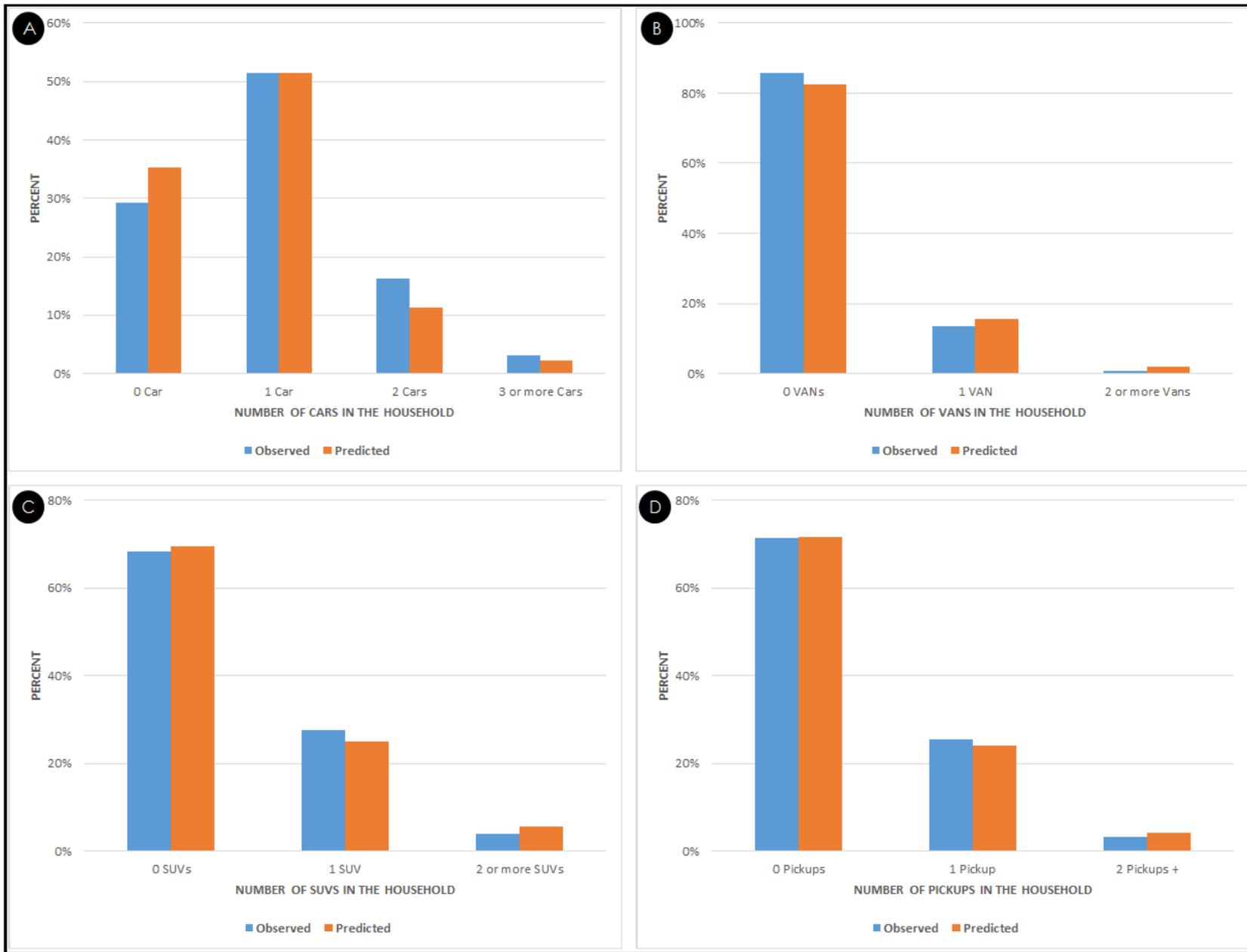
4 Estimation results for the count models are not furnished in the interest of brevity. In  
5 general, various household and zonal level characteristics were used in the count model  
6 specification. Annual mileage consumption for a particular body type was also used as an  
7 explanatory variable in these models. Low income households were found to own fewer vehicles  
8 of any body type, a result consistent with expectations. High income households had a greater  
9 propensity to own multiple cars and SUVs than other vehicle body types. Very small (household  
10 size = 1) and very large households (household size  $\geq 4$ ) tended to have fewer number of cars in  
11 their fleet. Both of these findings make intuitive sense in that smaller households might not own  
12 multiple vehicles at all, while larger households might choose to own a mix of vehicles (such as  
13 a car and a van), rather than multiple vehicles of a smaller size (such as cars). Households living  
14 in TAZs with low density (rural/suburban neighborhoods) tended to own multiple vans and pick-  
15 up trucks. Higher mileage consumption in a specific vehicle body type category was, as  
16 expected, associated with the ownership of multiple vehicles within that body type category.  
17 This feature of the model ensures that any large mileage allocation to a specific body type will be  
18 distributed appropriately across multiple vehicles.

19 Appropriate count models are applied whenever a household is simulated to have chosen  
20 a specific vehicle body type (from earlier components of the vehicle fleet composition model  
21 system). Results comparing observed and predicted counts for each vehicle body type are shown  
22 in Figure 5. The results from the count models constitute a test of the efficacy of the entire  
23 vehicle fleet composition model system (as this is a sequential application process). Results from  
24 the count model predictions show that the fleet composition model system as a whole is able to  
25 accurately predict the vehicle fleet mix, annual mileage consumption, as well as count of  
26 multiple vehicles of the same body type-age category.

## 27 28 **5. CONCLUSIONS**

29 Concerns about air quality and energy security, and the desire to promote sustainable  
30 communities and cities motivates the accurate prediction of emissions and energy consumption  
31 resulting from transportation-related activity. Traditional travel demand modeling systems often  
32 include information about household vehicle ownership and miles of travel, but do not include  
33 information about the types of vehicles owned by households and the extent to which the  
34 different types of vehicles are driven (utilized). As a result, estimates of energy consumption  
35 and emissions do not necessarily reflect the true distribution of vehicle types in the population  
36 and their utilization patterns. In addition, the advent of a variety of new vehicle technologies and  
37 fuel types hold considerable promise in reducing the energy and emissions footprint of  
38 household travel. However, in the absence of a model that forecasts household vehicle holdings  
39 by type under alternative technology, fuel, and policy scenarios, it is virtually impossible to  
40 accurately predict energy and emissions under alternative futures.

41 This paper describes the structure and algorithms underlying a comprehensive vehicle  
42 fleet composition model system that can be used to simulate the mix of vehicle types that a  
43 household owns, and estimate the extent to which each vehicle in the fleet would be utilized  
44 (driven). Such a model system can be effectively integrated in regional travel model systems to  
45 better estimate energy and emissions in specific locations, along specific corridors, and in  
46 subareas of interest.



**FIGURE 5 VFC model system replication results - part II: (A) car count model, (B) van count model, (C) SUV count model, and (D) pick-up count model**

1  
2  
3

The model system is estimated and tested for the Greater Phoenix metropolitan region in the United States. The model system includes a multiple discrete continuous extreme value (MDCEV) model of vehicle fleet mix that is capable of estimating the alternative vehicle types or alternatives that a household would own. This model is combined with other model components, and applied in forecasting mode using a practical heuristic algorithm, to predict the complete vehicle fleet composition details of a household. The components of the model system are estimated on a sample of 4,262 households belonging to the Greater Phoenix region in the National Household Travel Survey data set. The model efficacy is tested by applying the model to the same estimation data set and comparing the model predictions to the observed vehicle fleet composition patterns in the data set. Although not a true validation of the model system, the quality of the sample replication is very good, suggesting that the vehicle fleet composition model system may be suitable for integration in regional travel model systems. As activity-based travel models are increasingly deployed in various metropolitan areas around the world, vehicle fleet composition models such as that described in this paper can be integrated effectively to better estimate household vehicle ownership by vehicle type and assign a specific vehicle in the household to different drivers, and various trips and tours. By tracking individual vehicular movements through time and space, it is possible to obtain very accurate estimates of energy consumption and emissions.

Ongoing efforts are focused on adding a fleet evolution module to address vehicle ownership dynamics with replacement/addition of vehicles in a household's vehicle fleet over time. This could not be done in the current effort as the data used for model development is cross sectional in nature and contains no information about vehicle acquisition/disposal/replacement. The current model system does not account for fuel type in the vehicle type classification scheme; vehicle types are purely defined by body type and vintage. The inclusion of fuel type was not possible due to the very limited number of observations for alternative fuel vehicles in the data set. However, the inclusion of fuel type in the vehicle classification scheme would help planning agencies test policy scenarios pertaining to hybrid and electric vehicles. The model system is also currently undergoing a thorough validation process using a variety of secondary data sources, including Motor Vehicle Division records on vehicle registrations.

## REFERENCES

- Berkovec, J., and J. Rust. A Nested Logit Model of Automobile Holdings for One Vehicle Households. *Transportation Research Part B*, Vol. 19, No. 4, 1985, pp. 275-285.
- Bhat, C. R. A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B*, Vol. 39, No. 8, 2005, pp. 679-707.
- Bhat, C. R., and S. Sen. Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-Continuous Extreme Value (MDCEV) Model. *Transportation Research Part B*, Vol. 40, No. 1, 2006, pp. 35-53.
- Bhat, C. R., S. Sen, and N. Eluru. The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. *Transportation Research Part B*, Vol. 43, No. 1, 2009, pp. 1-18.
- Brownstone, D., D. S. Bunch, and K. Train. Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles. *Transportation Research Part B*, Vol. 34, No. 5, 2000, pp. 315-338.

- Choo, S., and P. L. Mokhtarian. What Type of Vehicle do People Drive? The Role of Attitude and Lifestyle in Influencing Vehicle Type Choice. *Transportation Research Part A*, Vol. 38, No. 3, 2004, pp. 201-222.
- EIA, *Oil: Crude and Petroleum Products Explained*. November 5, 2013. [http://www.eia.gov/energyexplained/index.cfm?page=oil\\_home#tab2](http://www.eia.gov/energyexplained/index.cfm?page=oil_home#tab2). Accessed July 24, 2014.
- EPA, *Sources of Greenhouse Gas Emissions*. July 22, 2014. <http://www.epa.gov/climatechange/ghgemissions/sources/transportation.html>. Accessed July 24, 2014.
- Golob, T. F., and L. D. Burns. Effects of Transportation Service on Automobile Ownership in an Urban Area. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 673, Transportation Research Board of the National Academies, Washington, D.C., 1978, pp. 137-145.
- Hocherman, I., J. N. Prashker, and M. E. Ben-Akiva. Estimation and Use of Dynamic Transaction Models of Automobile Ownership. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 944, Transportation Research Board of the National Academies, Washington, D.C., 1983, pp. 134-141.
- Kain, J. F., and G. R. Fauth. The Impact of Urban Development on Auto Ownership and Transit Use. *Real Estate Economics*, Vol. 6, No. 3, 1978, pp. 305-326.
- Kitamura, R., T. F. Golob, T. Yamamoto, and G. Wu. Accessibility and Auto Use in a Motorized Metropolis. Presented at the *79th Transportation Research Board Annual Meeting*, Washington, D.C, 2000.
- Lerman, S., and M. E. Ben-Akiva. A Behavioral Analysis of Automobile Ownership and Mode of Travel, Vol. 2. *Technical Summary*. U.S. Department of Transportation Washington, D.C., 1976.
- Lund, H. Testing the Claims of New Urbanism: Local Access, Pedestrian Travel, and Neighboring Behaviors. *Journal of the American Planning Association*, Vol. 69, No. 4, 2003, pp. 414-429.
- Mannering, F., and C. Winston. A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization. *The RAND Journal of Economics*, Vol. 16, No. 2, 1985, pp. 215-236.
- Manski, C. F., and L. Sherman. An Empirical Analysis of Household Choice among Motor Vehicles. *Transportation Research A*, Vol. 14, No. 5, 1980, pp. 349-366.
- Musti, S., and K. M. Kockelman. Evolution of the Household Vehicle Fleet: Anticipating Fleet Composition, PHEV Adoption and GHG Emissions in Austin, Texas. *Transportation Research Part A*, Vol. 45, No. 8, 2011, pp. 707-720.
- Paleti, R., N. Eluru, C. R. Bhat, R. M. Pendyala, T. J. Adler, and K. G. Goulias. Design of Comprehensive Microsimulator of Household Vehicle Fleet Composition, Utilization, and Evolution. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2254, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 44-57.
- Pendyala, R., C. R. Bhat, K. G. Goulias, R. Paleti, K. C. Konduri, R. Sidharthan, H. Hu, G. Huang, and K. P. Christian. The Application of a Socio-Economic Model System for Activity-Based Modeling: Experience from Southern California. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2303, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 71-80.

- Pinjari, A. R., and C. R. Bhat. Computationally Efficient Forecasting Procedures for Kuhn-Tucker Consumer Demand Model Systems: Application to Residential Energy Consumption Analysis. Technical paper, Department of Civil and Environmental Engineering, University of South Florida, July 2011.
- World Research Institute, *CAT1 2.0: WRIs Climate Data Explorer*. July 2014. [http://cait2.wri.org/wri/Country%20GHG%20Emissions?indicator\[\]=Total%20GHG%20Emissions%20Excluding%20Land-Use%20Change%20and%20Forestry&indicator\[\]=Total%20GHG%20Emissions%20Including%20Land-Use%20Change%20and%20Forestry&year\[\]=2011&chartType=geo](http://cait2.wri.org/wri/Country%20GHG%20Emissions?indicator[]=Total%20GHG%20Emissions%20Excluding%20Land-Use%20Change%20and%20Forestry&indicator[]=Total%20GHG%20Emissions%20Including%20Land-Use%20Change%20and%20Forestry&year[]=2011&chartType=geo). Accessed July 24, 2014.
- You, D., V. M. Garikapati, R. M. Pendyala, C. R. Bhat, S. Dubey, K. Jeon and V. Livshits. Development of a Vehicle Fleet Composition Model System for Implementation in an Activity-Based Travel Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2430, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 145–154.