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

Venu M. Garikapati ***(corresponding author)***
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355.
Tel: (480) 522-8067; Email: venu.garikapati@gatech.edu

Daehyun You
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355.
Tel: (480) 626-3716; Email: daehyun.you@gatech.edu

Ram M. Pendyala
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-894-2201; Fax: 404-894-2278; Email: ram.pendyala@ce.gatech.edu

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

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

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

Submitted for Presentation and Publication

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

Committee on Transportation Planning Applications (ADB50)

August 2015

ABSTRACT

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

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

1. INTRODUCTION

Understanding 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 attention is being paid to the role that household vehicle ownership and fleet composition play in shaping 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. 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 behavior) in the context of vehicle type choice. This is an important factor to be considered in fleet composition modeling as household often own multiple vehicle types and use them to varying levels.

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

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

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

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

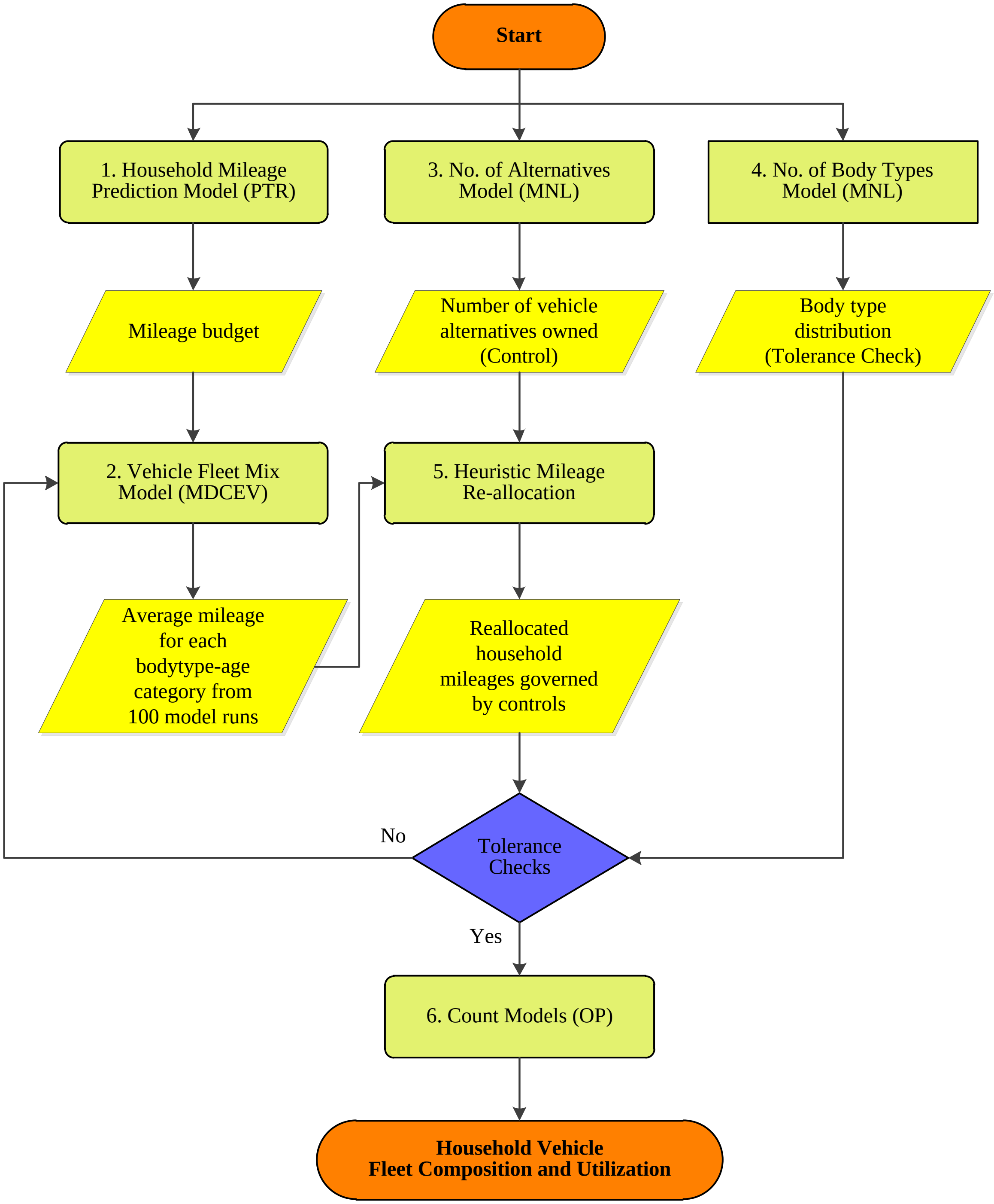
2. DATA

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

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

3. FRAMEWORK FOR VEHICLE FLEET COMPOSITION MODELING

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



PTR: Power Transformed Regression
MNL: Multinomial Logit
MDCEV: Multiple Discrete Continuous Extreme Value
OP: Ordered Probit

FIGURE 1 Vehicle fleet composition model framework

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

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

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

The mileage reallocation algorithm needs (as input) information on the number of distinct categories of vehicle alternatives owned by each household in the synthetic population. For this purpose, the vehicle fleet composition modeling framework includes an MNL model of the number of vehicle alternatives. If a household owns a car 0-5 years old, a car 6-11 years old, and a van 6-11 years old, the number of vehicle alternatives owned by this household is three. Thus, the HMR algorithm takes the outputs of the MDCEV model and the MNL model of number of vehicle alternatives as inputs. The logic followed by HMR algorithm is shown in Figure 2. The output of the MNL model of number of vehicle alternatives provides information on the number of vehicle types a household owns. Based on the output of the MDCEV model, a cumulative mileage distribution is computed for each household. Based on a Monte Carlo random draw, a vehicle alternative is selected from the cumulative distribution as ‘owned’ by the household. The chosen alternative is removed from the dataset, thereby eliminating the choice of the same alternative multiple times. This process is carried out '*k*' times, where *k* is the number of vehicle alternatives predicted by the separate multinomial logit model. Figure 3 presents a schematic illustrating the application of the Heuristic Mileage Reallocation (HMR) algorithm. In the sample illustration, a household is predicted to own three vehicle alternatives. Therefore, three random draws (without replacement) are performed to identify the specific alternatives chosen by a household. The mileage budget is scaled appropriately to ensure that the final mileage total matches the original mileage budget predicted for the household.

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

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

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

4. MODEL ESTIMATION AND APPLICATION RESULTS

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

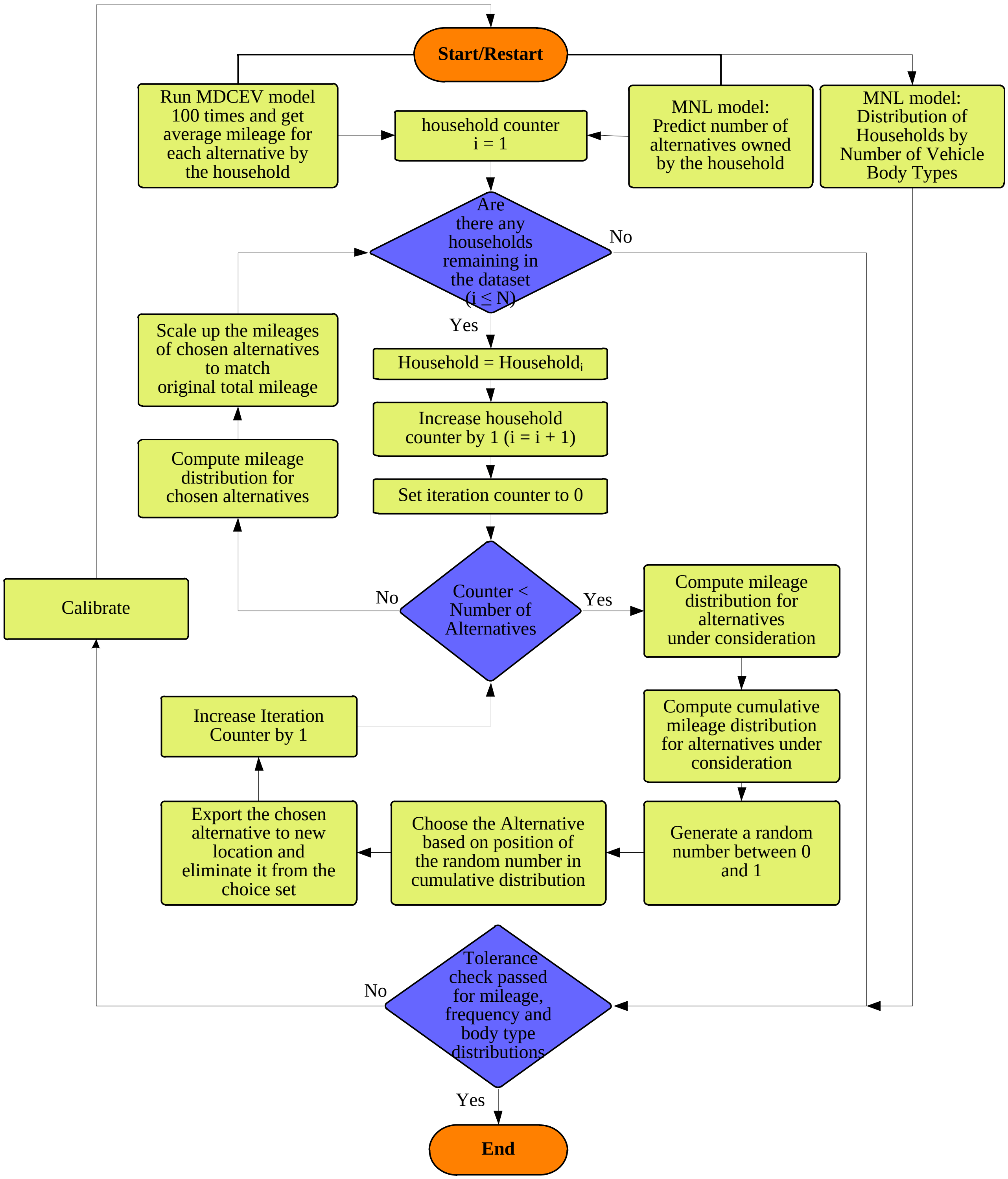


FIGURE 2 Heuristic mileage reallocation algorithm

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

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

4.1 Mileage Prediction Model

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

4.2 MDCEV Model of Fleet Mix

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

4.3 MNL Model of Number of Vehicle Alternatives

The MDCEV fleet mix model has a total of 13 motorized alternatives (4 body types x 3 vintage categories + motorbike). The MNL model of number of vehicle alternatives predicts the number of distinct vehicle categories owned by a household. Ideally the model should allow for a total of 13 alternatives, but observations from data revealed that the maximum number of distinct alternatives owned by any household in the dataset is five. Therefore, an MNL model is estimated with 6 categories (0, 1, 2, 3, 4, and ≥5vehicle alternatives) and the final category (≥5 vehicles) is considered as the base alternative. The purpose of this model is to provide a control total to the HMR algorithm in order to reallocate the mileage consumption distribution obtained by averaging the results from 100 MDCEV model simulations. Model estimation results for this component are presented in Panel A of Table 1.

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

4.4 MNL Model of Number of Body Types

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

4.5 Heuristic Mileage Reallocation (HMR) Algorithm

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

TABLE 1 Estimation Results for Multinomial Logit Models.

Panel A. MNL Model of Number of Vehicle Alternatives

Category	Explanatory Variable	Coefficient (t-stat)
Zero Alternatives	Constant	1.73(4.2)
	Lowest income household (< $25,000)	2.37(9.7)
	Low income household ($25,000 - $49,999)	0.82(3.23)
	Housing unit owned (from variable HOMEOWN)	-1.65(-9.95)
	Household size = 1	2.12(10.83)
	Zero worker household	1.21(6.42)
	Population density of household TAZ	0.0001(4.16)
One Alternative	Constant	4.27(13.72)
	Lowest income household (< $25,000)	1.43(11.59)
	Low income household ($25,000 - $49,999)	0.96(10.39)
	Household size = 1	2.22(18.38)
	Proportion of multi-family housing units in the TAZ	0.4(2.05)
	Two worker household	-0.86(-6.37)
Two Alternatives	Constant	5.31(17.14)
	Household with 2+ adults, youngest child 0-5	0.42(3.76)
	Medium income household ($50,000 - $74,999)	-0.21(-2.29)
	Two worker household	-0.34(-2.95)
	Households in lowest income quintile (Q1)	-0.0003(-2.03)
Three Alternatives	Constant	1.69(3.36)
	Housing unit owned (from variable HOMEOWN)	1(3.01)
	Count of adult HH members at least 18 years old	0.48(5.4)
	Three or more worker household	1.02(4.52)
	Population density of household TAZ	-0.0001(-3.49)
	Presence of children in the household	0.38(3.25)
	Household with 2+ adults, youngest child 16-21	0.5(2.5)
Four or More Alternatives	Constant	1.18(3.19)
	Highest income household (>= $100,000)	0.72(3)
	Household with 2+ adults, youngest child 16-21	1.63(5.47)
	Household size = 4 or more	1.51(6.25)
	Two worker household	-0.7(-2.59)

Goodness of Fit Statistics		
Sample Size (Number of Households)	Likelihood Ratio	$\chi^2_{(25,0.001)}$
4,262	2099.10	52.62

Panel B. MNL Model of Number of Vehicle Body Types

Category	Explanatory Variable	Coefficient (t-stat)
Zero Body Types	Constant	0.68(2.09)
	Lowest income household (< $25,000)	2.09(8.42)
	Low income household ($25,000 - $49,999)	0.61(2.38)
	Housing unit owned	-1.43(-8.58)
	Household size = 1	2.11(10.45)
	Zero worker household	1.14(6.15)
	Population density of the TAZ that the household resides	0.0001(4.08)
One Body Type	Constant	3.67(19.69)
	Lowest income household (< $25,000)	1.02(7.95)
	Low income household ($25,000 - $49,999)	0.65(6.86)
	Household size = 1	1.99(15.62)
	Proportion of multi-family housing units in the TAZ	0.26(1.39)
	Presence of children in the household	-0.33(-3.2)
Two Body Types	Constant	2.96(13.01)
	Household resides in rural area	0.19(2.01)
	High income household ($75,000 - $99,999)	0.35(3.34)
	Highest income household (≥ $100,000)	0.23(2.37)
	Household size = 4 or more	0.3(2.73)
	Housing unit owned	0.98(6.47)
Three or More Body Types	Housing unit owned	1.42(5.9)
	Count of adult HH members at least 18 years old	0.55(6.4)
	Three or more worker household	0.46(1.96)
	Population density of the TAZ that the household resides	-0.0001(-2.82)
	Presence of children in the household	0.36(2.55)

Goodness of Fit Statistics		
Sample Size (Number of Households)	Likelihood Ratio	$\chi^2_{(21,0.001)}$
4,262	1658.30	46.80

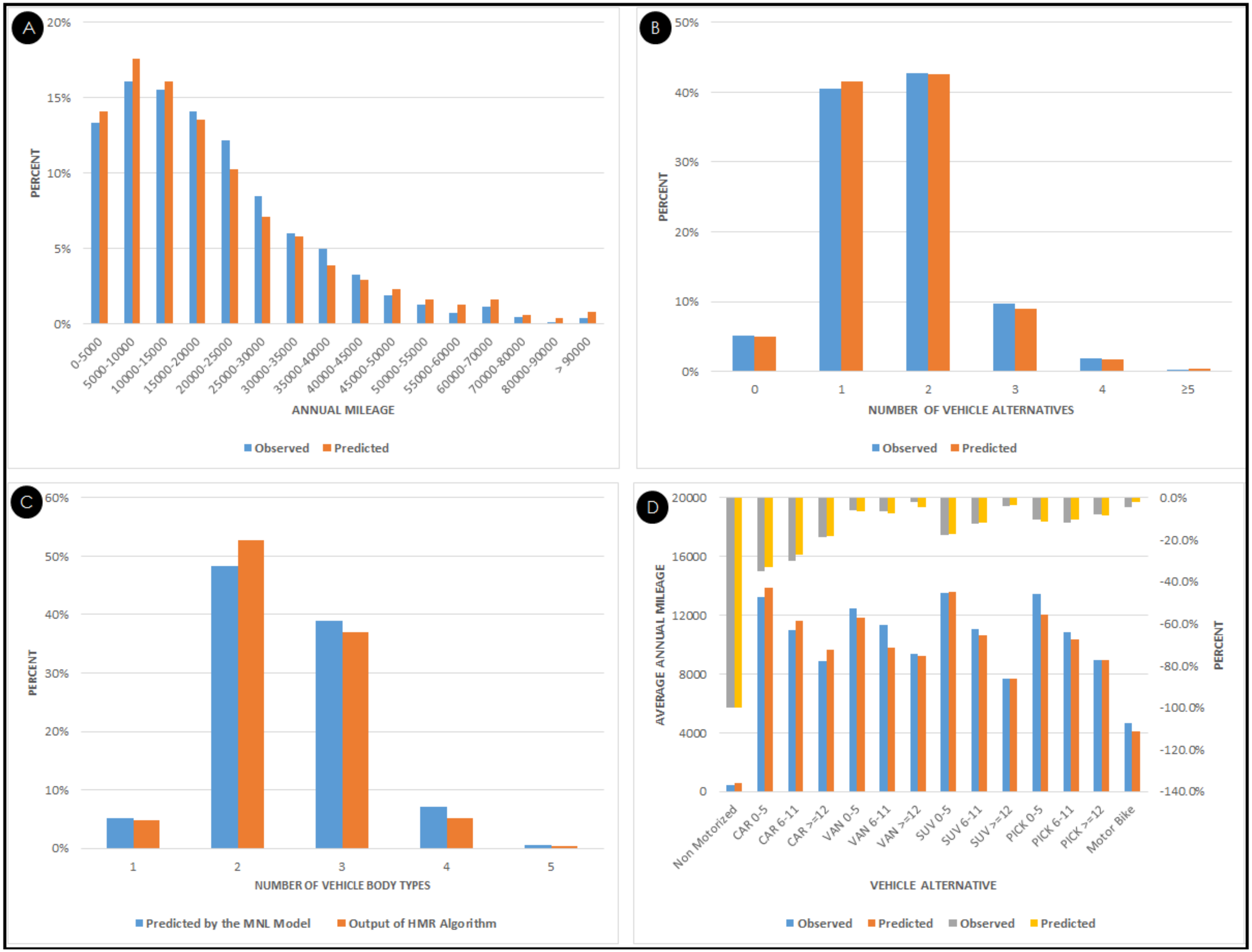


FIGURE 4 Model system replication results - part I: (A) mileage prediction model, (B) MNL model of number of alternatives, (C) MNL model of number of body types, and (D) MDCEV model of vehicle fleet mix

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

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

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

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

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

4.6 Count Models

Count models are necessary due to the aggregate categorical treatment of vehicle age (0-5 years, 6-12 years, 12 years or more) in the MDCEV model. If a household owns multiple vehicles of the same body type-vintage combination, the MDCEV model is not able to return the number of vehicles owned by a household within that alternative. The MDCEV model is only able to indicate that the household owns that vehicle type without providing any indication of the count of vehicles owned within the specific vehicle type chosen. Count models are applied for all vehicle alternatives that have non-zero mileage consumptions for a household. Ordered probit count models are estimated for car, van, SUV and pick-up truck body types. The maximum number of cars that a household may own is set at three in the current model specification due to data limitations. Similarly, the maximum number of vans, SUVs, and pick-up trucks that a household may own is set at two (again, due to data limitations). These values are set based on distributions of vehicle counts of different types in the estimation dataset. Motorbike category does not have a corresponding count model as the number of households owing multiple motorbikes is found to be very low.

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

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

5. CONCLUSIONS

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

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

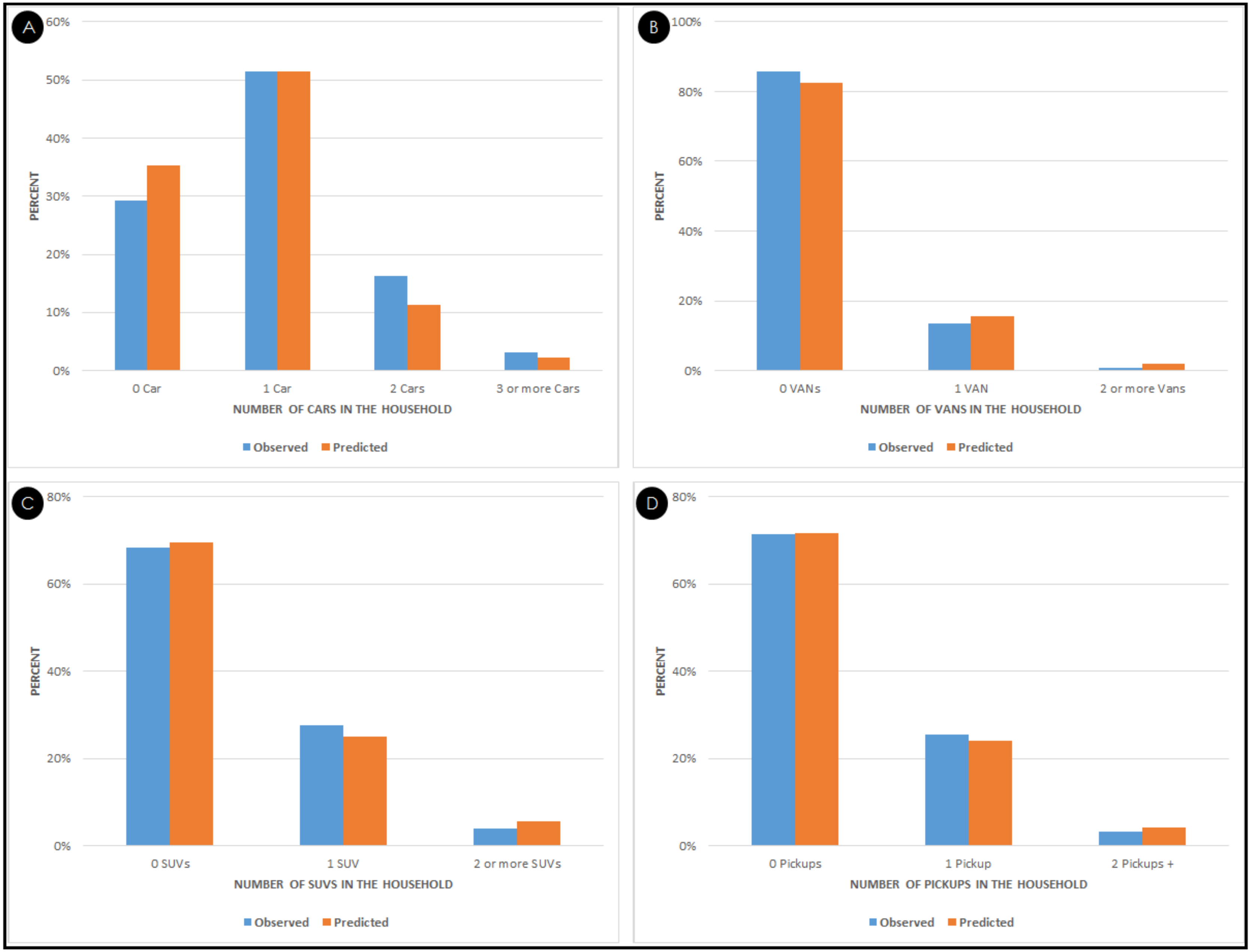


FIGURE 5 VFC model system replication results - part II: (A) car count model, (B) van count model, (C) SUV count model, and (D) pick-up count model

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 through 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. 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, *CATI 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. 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.