THE DESIGN OF A COMPREHENSIVE MICROSIMULATOR OF HOUSEHOLD VEHICLE FLEET COMPOSITION, UTILIZATION, AND EVOLUTION

Rajesh Paleti
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744, Email: rajeshp@mail.utexas.edu

Naveen Eluru
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744, Email: naveeneluru@mail.utexas.edu

Chandra R. Bhat* *(corresponding author)*
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744, Email: bhat@mail.utexas.edu

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
Room ECG252, Tempe, AZ 85287-5306
Phone: (480) 727-9164, Fax: (480) 965-0557, Email: ram.pendyala@asu.edu

Thomas J. Adler
Resource Systems Group, Inc.
55 Railroad Row, White River Junction, VT 05001
Phone: 802-295-4999, Email: tadler@rsginc.com

Konstadinos G. Goulias
University of California
Department of Geography
Santa Barbara, CA 93106-4060
Phone: 805-308-2837, Fax: 805-893-2578, Email: Kostas@geog.ucsb.edu

August 1, 2010

ABSTRACT
This paper describes a comprehensive vehicle fleet composition, utilization, and evolution simulator that can be used to forecast household vehicle ownership and mileage by type of vehicle over time. The modeling of vehicle ownership and utilization by type over time is of considerable interest in the current planning context where much attention is being paid to energy sustainability and environmental stewardship. The simulator framework presented in this paper includes two modules that unify existing streams of research in vehicle ownership and transactions modeling. The first module includes a copula based discrete-continuous model of vehicle fleet composition and utilization and is capable of simulating the types of vehicles (defined by body type, fuel type, and age) that households own, and the level of usage (mileage) associated with each vehicle owned. The second module includes a set of binary logit models capable of capturing household vehicle replacement, addition, and disposal decision processes over time. The simulator can be applied in annual time steps to evolve a base year vehicle fleet over time and offers the ability to produce detailed vehicular use profiles by type of vehicle that are necessary to accurately estimate energy consumption and greenhouse gas emissions under a wide range of scenarios. The components of the simulator are developed in this research effort using detailed revealed and stated preference data on household vehicle fleet composition, utilization, and planned transactions collected for a large sample of households in California. Results of the model development effort show that the simulator holds promise as a tool for simulating vehicular choice processes in the context of activity-based travel microsimulation model systems.

Keywords: vehicle fleet composition, household vehicle ownership, vehicle transactions and evolution, transportation demand forecasting, disaggregate microsimulation, behavioral choice model

1. INTRODUCTION

Activity-based travel demand model systems are increasingly being considered for implementation in metropolitan areas around the world for their ability to microsimulate activity-travel choices and patterns at the level of the individual decision-maker such as a household or individual. Due to the microsimulation framework adopted in these models, they are able to provide detailed information about individual trips, which in turn can result in substantially improved forecasts of greenhouse gas (GHG) emissions and energy consumption (Roorda *et.*, 2008). Concerns about energy sustainability, community vitality and livability, and the impact of GHG emissions on global climate have resulted in the profession focusing much attention on the ability of transportation models to accurately replicate the evolution of household and personal activity-travel choices over time for a wide range of scenarios.

One of the critical choice dimensions that have direct impact on energy consumption and GHG emissions is that of household vehicle fleet composition and utilization (Fang, 2008). Household vehicle ownership has long been considered an important determinant of travel demand, and there has been extensive research on the development of vehicle ownership models that allow one to estimate the number of vehicles owned/leased by households and/or their utilization (Bhat *et al.*, 2009; Brownstone and Golob, 2009). More recently, in light of the energy and emissions concerns, attention has been given to the types of vehicles owned by households – the type of vehicle being defined by the body type or size, the age of the vehicle, and the fuel type – as well as the mileage (utilization) of the vehicles (Mohammadian and Miller, 2003a; Cao *et al.*, 2006; Choo and Mokhtarian, 2004). These studies explicitly recognize that energy consumption and GHG emissions are not only dependent on the number of vehicles owned by households, but also on the mix of vehicle types and the extent to which different vehicle types are utilized (driven). The multiple discrete-continuous extreme value (MDCEV) modeling framework has proven valuable in the ability to model the ownership of multiple vehicle types and the travel mileage associated with each vehicle (Bhat and Sen, 2006; Bhat *et al.*, 2009).

The literature has recognized, however, that household vehicle ownership (or fleet composition and utilization) models are only capable of providing a snapshot of vehicle holdings and mileage as such models are routinely estimated on cross-sectional data sets that offer little to no information on vehicle transactions over time (Hensher and Le Plastrier, 1985; de Jong and Kitamura, 1992). As the focus of transportation planning is largely on forecasting demand over time, it is desirable to have a vehicle fleet evolution model that is capable of evolving a household's vehicle fleet over time (say, on an annual basis). Such a model would be akin to a demographic evolution model that purports to evolve or age a synthetic baseline population over time. The vehicle evolution model system should be sensitive to a range of socio-economic and policy variables to reflect that vehicle transaction decisions are likely influenced by the types of vehicle technologies that are and might be available, public policies and incentives associated with acquiring fuel-efficient or low/zero-emission vehicles, and household socio-economic and location characteristics (Brownstone *et al.*, 2000; de Haan *et al.*, 2009; Mueller and de Haan, 2009).

Unfortunately, however, the development of dynamic transactions models has been hampered by the paucity of longitudinal data on vehicle transactions that inevitably occur over time. Mohammadian and Miller (2003b) use about 10 years of data to model vehicle ownership by type and transaction decisions over time, but do not include fuel type as one of the attributes of vehicles. Yamamoto *et al.* (1999) use panel survey data to model vehicle transactions using hazard-based duration formulations as a function of changes in household and personal demographic attributes. Their study also shows the role of history dependency in vehicle transaction decisions with a preceding decision in time affecting a subsequent transaction decision. It is impossible to present a comprehensive literature review on this topic within the scope of this paper (see de Jong *et al.*, 2004 and Bhat *et al.*, 2009 for reviews), but suffice it to say that studies of dynamic vehicle transactions behavior emphasize the need for simulating vehicle fleet composition and utilization over time to accurately estimate energy consumption and GHG emissions arising from human activity-travel choices. The development and implementation of such model systems calls for the use of longitudinal data on vehicle holdings of households.

This paper offers a comprehensive vehicle fleet composition, utilization, and evolution framework that can be easily integrated in activity-based microsimulation models of travel demand. The model includes several components that allow one to not only predict current (baseline) vehicle holdings and utilization, but also simulate vehicle transactions (including addition, replacement, or disposal) over time. The usual data limitation is overcome in this study through the use of a unique large sample survey data set collected recently in California. The survey consisted of several components; a revealed choice (RC) component that gathered information on current vehicle holdings (RC data), a stated intentions (SI) component that gathered information on household plans (if any) on upcoming vehicle transactions in future years, including vehicle type purchase characteristics (SI data), and a special stated preference (SP) component that asked households to choose the type of vehicle that they would acquire/replace/dispose under a range of hypothetical scenarios (SP data). Pooling the data from these three survey components offered a rich data set for developing the comprehensive vehicle fleet composition and evolution simulator proposed in this paper.

The next section describes the proposed vehicle simulator framework. The third section provides an overview of the data set and survey sample. The fourth section presents the methodology. The fifth section discusses model estimation results, while the sixth section provides model evaluation statistics. The final section offers concluding thoughts.

2. VEHICLE FLEET COMPOSITION AND EVOLUTION FRAMEWORK

This section presents a brief outline of the vehicle fleet simulator that is capable of modeling baseline vehicle fleet composition and utilization as well as evolving the fleet of vehicles over time. The overall framework is presented in Figure 1. The simulator includes several components that are briefly described in this section.

First, there is a base year (baseline) model capable of predicting the current vehicle fleet composition and utilization of a household. The vehicle fleet is characterized by the mix of vehicles defined by body type (or size), fuel type, and vintage. The utilization of vehicles is defined by the travel mileage allocated to each vehicle in the fleet. In order to recognize the fact that the vehicles owned by a household at any given point in time are not acquired contemporaneously, the household is deemed to have acquired the vehicles on multiple choice occasions. Based on extensive analysis of travel survey data sets, it has been found that the number of vehicles owned by a household is virtually never greater than the number of adults in the household plus two. So, each household is assumed to have a number of choice occasions (on which to acquire a vehicle) equal to the number of household adults plus two. In the figure, an example is shown for a two-adult household with four possible choice occasions. In each choice occasion, a household may acquire a vehicle and associate an amount of mileage (utilization) to it, or may not acquire a vehicle at all. In this framework of a two-adult household, the household may have anywhere from zero to four vehicles depending on the choice made on each occasion. This choice occasion based approach to modeling vehicle fleet composition and utilization for a base year was successfully employed by Eluru *et al.* (2010). Note that the prediction of base year vehicle characteristics for each household are based on the vehicle selection module estimation (as just discussed) on a combined dataset that includes the vehicle type and usage choices provided by respondents in the California survey from the RC, SI and SP data components (*i.e.,* the estimation of the vehicle selection module within the dotted square of Figure 1 is undertaken from the RC, SI and SP data).

Once the base year fleet composition and utilization has been established for each household, the simulator turns to the evolution component. The evolution component works on an annual basis with households essentially faced with a number of possible choice alternatives (decisions). For each vehicle in the household, a household may choose to either dispose the vehicle (without replacing it) or replace the vehicle (involving both a disposal and an acquisition). If the choice is to replace the vehicle, then the vehicle selection module model estimation results can be applied to determine the type of vehicle that is acquired and the mileage that is allocated to it. Finally, a household may also choose to add a net new vehicle to the household fleet. In the case of an addition, once again the vehicle type choice and utilization model from the first simulator component can be applied to the vehicle acquired. Note that multiple transactions are possible in the same year. For example, consider a three vehicle household. One vehicle may be disposed, reducing the fleet size to two vehicles. Among the remaining two vehicles, one vehicle may be replaced with a new vehicle, while the second vehicle is simply retained (no disposal and no replacement). Thus, in any given year, the number of possible transactions is equal to the number of household vehicles currently owned by the household plus one (assuming that a household will not add more than one net vehicle to the fleet in any year). This framework overcomes the limitations of past studies that generally allowed only one possible transaction in any given year.

3. DATA

The data for the current study is derived from the residential survey component of the California Vehicle Survey data collected in 2008-2009 by the California Energy Commission (CEC) to forecast vehicle fleet composition and fuel consumption in California. The survey included three components, which are briefly discussed in turn in the next three paragraphs.

The revealed choice (RC) component of the survey collected detailed information on the current household vehicle fleet and usage. This included information about the vehicle body type, make/model, vintage, and fuel type for each vehicle. In addition, the annual mileage that each vehicle is driven/utilized and the identity of the primary driver of each vehicle are also collected. The usual set of household and personal socio-economic and demographic characteristics, as well as some household location attribute variables, is also gathered in this component of the survey.

The survey then included a set of questions to probe whether a household intended to replace an existing vehicle or acquire a net new additional vehicle in the fleet, and the characteristics of the vehicle(s) intended to be replaced or purchased (SI data). Essentially, the stated intention component of the survey gathered detailed information on replacement plans for each vehicle in the household fleet, and plans for adding net new vehicles (within the next five year period).

Finally, households that intended to purchase a vehicle either as a replacement or addition, and for whom there was adequate information on current revealed choices, were recruited for participation in a stated preference exercise (SP data). The SP exercises included several vehicle types and fuel technology options not currently available in the market, thus providing a rich data set for modeling vehicle transaction choices in a future context. The exercises involved the presentation of eight choice scenarios with four alternatives in each scenario. Attributes considered in describing each alternative included the vehicle type, size, fuel type, and vintage; a series of vehicle operating and acquisition cost variables; fuel availability, refueling time, and driving range; tax, toll, and parking incentives or credits; and vehicle performance (time to accelerate 0-60 mph).

The revealed choice (RC) and stated intentions (SI) data on current vehicle fleet composition and utilization was collected for a sample of 6577 households. Among these households, the stated preference (SP) component was administered to a sample of 3274 households who indicated that they would undertake at least one transaction in the future. The development of models for the vehicle simulator involved pooling the revealed choice (RC), stated intentions (SI) and stated preference (SP) components of the data. The pooled data set is ideal for modeling vehicle holdings and utilization, as well as vehicle transactions over time. The pooled data set allows one to include a range of vehicle types (including those not commonly found in the market place) in a vehicle type choice model, and provides a range of policy variables (from the stated preference component) that can be used to test the effects of such variables on vehicle fleet composition, utilization, and evolution decisions.

Extensive effort was expended in the preparation of the estimation data set. The vehicle selection module estimation was undertaken using a random sample of 1165 respondent households with complete information. Care was taken to ensure that the distributions of vehicle types, fuel type and vintage in the estimation data set were the same as those in the original data set of 6577 observations. The discrete dependent variable in the vehicle selection module estimation is a combination of six vehicle types (compact car, car, small cross utility vehicle, sport utility vehicle or SUV, van, and pick-up truck), six fuel types (gasoline, diesel, hybrid electric, fully electric, natural gas, and flex fuel), and five age categories (new, 1-2 years, 3-7 years, 8-12 years, and more than 12 years old). In addition, the no-vehicle choice category exists as well. Thus, there are a total of 211 alternatives in this choice process. The continuous dependent variable in the vehicle selection module estimation is the mileage traveled using each vehicle.

The vehicle evolution component of the model system developed in this paper includes the choice of replacement or addition of a vehicle. No information was collected on vehicle disposal plans and hence this choice dimension could not be considered using this data set. Of the 1165 household sample used for estimating the vehicle selection module, 915 households had complete information on vehicle transaction details (SI data). The survey asked each household to identify the number of years in which it plans to replace each vehicle in the household. Thus, the replacement choice process could be easily represented as an annual decision for each household, with replacement decisions beyond five years grouped into a single category of “five or more years”. Although the population is aged in the model estimation data set, many demographic changes are not taken into account (such as changes in number of workers, household income, household size, *etc*.) in the current effort; in ongoing work, the vehicle simulator described here is being integrated with a demographic evolution simulator to fully evolve households and their vehicle fleets over time.

4. METHODOLOGY

The comprehensive econometric microsimulator of vehicle fleet composition, utilization, and evolution includes two components. The first component is a vehicle selection module while the second component is the vehicle evolution module. In this section, the modeling methodology adopted in each component is described in brief.

4.1 Vehicle Selection Module

The vehicle selection module employs the traditional discrete-continuous framework for modeling the base year vehicle fleet composition and utilization. The vehicle fleet is described by the vehicle body type, fuel type, and vintage, and mileage is modeled using the logarithmic form. The methodology is the same as that described in Eluru *et al.* (2010) which is suitable to accommodate the many dimensions of vehicle fleet and usage decisions. The vehicle fleet and usage decisions are assumed to occur through a series of unobserved choice occasions. Based on the data, it is found that 99.5 percent of all households own less than or equal to N+2 vehicles, where N is the number of adults in the household. Therefore, N+2 possible choice occasions are created for each household from the RC data. At each choice occasion, the traditional multinomial logit model was used to represent the vehicle type choice and regression was used for the usage component. A joint copula (Eluru *et al.*, 2010) based framework is used to account for the fact that there might be common unobserved factors which affect both vehicle type and usage decisions.

Let q be the index for the households, q = 1, 2, 3,…., Q and let i be the index for the vehicle type alternatives. Let j be the index for the vehicle choice occasion j = 1, 2, …., J_q where J_q is the total number of choice occasions for a household q which is equal to N+2 (from RC data), plus the number of choice occasions where a replacement/addition decision was observed/reported (from SI data), plus up to eight choice occasions from the stated preference questionnaire (from SP data). For example, if a household with two adults (4 choice occasions) owns two vehicles, and it plans to replace the first vehicle with a sedan and add a SUV in the future (2 choice occasions), and if vehicle replacement decisions of the first vehicle are observed in the SP questionnaire eight times (eight choice occasions), then there are a total of 14 choice occasions for this household. With this notation, the vehicle type choice discrete component takes the following form:

$$u^*_{qij} = \beta' x_{qij} + \varepsilon_{qij} \tag{1}$$

u^*_{qij} is the latent utility that the qth household obtains from choosing alternative i at the jth choice occasion. x_{qij} is a column vector of known household attributes at choice occasion j (including household demographics and vehicle fleet characteristics before the jth choice occasion), β is the corresponding coefficient column vector of parameters to be estimated, and ε_{qij} is an idiosyncratic error term assumed to be independently and identically type-I extreme value distributed across alternatives, individuals, and choice occasions. Its scale parameter is normalized to one for revealed preference (RP) choice occasions and specified as $\frac{1}{\lambda}$ for the stated intention (SI) and stated preference (SP) choice occasions.

Then, the household q chooses alternative i at the jth choice occasion if the following condition holds:

$$u^*_{qij} > \max_{s=1,2,\ldots,I,s\neq i} u^*_{qsj} \tag{2}$$

The above condition can be written in the form of a series of binary choice formulations for each alternative i (Lee, 1983). Let R_{qij} be a dichotomous variable that takes the values 0 and 1, with R_{qij} =1 if the ith alternative is chosen by the qth household at the jth choice occasion, and R_{qij} =0 otherwise. Then, Equation (2) can be written as follows:

$$R_{qij} = 1 \text{ if } \beta' x_{qij} > v_{qij}, \ (i = 1, 2, \ldots, I) \tag{3}$$

where $$v_{qij} = \left\{ \max_{s=1,2,\ldots,I,s\neq i} u^*_{qsj} \right\} - \varepsilon_{qij} \tag{4}$$

Although the GEV (generalized extreme value) distribution assumption on the ε_{qij} terms may be used, independent and identically distributed errors across households, alternatives and choice occasions were assumed in this work, leading to the traditional multinomial logit model. This assumption can be relaxed using either GEV based models or mixing distributions. Such an effort is left for future research.

The vehicle mileage component takes the form of a classical log-linear regression as follows:

$$m^*_{qij} = \alpha_i' z_{qij} + \eta_{qij}, \quad m_{qij} = 1 \quad [R_{qij} = 1] m^*_{qij} \tag{5}$$

In the above equation, m^*_{qij} is a latent variable representing the logarithm of annual mileage for the vehicle type i if it had been chosen at the jth choice occasion. z_{qij} is the column vector of household attributes, α_i' is the corresponding column vector of parameter to be estimated, and η_{qij} is an error term assumed to be independent and identically distributed across households q and choice occasions j, and identically distributed across alternatives i. Also, since the annual mileage is observed only for the chosen vehicle type at each choice occasion, any dependence between the m^*_{qij} terms across alternatives is not identified,

The two model components discussed above are brought together in the following equation system:

$$R_{qij} = 1 \text{ if } \beta' x_{qij} > v_{qij}, \ (i = 1, 2, \ldots, I) \ (j = 1, 2, \ldots, J)$$

$$m^*_{qij} = \alpha_i' z_{qij} + \eta_{qij}, \quad m_{qij} = 1 \quad [R_{qij} = 1] m^*_{qij} \tag{6}$$

Copula based methods are used to determine the dependencies between the two stochastic terms v_{qij} and η_{qij}. In this method, the stochastic error terms are transformed into uniform distributions using their inverse cumulative distribution functions which are subsequently coupled into multivariate joint distributions using copulas (Eluru *et al*., 2010). The expression for the log-likelihood is similar to the one in Eluru *et al.* (2010). Six different copulas were used in this paper: (1) Gaussian copula, (2) Farlie-Gumbel-Morgenstern (FGM) copula, (3) Clayton, (4) Gumbel, (5) Frank, and (6) Joe copulas (Bhat and Eluru, 2009).

To complete the model specification, a normal distribution with zero mean and variance σ_η^2 is assumed for the η_{qij} terms. The variance of η_{qij} is assumed to be the same across all alternatives unlike in Eluru *et al.* (2010) because there are 210 vehicle type alternatives in this effort, and estimation of variance specific to each alternative is not feasible. Moreover, instead of estimating parameters specific to each alternative (α_i') in the regression equation, only one set of parameters (α_i) was estimated for efficiency.

4.2 Vehicle Evolution Module

The vehicle selection module results are used even in the vehicle evolution module for predicting vehicle type and usage. In addition, a binary logit model form is used for modeling both the vehicle replacement and addition decisions (on an annual basis). Let q be the index for the households, $q = 1, 2, 3, \ldots, Q$, let i be the index for the vehicle in the household and let j be the index for the vehicle replacement/addition occasion $j = 1, 2, \ldots, J_q$ where J_q is the total number of choice occasions for a household q which is equal to $\min\{t_{qi}, 5\}$, where t_{qi} is the number of years in which the household is planning to replace/add a vehicle i. For example, if a household with two vehicles plans to replace its first vehicle in two years, replace its second vehicle in five years, and add a vehicle in three years, then two choice occasions were created for the replacement decision of the first vehicle (0,1), five choice occasions for the replacement decision of the second vehicle (0,0,0,0,1), and three choice occasions for the addition decision (0,0,1), where 1 corresponds to an addition/replacement decision and 0 corresponds to a do-nothing option. With this notation, the vehicle evolution models take the following form:

$$l_{qij}^* = \gamma' w_{qij} + \vartheta_{qij}, \quad l_{qij} = 1 \text{ if } l_{qij}^* > 0 \text{ ; } l_{qij} = 0 \text{ otherwise} \tag{7}$$

l_{qij}^* is the latent utility that the qth household obtains from choosing to replace/add vehicle i at the jth choice occasion. w_{qij} is a column vector of known household attributes at choice occasion j (including household demographics and vehicle fleet characteristics before the jth choice occasion), γ is the corresponding column vector of parameters to be estimated, and ϑ_{qij} is an idiosyncratic error term assumed to be independently and identically type-I extreme value distributed across alternatives, individuals, and choice occasions.

5. MODEL ESTIMATION RESULTS

This section presents a few sample model estimation results to illustrate the nature of the model specifications. For the sake of brevity, complete model estimation results are not furnished in this paper (the complete model estimation results are available in Paleti *et al*., 2010). However, the discussion in this section provides an overview of results obtained from all model components. A sample of 1165 households with complete information provided the basis for estimating the model components. Descriptive statistics for this sample of households (as obtained from RC data) are shown in Table 1. Car, van, and SUV are the predominant vehicle types; annual mileage driven tends to be larger for larger vehicles than for cars, presumably because households use larger vehicles for longer trips. Less than two percent of the households report having no vehicle. All of the other descriptive statistics are self-explanatory and generally show a reasonable distribution of attributes that makes the sample suitable for estimating choice models.

5.1 Vehicle Selection Module

The vehicle selection module includes the vehicle type choice model component and the vehicle mileage component. For the sake of brevity, the model estimation results table is not shown for this model system. The results are similar to those reported in Eluru *et al.* (2010) and offered behaviorally intuitive interpretations. The constant associated with the no-vehicle category was largest, consistent with the fact that virtually all households have at least one vehicle. Also, there is a greater preference to own a compact car or car, in comparison to other vehicle types. Gasoline fuel vehicles are the most preferred, while compressed natural gas (CNG) and fully electric vehicles are the least preferred. As expected, households have a strong preference for newer cars.

A range of policy sensitive variables were included in the model. These are all generic variables, with a single effect estimated for each variable across all alternatives. All of the cost variables had negative coefficients indicating that as cost increases, the preference for a vehicle type decreases. Vehicle performance (time to accelerate from 0 to 60 mph) had a positive impact on the utility of an alternative, as did fuel efficiency. In the category of CNG and electric vehicles, those with medium (150-200 miles) and high (>200 miles) driving ranges are preferred over those with lower ranges. Interestingly, it was found that policy variables that offered incentives such as car pooling, free parking, $1000 tax credit, 50 percent reduction in tolls, and $1000 off the purchase price all had about the same effect on enhancing the utility of various alternatives. In other words, one policy incentive did not clearly outshine the others in terms of influencing vehicle type choice. But, all these policy variables are statistically significant in the final model. Although refueling time was not found to be statistically significant, the availability of refueling stations was a significant factor affecting vehicle type choice.

As expected, a range of household socio-economic and demographic variables significantly affected vehicle type choice. Households with more male adults had a preference for larger vehicles as opposed to compact cars and small cross utility vehicles. Interestingly, these households had a lower preference for hybrid electric vehicles than households with more females. Households with younger children tend to prefer larger vehicles, consistent with the notion that families probably like the room offered by such vehicles. Households with older children were found to have a preference for acquiring older vehicles, perhaps because parents get teenagers older vehicles when they first begin driving. On the other hand, households with senior adults tend to prefer newer vehicles, possibly because these households want trustworthy cars that are perceived to be safe.

As the household income increases, the inclination to get older vehicles decreases. These households are likely to be able to afford newer vehicles and have a preference to do so. Also, these households show a preference for a mix of vehicle body types including both small and large vehicles. This may suggest that these households are able to afford a mix of vehicle body types for different types of trips. Households located in rural and suburban regions are more inclined to own regular gasoline or diesel fuel vehicles. However, households in rural areas were also found to have a preference for hybrid electric vehicles. Those with a higher education level tended to have a preference for newer vehicles and alternative fuel vehicles. It is possible that these individuals tend to be more environmentally sensitive, leading to their preference for these types of less polluting vehicles. A set of findings hard to explain is that Caucasian households were more likely to prefer cars over larger vehicles, older vehicles over newer vehicles, and traditional fuel vehicles over alternative fuel vehicles. It is not immediately clear why these preferences exist for this group in comparison to other groups.

The existing household vehicle fleet has a significant impact on vehicle type choice/selection. Households tend to prefer less any vehicle type that already exists in their fleet. With respect to replacement, households are more prone to replace a vehicle in the fleet with the same type of vehicle. When the replaced vehicle is a SUV, households tend to replace it with a newer vehicle rather than an older one. Households which replace a gasoline fuel vehicle are more likely to replace it with an alternative fuel vehicle rather than a diesel fuel vehicle. This suggests that households looking to replace an existing gasoline vehicle are likely to consider newer alternative fuel vehicles; public policies aimed at offering incentives may provide the needed impetus to move in the direction of a greener fleet.

The vehicle usage (mileage) model component yielded largely intuitive results as well. Households with higher incomes are associated with higher travel mileage, consistent with the notion of more financial freedom to engage in out-of-home discretionary pursuits. Households with small children tend to have larger mileage, perhaps because these households have errands to run and serve-child trips that accumulate miles. Households in suburban regions also travel more than other households, possibly because suburban locations are more auto-oriented. Households with senior adults greater than 65 years of age tend to have lower mileage, presumably because these households consist of retired individuals living in empty nests. Households with more vehicles have lower mileage on a per vehicle basis, a manifestation of the ability to divide total household travel among multiple vehicles. Households with more workers have larger mileage, presumably due to greater levels of work travel. Similarly, households in which individuals are farther from their work places accumulate more mileage on their vehicles. Higher mileage values are associated with cars and larger vehicles such as SUV and van, but lower mileage values are associated with smaller cross utility vehicles and older vehicles.

5.2 Vehicle Evolution Models

The vehicle evolution model component consists of a replacement decision model and an addition decision model. Due to the absence of requisite data, it was not possible to estimate a disposal decision model with the data set used in this study. Estimation results for the models are not presented in the interest of brevity, but are discussed here.

The replacement model is a binary logit model that was found to offer plausible behavioral findings. The constant is significantly negative suggesting that households have a baseline preference to not replace their vehicles; this is consistent with the notion that vehicle transactions are infrequent events often spaced years apart. Caucasian and Hispanic households are more likely to replace a vehicle than households of other races. Larger households are more likely to replace a vehicle, presumably because they have greater turnover; however, if the vehicle being replaced is a SUV, then the probability of replacing it reduces. As expected, higher income households are more likely to replace a vehicle, while those with young children are less inclined to replace a vehicle. It is possible that households with young children are dealing with new expenses and do not feel the need to replace a vehicle. Households with older children are more likely to replace a vehicle, possibly because their fleet is getting old or because they are getting ready for the day when one or more children begins to drive. Sports cars and cross-utility vehicles are the least likely to be replaced; van, SUV, and pick-up truck are all equally likely to be replaced as the coefficients are extremely similar to one another. Older vehicles are more likely to be replaced than newer ones, although the coefficient for the 12 years or older category is less positive than for the 8-12 year old category. It is possible that vehicles 12 years or older have either been maintained very well, had parts replaced, or simply hold an emotional attachment that reduce the likelihood of replacement compared to the 8-12 year old category. Gasoline fuel vehicles are the most likely vehicle fuel type to be replaced, a finding consistent with the fact that gasoline vehicles are the predominant vehicle type in the population.

The vehicle addition model is also a binary logit model. Hispanic households were found to be the least likely to add a vehicle. Caucasians were found to be the second least likely to add a vehicle. Households with more adults and larger number of persons were more likely to add a new vehicle to their fleet. Lower income households were found to be more likely to add a vehicle in comparison to other higher income categories. It is possible that lower income households do not currently have the desired number of vehicles and hence desire to add a net additional vehicle to the fleet. Higher income households probably have the desired number of vehicles and so, rather than add a net additional vehicle, merely wish to replace an existing vehicle over time. Households with senior adults are less inclined to add a vehicle, while households with children aged 12-15 years are more likely to add a vehicle presumably because they are getting to acquire a vehicle for the new driver in the household. Households in rural regions appeared more likely to add a vehicle. As current vehicle fleet size increased, the less likely it was for a household to add a net additional vehicle. This was true across all vehicle type categories.

6. MODEL ASSESSMENT AND EVALUATION

As it is envisioned that this simulator will be used for vehicle fleet composition and evolution simulation in an activity-based travel model system setting, it was considered important to undertake a model assessment and validation effort. In the context of pooling RC, SI and SP data, one is often concerned with the possibility that the choice process exhibited in the RC data is different from that exhibited in the SI and SP data. For this reason, a scale parameter was estimated in the vehicle type choice – usage model to adjust model parameters in the joint RP-SI-SP model system. The RP to SI-SP scale parameter (λ) was estimated to be 0.5544 with a t-statistic of 24.15 (against a value of 1 which corresponds to the case when the variance of unobserved factors in the RP and SI-SP contexts are equal). This scale parameter is significantly smaller than unity, indicating that the error variance in the SI-SP choice context is higher than in the RP choice context (see Borjesson, 2008 for similar result).

In the vehicle selection module component, both a joint model that employed the copula-based estimation approach, and an independent model that models vehicle type choice and usage separately, was estimated. The log-likelihood of the independent model which ignores dependency between vehicle type choice and mileage is -29418.68. All of the models with different copula dependency structures are non-nested models. The Bayesian Information Criterion (BIC) (Trivedi and Zimmer, 2007) may be used to select the best model. However, as all competing models have the same number of exogenous variables, the use of the BIC reduces to choosing the model with the highest log-likelihood function value at convergence. Among all the copula structures considered, the Frank copula model offered the best statistical fit with a log-likelihood value of -29232. The improvement in fit, relative to the independent model, is readily apparent.

The copula dependency parameter (θ) is estimated to be equal to -3.367 with a t-statistic of -9.39. This shows that there is significant dependency between the vehicle type choice and usage dimensions. The Kendall's measure (τ) which is similar to the standard correlation coefficient was computed using the expression:

$$\tau = 1 - \frac{4}{\theta}\left[1 - \frac{1}{\theta}\left[\int_{t=0}^{\theta} \frac{t}{e^t - 1} dt\right]\right]$$

The value of τ was found to be -0.3622. The error term ν_{qij} enters Equation (3) with a negative sign. Thus, a negative sign on the Kendall's measure indicates that the unobserved factors which increase the propensity to choose a certain vehicle type also increase the propensity to accumulate more mileage on that vehicle. This is consistent with expectations. For example, if there are unobserved factors that make a household prefer higher performance sports cars, then it is likely that those very same factors also influence a household to drive (and enjoy) that vehicle type to a greater extent.

A model evaluation exercise was undertaken to test the comparative ability of the joint model and the independent model to replicate vehicle fleet composition choices in a random hold-out sample of 500 households not included in the estimation sample (see Table 2). The predicted log-likelihood function values of the independent and copula-based joint models were compared for different segments of the hold-out sample. The overall predictive log-likelihood ratio test values for comparing the copula based joint model with the independent model indicate that the copula based joint model is statistically significantly better than the independent model in all cases, except for households with no vehicles and households that have children aged 5-11 years where there is no appreciable difference in predictive power between the two models. The results clearly demonstrate the superiority of the joint model in predicting vehicle fleet composition and utilization, relative to the independent model. Ongoing research includes additional and extensive validation and sensitivity tests of the vehicle fleet simulator developed in this study.

7. CONCLUSIONS

The modeling and analysis of household vehicle ownership and utilization by type of vehicle has gained added importance in recent years in the face of rising concerns about global energy sustainability, greenhouse gas (GHG) emissions, and community livability in urban areas around the world. Households may choose to own and drive (utilize) a variety of different vehicle types and the ability to accurately forecast these choice dimensions is undoubtedly of much interest in the current planning context which is dominated by efforts on the part of planners and policy makers to minimize the adverse impacts of automobile use on the environment.

This paper presents the design and formulation of a comprehensive vehicle fleet composition and evolution simulator that is capable of simulating household vehicle ownership and utilization decisions over time. The simulation framework consists of two main modules – one module that models the current (baseline) fleet composition and utilization for a household and another module that evolves the baseline fleet over time by considering the acquisition, replacement, and disposal processes that households may undertake as they turnover their fleet.

One of the major impediments thus far to the development of such a vehicle fleet evolution simulation system has been the availability of longitudinal data on the dynamics of household vehicle ownership and utilization by type of vehicle. This issue is overcome in this study through the use of a large sample data set collected as part of a survey undertaken by the California Energy Commission in California. The survey includes a revealed choice (RC) component that captures information about current vehicle fleet information for the respondent households, a stated intentions (SI) component that captures information on the plans of respondent households to replace existing household vehicles or add net additional vehicles to the fleet (and the timing of such potential transactions), and a stated preference (SP) component that captures information on the vehicle type likely to be chosen by households when faced with a set of hypothetical choice scenarios. Data from these three survey components are pooled together to obtain a rich data set that can be used to model the full range of vehicle ownership and transactions decisions of households.

The paper includes a detailed description of the simulator framework, the modeling methodologies employed in various modules of the framework, and estimation results for various model components. In general, it is found that socio-economic characteristics, vehicular costs and performance measures, government incentives, and locational attributes are all important in predicting vehicle fleet composition, utilization, and evolution. The joint modeling framework is applied to predict vehicular choices for a random holdout sample of households and shown to perform substantially better than an independent set of model components that ignore common unobserved factors that impact both vehicle fleet composition and utilization.

The approach presented in this paper offers the ability to generate vehicle fleet composition and usage measures that serve as critical inputs to emissions forecasting models. The novelty of the approach is that it accommodates all of the dimensions characterizing vehicle fleet/usage decisions, as well as all of the dimensions of vehicle transactions (*i.e.*, fleet evolution) over time. The resulting model can be used in a microsimulation-based forecasting model system to obtain the fleet composition for a future year and/or examine the effects of a host of policy variables aimed at promoting vehicle mix/usage patterns that reduce GHG emissions and fuel consumption. Further work involves the implementation of the vehicle simulator in the activity-based travel demand model system for the Southern California region.

ACKNOWLEDGMENTS

The authors would like to thank the California Energy Commission for providing access to the data used in this research, and the Southern California Association of Governments for facilitating this research. The authors are also grateful to Lisa Macias for her help in formatting this document.

REFERENCES

Bhat, C.R. and N. Eluru (2009) A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, 43(7), 749-765.

Bhat, C.R. and S. Sen (2006) Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B*, 40(1), 35-53.

Bhat, C.R., S. Sen, and N. Eluru (2009) The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B*, 43(1), 1-18.

Borjesson, M., (2008) Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E*, 44, 1025-1038.

Brownstone, D. and T.F. Golob (2009) The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.

Brownstone, D., D.S. Bunch, and K. Train (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B*, 34(5), 315-338.

Cao, X., P.L. Mokhtarian, and S.L. Handy (2006) Neighborhood design and vehicle type choice: evidence from Northern California. *Transportation Research Part D*, 13(2), 133-145.

Choo, S. and P.L. Mokhtarian (2004) What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A*, 38(3), 201-222.

de Haan, P., M.G. Mueller, and R.W. Scholz (2009) How much do incentives affect car purchase? Agent-based microsimulation of consumer choice of new cars, part II: Forecasting effects of feebates based on energy-efficiency. *Energy Policy*, 37(3), 1083-1094.

de Jong, G. and R. Kitamura (1992) A review of household dynamic vehicle ownership models: holdings models versus transactions models. *Proceedings of Seminar E, 20th PTRC Summer Annual Meeting,* PTRC Education and Research Services Ltd., London, 141-152.

de Jong, G.C., J. Fox, A. Daly, M. Pieters, R. Smit (2004) A comparison of car ownership models. *Transport Reviews*, 24(4), 379-408.

Eluru, N., C.R. Bhat, R.M. Pendyala, and K.C. Konduri (2010) A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation*, 37(4), 603-626.

Fang, A. (2008) Discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B*, 42(9), 736-758.

Hensher, D.A. and V. Le Plastrier (1985) Towards a dynamic discrete-choice model of household automobile fleet size and composition. *Transportation Research Part B*, 19(6), 481-495.

Lee L.-F. (1983) Generalized econometric models with selectivity. *Econometrica*, 51(2), 507-512.

Mohammadian, A. and E.J. Miller (2003a) Empirical investigation of household vehicle type choice decisions. *Transportation Research Record*, 1854, 99-106.

Mohammadian, A. and E.J. Miller (2003b) Dynamic modeling of household automobile transactions. *Transportation Research Record*, 1837, 98-105.

Mueller, M.G. and P. de Haan (2009) How much do incentives affect car purchase? Agent-based microsimulation of consumer choice of new cars-part I: Model structure, simulation of bounded rationality, and model validation. *Energy Policy*, 37(3), 1072-1082.

Paleti, R., N. Eluru, C.R. Bhat, R.M. Pendyala, and T.J. Adler (2010) Estimation Results: The Design of a Comprehensive Microsimulator of Household Vehicle Fleet Composition, Utilization, and Evolution. Supplementary Material, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin. http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Estimation_Results.pdf

Roorda, M.J., J.A. Carrasco, and E.J. Miller (2008) An integrated model of vehicle transactions, activity scheduling and mode Choice. *Transportation Research Part B,* 43(2), 217-229.

Trivedi, P. K. and D. M. Zimmer (2007) Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, 1(1), Now Publishers, Inc., 1-111.

Yamamoto, T., R. Kitamura, and S. Kimura (1999) Competing-risks duration model of household vehicle transactions with indicators of changes in explanatory variables. *Transportation Research Record*, 1676, 116-123.

LIST OF TABLES

LIST OF FIGURES

TABLE 1 Sample Characteristics

Variable	**Sample Share (%)**	**Mean Mileage**
Vehicle Type		
Compact Car	25.6	11894.36
Car	29.3	11887.08
Sports or Small Cross-utility Vehicle	4.8	11612.97
SUV	18.5	13099.24
Van	5.9	13019.13
Pickup	16.0	12310.61
Number of vehicles		
Zero	1.8	
One	28.4	
Two	50.0	
Three	14.2	
Four or more	5.6	
Number of adults		
One	18.5	
Two	64.3	
Three	10.7	
Four	4.9	
Five or more	1.5	
Number of workers		
Zero	18.3	
One	34.5	
Two	39.8	
Three	5.5	
Four or more	1.9	
Location		
Urban	48.2	
Suburban	47.8	
Rural	4.0	
Presence of senior adults	22.1	
Presence of children		
0-4 years	12.8	
5-11 years	14.9	
12 to 15 years	10.4	
Household Income		
<$20k	3.3	
Between $20 and $40K	13.1	
Between $40 and $60K	16.0	
Between $60K and 80K	18.3	
Between $80K and $100K	14.8	
Between $100K and $120K	10.8	
> $120K	23.7	
Educational Attainment		
High school	8.2	
College (with/without degree)	58.0	
Post Graduate	33.8	
Total Sample Size	1165	

TABLE 2 Disaggregate Measures of Fit for the Validation Sample

Sample details	**Number of households**	**Independent model predictive likelihood**	**Copula based joint model predictive likelihood**	**Predictive likelihood ratio test ($\chi^2_{1,0.05} = 3.84$)**
Full validation sample	500	-14168.33	-14106.56	123.53
Number of vehicles				
Zero	6	-156.77	-156.39	0.77
One	152	-3024.61	-3016.16	16.89
Two	225	-6325.87	-6309.33	33.07
Three	89	-3292.37	-3260.62	63.49
Four or more	28	-1368.71	-1364.06	9.31
Number of workers				
Zero	90	-2121.81	-2119.69	4.23
One	171	-4509.36	-4494.66	29.40
Two	196	-5841.05	-5812.66	56.78
Three	37	-1383.00	-1369.84	26.32
Four or more	6	-313.10	-309.70	6.80
Highest Educational Attainment				
High school	43	-1117.03	-1108.82	16.43
College (With/without degree)	271	-7753.09	-7726.33	53.51
Post Graduate	186	-5298.21	-5271.41	53.60
Presence of children				
0-4 years	57	-1672.28	-1658.65	27.24
5-11 years	74	-2187.88	-2187.10	1.56
12-15 years	58	-1913.27	-1891.21	44.13
Presence of senior adults (Age≥65 years)	113	-2900.29	-2888.63	23.32
Region				
Urban	241	-6701.42	-6663.08	76.68
Sub-urban	235	-6764.65	-6746.77	35.76
Rural	24	-702.25	-696.71	11.10

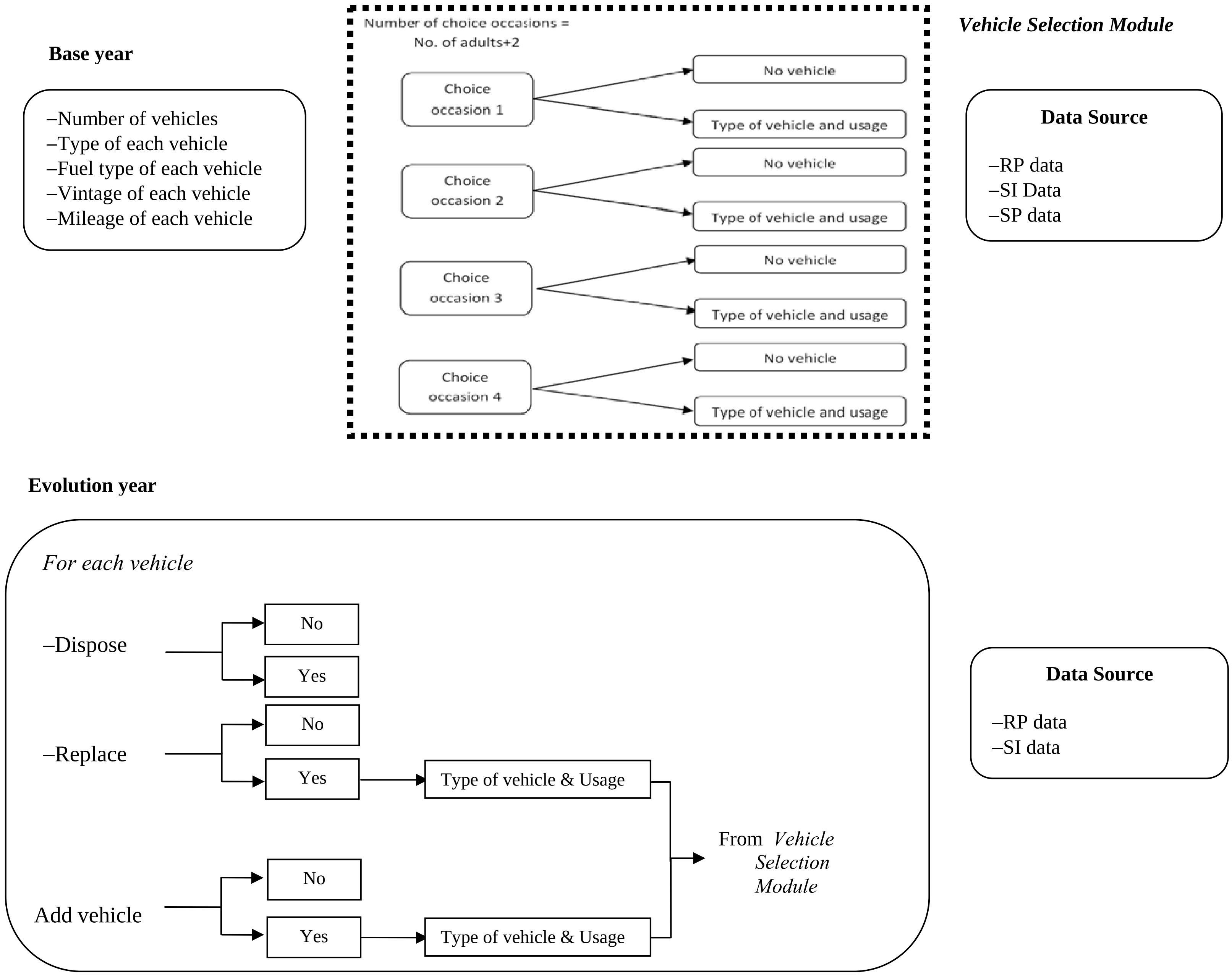

FIGURE 1 Vehicle fleet composition, utilization, and evolution simulator framework.