AN ANALYSIS OF HOUSEHOLD VEHICLE TYPE ACUQISITION

USING MULTINOMAIL LOGIT MODEL

by

Kim, Sang-heun

Advisor: Guo, Y. Jessica

A research paper submitted in partial fulfillment of

the requirements for the degree of

Master of Science

Department of Civil Engineering and Environment

at the University of Wisconsin-Madison

Table of Contents

CHAPTER 1 INTRODUCTION

1.1 Background

The US is a highly motorized society and personal vehicles are clearly an integral part of American life in that people are used to rely on their personal vehicles when both traveling and commuting. According to the 2001 National Household Transportation Survey (NHTS), household travel generates more than 80 percent of the total vehicle miles of travel in the United States and 87% of the daily trips are made by personal vehicles. Overall personal vehicles have been outnumbering licensed drivers since 1972 at an ever increasing rate. Each year more than 100 new mainstream models are manufactured and millions of vehicles are sold in America. Naturally, people ponder which type of vehicle to select based on their own preferences and needs out of a wide range of makes and models. It can be varied with household demographics, residence location, or primary driver's characteristics. Historically, there has been a trend of a popular type of vehicle in various time spans such as compact cars in the mid-1970s, minivans in the 1980s, and pickups/sport utility vehicles in the 1990s. Here, an important question arises; what factors determined or contributed to the choice of a certain type of vehicle? In order to find this kind of question, many researchers have devoted to identifying the factors that have an influence on peoples' vehicle choosing behaviors. As a result, a variety of models have been developed for explaining the vehicle type choice and disaggregate choice models such as multinomial logit model (e.g. Lave and Train, 1977; Mannering and Winston, 1985; Choo and Mokhtarian, 2004) and nested logit model (Berkovec, 1985; Feng et al., 2005; Mohammedian and Miller, 2003) for the vehicle type measure and continuous regression model for the vehicle use measure have been widely used by many eminent studies in this field. These models generally focused on several types of determinants such as vehicle attributes, household demographics, residence location factors, primary drivers' characteristics, or personalities and travel attitudes. Several earlier models with such those general factors basically show consistent results, however, failed to capture the unique features except for the intuitive results whatever methods they used. Of course, it is true that the earlier studies produced undoubtedly remarkable achievements with providing important insights into the factors affecting the choice of vehicle type from their perspectives.

1.2 Objectives

Unlike a traditional society in the past, American life nowadays has no longer a certain form. As such, only one-fourth of households in the US consist of married couple and their children these days. Increasing level of complexity in people's activity and diversity of a unit of household composition clearly contribute much to the shift of people's existing traditional travel patterns and vehicle selection behaviors. While vehicle type choice in households has so far been studied extensively, relatively few studies have been devoted to the effect of diverse household compositions on vehicle type choice. As vehicle ownership by vehicle type is of much interest to transportation planners and policy makers, the aim of this study is not only to understand the principle of vehicle type choice analysis by following existing methods such as multinomial logit model and nested logit model but also to provide further detailed analysis of vehicle type choice, especially about the influence of various household compositions, using the data from 2001 National Household Transportation Survey (NHTS). This study attempts to identify the feature of eight vehicle types; subcompact cars, compact cars, mid-size sedans, full-size sedans, SUVs, pickup trucks, vans, and sports cars, by providing descriptive analysis based on their attributes, resident location characteristics, household demographics, and primary drivers' characteristics.

In order to explore the current behaviors of choosing vehicle types in households, this paper will be limited to consideration of recently achieved vehicles. Furthermore, major eight vehicle types are considered; subcompact cars, compact cars, mid-size sedans, full-size sedans, SUVs, pickup trucks, vans, and sports cars.

1.3 Organization

The remaining part of the report has been organized as follows. Chapter 2 starts with a literature reviews summarizing and comparing the data sets, choice sets, model structures, and explanatory variables they used in terms of their purpose and issues. Chapter 3 introduces the description of the data that have been used and furnishes details of the methodology of the modeling framework used in this study. Chapter 4 gives an overview of the results of vehicle type choice analysis using a multinomial logit model. Finally, conclusion and the scope for further research are summarized in Chapter 5.

CHAPTER 2 LITERATURE REVIEW

2.1 Overview of the literature

Within the society where automobile plays a pivotal role in daily life, the analysis and modeling of vehicle type choice in households has been much of interests to social scientists and transportation planners. To that end, since one of the initial disaggregate studies was published by Lave and Train (1977) in which they evaluated a multinomial logit model for the type of car that households will choose to buy, vehicle type choice has been extensively explored from a variety of technical and empirical perspectives including introduction of new model structures (Bhat and Sen, 2006; Fang, 2008), development of operational models (Mohammadian, 2003), investigation of the influence of attitudes and life style of individuals (Choo and Mokhtarian, 2004) as well as the study of the influence of the build environment (Cao, 2006; Potoglou, 2009) for past decades.

The studies to be reviewed are organized as follows. First, the dataset they used are presented in chronological order. The next subsection presents several types of choice set in terms of their purpose or nature of studies. Afterwards, the structures of models which are mainly applied are discussed. The following subsection considers the important explanatory variables including insignificant variables opposed to their initial hypotheses. Finally, the limitation of each study and development for future works are discussed.

2.2 Data set

Researches in an early stage were conducted under certain restriction due to the limitation of data which were limited to a certain area, not focused on the aim of household vehicle ownership analysis, didn’t include diverse aspects of information, and also had a relatively small number of sample cases. Lave and Train (1979) used 541 new car buyers data in seven cities as their sample for the research, Berkovec and Rust (1985) estimated a nested logit model of automobile holdings using data on 237 single vehicle households and they admitted the further promising approach was restricted by the lack of the amount on their dataset. Since 1990s several large scale of survey have been systematically gathered for either national or local purpose, thereafter more various and elaborated researches have been developed based on these ample information. Golob et al. (1997) used 1,869 households sample from a 1993 survey in most of the urbanized area of California for their modeling of the vehicle miles of travel for each vehicle in multi-vehicle households. Adjemian et al. (2010) estimated spatial interdependence in automobile type choice with the data from the 2000 San Francisco Bay Area Travel Survey (BATS) authorized by the Bay Area Metropolitan Transportation commission, and this data have been widely used by many other researchers in that the survey is comprehensively augmented with land use and network level of service attributes (Bhat, 2006; Spissu, 2009; Eluru, 2010). This survey, on the other hand, does not include information on fuel economy, fuel use, and fuel cost by vehicle make and model, which are important to examine the economic and environmental influences associated with household vehicle fleet holdings and usage. Another major data extensively used are that of the 2001 National Household Travel Survey data containing comprehensive data on travel and transportation patterns in the United States conducted by Bureau of Transportation Statistics (BTS), Federal Highway Administration (FHWA), and National Highway Traffic Safety Administration (NHTSA) (Feng, 2005; Fang, 2008). This dataset, unlike the previous dataset, has the advantages of not only having several automobile attributes in company with environmental properties but also including both national sample and nine add-on samples so that researchers use this dataset on their purpose suitably.

2.3 Choice set

The decision of choice set to be explained by the models is quite diverse depending upon the structure of the models and the objective of the study. Examining other vehicle classification schemes found in the previous literatures is useful in determining the vehicle type categories more effectively. The strategies are focused on vehicle size, function, or combination. Most of strategies classifies the vehicle type in order of vehicle size first, and then adds some specialized vehicle types such as pickup, SUV, and sports car. Vehicle types are classified into 10 categories including subdivisions within categories based on size and price: subsubcompact, sports, subcompact A and B, compact A and B, intermediate, standard A and B, and luxury (Lave and Train, 1979). The vehicles are classified into 600 different types by make, model, and vintage, but the models use only 26 alternative vehicle types which include the chosen alternative and 25 others randomly selected from the universal choice set (Manski and Sherman (1980). The vehicle types were classified by make, model, body type, and vintage (using vintage dummy variables for less than 2 years, 2-9 years, 10-14 years, and 15 years or older). In addition to the chosen alternative, 19 alternative vehicle types were randomly selected from 950 different types identified for the models (Hocherman et al., 1983). The vehicle types are classified into 15 categories based on size (subcompact, compact, intermediate, standard, and luxury/sports) and age (new (1977-78), mid (1973-76), and old (1967-1972)) (Berkovec and Rust (1985). Vehicles are classified into 131 different types based on make, model, and vintage plus an old car group of all pre-1969 vehicles (Berkovec, 1985). The vehicle types are classified by make, model, and year (e.g. Ford Maverick 1972). The dependent choice set includes the chosen alternative and nine others randomly selected from more than 2,000 different types (Mannering and Winston

(1985). The choice set is classified into six bundles: one choice is the number of vehicles (0, 1, or 2), and another choice for each vehicle is the type of vehicle (a car or an SUV) in that there are no necessity to model the choice among hundreds of vehicle types in order to characterize the vehicle types by emission rate (Feng et al., 2005). The vehicles were categorized into one of five vehicle types: passenger car, SUV, pickup truck, minivan, and van (Bhat and Sen, 2006). Vehicle are classified into ten types according to the classification in Consumer Report: compact, compact luxury, sedan mid-size, sedan full-size, sedan luxury, SUV small, SUV mid-size, SUV large, minivan, and pickup truck (Fang, 2005). The vehicle types are classified into six categories: coupe, SUV, pickup truck, vans (including minivans), compact sedans (including subcompact sedans) and large sedans (including mid-size sedans and station wagons). In addition to these six alternatives, there exists the “no vehicle” alternative (Eluru et al., 2010). The vehicles in BATS are classified into nine types according to those used by the auto information company Edmunds.com: coupe, compact sedan, mid-size sedan, large sedan, station wagon, SUV, pickup truck, minivan/van, and sports car (Adjemian et al., 2010). The nine vehicle categories used in Consumer Reports are classified. Specifically, the first five categories are classified in order of vehicle size and the other categories are added to represent specialized vehicles: small (including minicompact and subcompact), compact, mid-sized, large, luxury, minivan/van, sports, pickup, SUV.

2.4 Model structure

Many different types of models have been developed to address the various combinations of the issues and prevail against the existing problems caused by previous models.

Such disaggregate models as multinomial logit (Lave and Train, 1977; Mannering and Winston, 1985; Choo and Mokhtarian, 2004) and nested logit model (Berkovec, 1985; Feng et al., 2005; Mohammedian and Miller, 2003) are generally adopted to account for the vehicle type choice. Traditionally, these models can be further divided into two categories in terms of the purpose and point of view of the study, vehicle holdings and vehicle transaction models. The former can be distinguished as a model which deals with a household vehicle fleet at a single point in time, while the latter explains changes to the household vehicle fleet, such as replacement and disposal. Majority of early studies presented the vehicle holdings model which describes the likelihood that a household of a given attributes will hold a particular type or set of vehicles, and successfully accounted for the influences on vehicle demand of various vehicle attributes, household and primary driver attributes, and further personal attributes (Lave and Train, 1985; Mannering and Winston, 1985; Choo and Mokhtarian, 2004). In practice, however, the current set of vehicles the household owns is not acquired instantaneously but an outcome or process from a chain of transaction decisions to buy, selling, or retaining of household vehicle fleets. From this point of view, one can argue against the holdings model in terms of insufficiency to quantify such a sequential decision reflecting the long-term planning efforts of the household. Afterwards, several attempts to supplement and further compel this issue was made, and the transaction models which represent changes in a household’s vehicle ownership status, such as buying and/or selling a car, were developed. Even though there are many advantages in transaction models, only very few researches have been accomplished so far mostly in that the appropriate data which are required a longitudinal survey with the household transaction records are hard to be obtained (Mohammadian and Miller, 2003; Hocherman, Prashker and Ben-Akiva, 1983).

As mentioned above, both models can be separated into two types, multinomial logit model (MNL) and nested logit model (NL). The MNL is a general model to describe decision makers' choices among alternatives which gives the choice probabilities of each alternative as a function of the systematic portion of the utility of all the alternatives. But the MNL assumes proportional substitution patterns (IIA), i. e. the ratio of the choice probabilities of two alternatives is not dependent on the presence or absence of other alternatives in the model. To overcome these restrictive substitution assumptions between alternatives, various extensions of the MNL exist. The NL which groups similar alternatives into same branches and thus creates a hierarchical structure of the alternatives has been widely favored in this area (Ben-Akiva and Lerman, 1985; Train, 2003). For vehicle holding analysis, using nested logit model the type choice probabilities are estimated conditional on the chosen number of vehicles by jointly considering the vehicle types and the number of vehicles. The nested logit model for vehicle transaction examines the household's choice of the vehicle type conditional on transaction behavior when buying additional car and/or selling a current car.

Recently, some advances have been made in dynamic vehicle ownership models, which are more general models including vehicle type choice model. In particular, based on the idea that the amount a household drives affects both the number and type of vehicles that it chooses to own, some studies modeled the household's choice of how much to drive and the interrelation of this choice with the choice of how many and what types of vehicles to own. Even though some previous models considered the vehicle usage (vehicle miles travel) as one of influential factors in vehicle type choice, simultaneity bias cannot be excluded because most of them treated it as exogenous. In order to vest the vehicle type choice with such endogeneity, the model takes the form of a joint discrete-continuous structure in which the discrete component represents the vehicle type choice dimension and the continuous component represents the miles of travel (Bhat and Sen, 2006; Bhat, Sen and Eluru, 2009; Eluru et al., 2010, Spissu et al. 2009).

2.5 Explanatory variables

This subsection is devoted to a review of a diversity of explanatory variables for the vehicle type choice model by comparing and analyzing the effects of key attributes in the previous literatures as well as discussing and finding some interesting points. Several different types of variables were considered as determinants of vehicle type choice of household. These can be typically sorted into four groups; household demographics, residence location attributes, and vehicle attributes, and primary driver characteristics.

Due to the ultimate objective of most studies with this topic is to discover the influence of automobiles on households and comprehend their relationship no matter what they specially focus on, some variables in household demographics are used in common. That is, the fundamental factors such as household size and income are essentially included in almost all models and most of these findings are consistent with expectations. Intuitively, households with high income are more likely to choose relatively expensive vehicles such as luxury cars, SUVs, and less likely to choose pickups (Bhat and Sen, 2006; Lave and Train, 1977; Adjemian, Lin and Williams, 2010; Eluru et al, 2010; Choo and Mokhtarian, 2004; Spissu et al, 2009). Even though only the selection of van is somewhat divided among the models, this presumably appears to be a matter of classification regarding that such households prefer vans in the models which consider minivans and vans are in the same category while those are unlikely to choose vans in the counterparts. In addition to that, the majority of the models have a general agreement that households with more members tend to prefer vans suggesting that households needs bigger vehicles as their size grows (Adjemian et al., 2010; Eluru et al., 2010; Potoglou, 2008; Spissu et al., 2009). However, there might be much room for consideration that such this corresponding relationship between the size of household and the vehicle will still be effective under the analysis of secondary vehicles. Assuming that households already own the so-called family car, it could be more rational to choose a small auto considering practicality or economic efficiency when they buy an extra one. For the matter of the number of vehicles in households, Lave and Train (1977) showed that the variable indicating the households with more than two cars is positively associated with small cars while other studies presented the number of vehicles is positively associated with pickup trucks (Adjemian et al., 2010; Cao et al., 2006; Potoglou, 2008).

Besides, there are a bunch of variables frequently referred on household demographics in the previous models. Bhat and Sen (2006), Eluru et al. (2010), Cao et al. (2006), Spissu et al. (2009), Choo and Mokhatarian (2004) successfully showed that the number, at least the presence, of children in households is positively associated with SUVs and/or minivans as chosen vehicles. The number of males has a positive sign for pickup or large vehicle except for van (Bhat and Sen, 2006; Cao et al., 2006); conversely, females are less likely to choose pickup trucks than any others (Eluru et al., 2010; Choo and Mokhatarian, 2004). Similar with female, households with high education level appear to be reluctant to choose pickup trucks (Choo and Mokhatarian, 2004; Cao et al., 2006; Potoglou, 2008). The variable referring to the number of workers is associated with a negative coefficient on minivans, midsized or large sedans which are called as a family car. From an occupational point of view, Choo and Mokhatarian (2004) presented that the sales variable has a positive sign for mid-sized and luxury cars. Furthermore, Mohhammadian and Miller (2003) presented that managers are more likely to buy midsize vehicles, large vehicles, or SPVs, while professionals prefer to buy subcompact vehicles.

According to Tardiff (1980), in order to estimate the coefficients of a choice model, there must be variation in the independent variable and the vehicle characteristics used as explanatory variables are important because the choice and definition of vehicle attributes affects the estimation of the model. For this reason, a wide variety of variables are used in the literature to explain the difference in utility and these include horsepower, luggage space, length, width, fuel efficiency, engine size, number of seats, and weight. However, many of these variables are highly correlated with each other and this high multicolinearity between variables might cause problems with the identification of the influence of vehicle attributes on the utility model. To solve this problem, Lave and Train (1977) used several interaction terms between car characteristics and household demographics in the equation such as the family size and car size, age and performance, income and weight, and so on. As a result, they drew a quiet reasonable results that young people are much more influenced by an increase in performance than older people. On the other hand, Mohammadian and Miller (2003) utilized principle components analysis which enables to reform a set of observed independent variables into a new set of independent variables. They created two factors, the vehicle performance factor which consists of vehicle weight, engine displacement, and fuel intensity and the space factor which consists of size and space. Unlike the performance factor is positive for all alternatives, the space factor was found that only SPV and van have a positive sign.

Recently, reflecting great attention shown to the interaction between built environment and travel behavior, a number of residence location variables are used to answer some arising questions such as the influencing aspects of neighborhood design on vehicle type choice, the role of residential preference in vehicle type choice. In most studies, population density is negatively associated with pickup trucks and positively associate with compact cars in logical way (e.g. Bhat and Sen, 2006; Fang, 2008; Potoglou, 2008). Households in urban area are more likely to choose small or luxury cars (Eluru et al., 2010), while households in suburban area are more likely to choose minivans and pickup trucks (Potoglou, 2008). Eluru et al. (2010) and Spissu et al. (2009) presented that as land use mix increase, the likelihood of choosing pickup trucks decreases as consistent with expectations.

CHAPTER 3 METHODOLOGY

This chapter presents the methodology developed to study the factors influencing mode choice in households of an appropriate model. This chapter is organized into three main sections. The first section covers overall research design. The second elaborates on sources of data and characteristics of the data. The third discusses the statistical method.

3.1 Research Design

The purpose of this paper is to examine the relationship between household composition and a type of newly purchased vehicles by analyzing the multinomial logit model with regard to the recently purchased vehicles as a dependent variable. Because the scope for this study is a national level, the data for this analysis are geographically restricted to national sample only and all the add-on data are excluded by design. The newly purchased vehicles are defined as vehicles received within one year starting from March 2001.

3.2 Data source

3.2.1 Description of NHTS 2001

The National Household Travel Survey (NHTS) administrated by the Unites States Department of Transportation (USDOT) is an official source of the national data on the travel behavior of the American civilians for the sake of transportation planners and policy makers and has been widely used by means of quantifying the travel behaviors, relating travel behaviors to household demographics or residence environments. The 2001 NHTS updates information gathered in prior Nationwide Personal Transportation Surveys (NPTS) and the American Travel Survey (ATS). The 2001 NHTS was conducted using Computer-Assisted Telephone Interviewing

(CATI) technology over the 14-month period from March 2001 through May 2002. This is the unique data available at the national level, which includes the demographics of households, household members, the vehicles owned by the households and several residence related environmental information. Thus, the 2001 NHTS enables transportation planners or sociological scholars to study the vehicle ownership, vehicle type choice, fleet combination, and utilization pattern through linking and combining a number of useful variables so that the analysis provides a better understanding of a trend or preference for the type of vehicles in a certain circumstances with respect to transportation safety, energy consumption, environmental impact, and general health.

3.2.1 Original Data Sets

The 2001 NHTS data sets consist of four different data files; household file, person file, vehicle file and trip file. The household file, data collected once for the household (one record per household), contains variables describing the household characteristics and members that include the socio-demographic, geographic housing characteristics and age, income, driver status and working status of all household members. The person file, data items collected once for each interviewed household member (one record for each completed person interview) contains variables describing characteristics of persons, worker information, driver information and travel attitude toward current transportation system. The trip file contains the purpose, mode, distance and duration, origin and destination characteristics of all the daily trips (248,517) made by all the persons in the trip file. The vehicle file, data relating to each of the household’s vehicle (one record per household vehicle) contains variables describing the vehicle attributes such as make, model, year, and type, annual miles and information of primary drivers. The sample was comprised of 26,038 national samples and 43,779 households in nine Add-on areas, for a total of

69,817 interviewed households. Among these records, the national sample only was chosen for the study area because this sample fulfills both the aim of the study and the size needed to be modeled. Specifically, 26,038 households consist of 60,282 household members who own 53,275 vehicles as a total.

3.3 Data preparation

3.3.1 Dependent variable

The 2001 NHTS data set includes the variable ‘Vehicle type’ which divides the vehicle into eight categories: car, van, SUV, pickup truck, other truck, recreational vehicles, motorcycles and other. One might, however, argue that this classification is inappropriate or vague in this analysis in that a car which is usually recognized as an automobile is too broad to examine households’ vehicle type choice. Considering that households more than 50% own at least one automobile (56.8% from 2001 NHTS), the more specified and detailed classification in car class is recommended for better results. For this reason, automobile class was reformed and divided into 5 categories by referencing the Wikipedia automobile classification scheme with make, model and year variables in original vehicle file; subcompact, compact, mid-size, full-size and sports car. After excluding a class of other, consequently a total of eight categories were built in the vehicle type variable; subcompact, compact, mid-size, full-size, SUV, pickup truck, van (including minivan) and sports car. This classification is almost consistent with *Consumer Reports* classification system and, similar to *Consumer Reports*, most schemes of vehicle classification first group vehicles by size, and then special categories such as sports, SUV and pickup are added.

Table 1. Distribution of household vehicles recently acquired (less than 12 months)

Type of vehicles	Subcompact	Compact	Mid-size	Full-size	SUV	Pickup	Van	Sports	Grand Total
Count	286	1135	1307	561	1184	1305	673	200	6651
Percentage share	4.30%	17.07%	19.65%	8.43%	17.80%	19.62%	10.12%	3.01%	100.00%
Rank	7	4	1	6	3	2	5	8	-

3.3.2. Household characteristics

A number of household characteristics used for this analysis include the household income, size, location of the households, the number of particular household member, several household life cycles, the number of workers, the number of drivers, the number of owned vehicles, and a type of housing units.

Income variable is derived from total annual income for the household and classified into six groups which are consistent with Census 2000 classification. Life cycle variable is classified into ten groups and each code indicates specific household composition. In this analysis, each code is transformed as a dummy variable indicating corresponding household composition. Home ownership variable indicates whether a housing unit is owned or rented and is classified into four groups. Home type variable indicates a type of housing unit and is classified into six groups. Three interaction variables are applied in this analysis. Those are a ratio of household members over sixteen years old to the number of household vehicles, a ratio of household members over sixteen years old to the number of workers in the household, and a ratio of number of workers to the number of vehicles in the household.

3.3.3. Residence location attributes

Several household location variables are considered in this specification, and this includes household census division and region, population density (persons per square mile), size of MSA (Metropolitan Statistical Area), and urban/rural indicator.

Those two household census variables are derived from the geocode for the household's home address and used for the 2000 Census Division source. The division variable is divided into four groups; Northeast, Midwest, South, and West and the region variable is classified into nine groups. Population density variable was also derived from 2000 Census data and divided into 8 groups. MSASIZE variable is a population size category of MSA for household's home address and derived as 6 categories.

3.3.4. Vehicle attributes

As stated above, this paper aims to analyze the effect of household composition on the choice of recently purchased vehicle. However, the information of previously owned vehicles are also highly recommended for more precise outcome. This model includes not only attributes of recently purchased vehicles but also attributes of existing vehicles separately. This includes age of vehicles, annual vehicle-miles traveled, miles per gallon, and fuel cost. Fuel cost variable was estimated in cents per gallon in local area, based on the sample household location.

3.3.5. Primary driver characteristics

Using person IDs of primary driver from the original dataset, the characteristics of primary driver such as age, gender, employment status, education level, the number of vehicles for which the subject is the primary driver and several attitudes toward travel problems are selected from the common person file.

The education variable indicates the education level of the household respondent and is classified into 8 different groups. The employment status variables provide copious information as an example of the four major occupational categories, full/part time information, job requiring driving a motor vehicle, and transportation mode to work last week. Respondents' attitudes toward travel problems were randomly surveyed through the 2001 NHTS in the form of scale system where 1 means it is not a problem at all, and 5 means it is the worst travel problem. Among ten variables, only two variables were included in this model due to the matter of sample size; highway congestion issue and price of gasoline issue.

Table 2. Sample characteristics of vehicle type categories

Type of vehicles	**Avg. of HH income**	**Avg. of educ. level**	**Avg. of HH size**	**Avg. of HH vehicles**	**Avg. of VMT**	**Avg. of vehicle age**	**Avg. of driver's age**
Subcompact	3.88	3.92	2.91	2.66	14,169.53	6.50	36.40
Compact	3.94	3.74	2.98	2.64	13,560.40	5.77	39.68
Mid-size	3.97	4.12	2.78	2.48	14,852.36	5.63	44.30
Full-size	4.12	3.98	2.73	2.66	14,217.47	6.04	**51.22**
SUV	**4.66**	**4.33**	3.09	2.79	14,788.63	3.85	42.11
Pickup	4.07	3.26	2.98	3.03	**15,307.49**	6.29	42.96
Van	3.96	4.06	**3.49**	2.64	15,304.30	4.86	45.49
Sports	4.20	3.81	2.97	**3.21**	11,655.34	**9.61**	36.27
Grand Total	**4.12**	**3.89**	**2.99**	**2.73**	**14,576.52**	**5.58**	**43.01**

3.4 Modeling methodology

This section describes the fundamental mathematical framework, methodology of estimation and application of the model structures of multinomial logit modeling frameworks that are used for the analysis.

3.4.1 Multinomial logit modeling

The dependent variable, a vehicle type recently purchased, consists of nine mutually exclusive categories, so a multinomial logit model which is most widely used for estimating discrete choice models in this field is developed for vehicle type choice.

3.4.2 Random utility maximization

Discrete choice models are based on random utility maximization theory. The random utility theory which assumes that the decision-maker's preference for an alternative is captured by a utility, an indicator of value to an individual. The utility maximization rule is mentioned that an individual will select the alternative which maximizes his or her utility out of the available alternative set. Probability of choice 'i' is equal to the probability that the utility of alternative 'i' is greater than or equal to the utilities of all other alternatives in the choice set.

(or) $P(i|Cn) = Pr\ [U_{in} \geq U_{jn}\ ,\ \text{all}\ j \in Cn]$

where, Cn is the set of alternatives available for the nth choice maker (choice set).

The utility maximization rule also implies there is no uncertainty in the individual's decision process, that is, the decision maker is certain to choose the highest ranked alternative under the observed condition. However, the analyst is unlikely to know specific circumstances of the individual's decision or the individual may have incomplete or incorrect information about the attributes of alternatives. To take account of the lack of information on the part of analyst, random utility models are applied by introducing an error term in the utility of each alternative. Thus, the utility of an alternative 'U_i' is split into a deterministic term 'V_i' and a random error term 'ε_i'. Then,

$$P(i|Cn) = Pr\ [V_{in}+\varepsilon_{in} \geq V_{jn}+\varepsilon_{jn},\ \text{all}\ j \in Cn]$$

The deterministic utility V_{in} expressed as linear function of explanatory variables is given by:

$V_{t,i} = V(S_t) + V(X_i) + V(S_t, X_i)$

where $V_{t,i}$ is the deterministic utility of alternative *i* for individual *t*,

$V(S_t)$ is the utility associate with characteristics of individual *t*,

$V(X_i)$ is the utility associated with attributes of alternative *i*,

$V(S_t, X_i)$ is the utility which results from interaction between attributes of alternative *i* and characteristics of individual *t.*

Alternative distributional assumptions about the joint probability distribution of the full set of disturbances (error terms) yield different probabilistic choice models. Assumption that the disturbances are 'Gumbel' distributed leads to the multinomial logit model with the 'Independence of Irrelevant Alternatives' (IIA) property. Multinomial logit model gives the choice probabilities of each alternative as a function of the systematic portion of the utility of all the alternatives. The general expression for the probability of choosing an alternative 'i' (i=1,2,…,J) from a set of J alternatives is:

$$Pr(i) = \frac{exp(V_i)}{\sum_{J=1}^{J} exp(V_j)}$$

Where $\Pr(i)$ is the probability of the decision maker choosing alternative *i* and

V_i is the systematic component of the utility of alternative *j.*

3.4.3 Independence of irrelevant alternatives (IIA) property

The multinomial logit model is perhaps one of the most widely used discrete models in several fields including transportation area. However, violation of Independence of irrelevant alternatives (IIA) property may lead unrealistic results in a certain choice situation. The IIA is a property which implies that the relative odds between two alternatives are the same no matter what other alternatives are available. In reality, there are a number of outcomes that violate the IIA property such as (i) predicting the outcome of multicandidate elections or (ii) when the choice made by humans is involved. According to Lave and Train (1977), if alternative specific constants are included as explanatory variables in MNL model, then these constants partially "correct" for violations of the IIA property. However, we still must conduct a test of IIA to prevent that will results in erroneous predictions of choice probabilities in a case of vehicle type choice analysis. There are a number of statistical tests that are conducted to test for IIA violations. One of the basic types of tests is the Hausman-McFadden test on a subset of alternatives (Hausman and McFadden, 1984) which compares the estimates from the full and restricted model. In this test, if IIA holds, the two sets of estimates should not be statistically different. If the estimates are significantly different, the assumption of IIA is rejected, and consequently, the MNL model is inappropriate.

Since this limitation of MNL model comes from the assumption of the independence of error terms in the utility of alternatives, different models which derived through the different assumptions of error distributions of alternative utilities would not obey the IIA property. And, the Nested Logit model which is characterized by grouping subsets of alternatives that are more similar with each other is the simplest and widely used model among several possibilities. In the nested logit model, when two alternatives are in the same nest, the IIA property holds since the ratio of probabilities is independent of the existence of other alternatives. However, if two alternatives are placed in different nests, then the IIA property no longer holds.

3.4.4 Estimation

Maximum Likelihood method is used to estimate the parameters (coefficients) of multinomial logit model. Maximum likelihood estimates are the value of the parameters for which the observed sample is most likely to have occurred. The likelihood function for a general multinomial choice model is

$$L(\beta)=\prod_{n=1}^{N}\prod_{i\in C_n} P_n(i)^{y_{in}}$$

where, N denotes the sample size,

y_{in} =1 if i is chosen by decision maker n,

=0 otherwise,

where, for the linear in parameters model:

$$P_n(i)=\frac{\exp(X_{in}'\beta)}{\sum_{j'}\exp(X_{jn}'\beta)}$$

Taking the logarithm provides the log-likelihood function as followed:

$$L=\sum_{n=1}^{N}\sum_{i\in C_n} y_{in}(\beta x_{in}-\ln\sum_{i\in C_n} e^{\beta x_{jn}})$$

Now, we can solve to estimate the parameters which maximize L.

CHAPTER 4 ESTIMATION AND RESULTS

In the previous section, we verified the separated average values of several variables represent household's vehicle type preference to some degree. In reality, however, the relationship of one variable to vehicle type can be affected by other variables. The relationship of one variable to vehicle type may be significant in isolation, but disappear or diminish in importance when the impact of a related variable is accounted for. Conversely, a simple pairwise relationship may appear insignificant due to the counteracting influences of other variables, but become significant when those other variables are controlled for. Thus, a disaggregate discrete choice model was developed to estimate the probability of choosing each vehicle type based on the collective effect of factors such as vehicle attributes, household characteristics, residence location variables, and primary driver's characteristics. These parameters are estimated for seven alternatives (i.e., subcompacts, compacts, mid-size sedan, full-size sedan, pickup truck, SUV, or van), and the sports car is considered as the base alternative. The coefficients of a certain vehicle type in the multinomial logit model indicate the propensity to acquire a corresponding vehicle type with positive values indicating an increase in the probability and negative values indicating a decrease in the probability in comparison with the probability of choosing sports car. The statistical software package NLOGIT 4.0 was used to estimate the model.

Based on the initial model specification, statistically insignificant variables were eliminated in stages, and then variations on the subsequent specifications were tested to obtain a final model. When initial estimations suggested that some variables had a similar impact on more than one vehicle type, we then constrained those coefficients to be equal for the sake of parsimony and to increase the degrees of freedom available in the sample. Through this procedure, the final model with seven alternative specific constants (ASCs) and 73 alternative-specific variables (ASVs), representing 13 different variables, was developed. As shown in Table 3, most of explanatory variables were statistically significant and conceptually interpretable. Additionally, as a goodness of- fit test statistic, the Chi-Square Test value ($\chi 2$) of 6548.05 shows that the final model significantly differs from the equally likely model (in which all coefficients are equal to zero) at $\alpha << 0.005$. The adjusted ρ^2 value of the final model is .294 which means this model can explain about 30% of the information in the data.

Turning to the explanatory variables in the final model, the estimation results are similar with the sample characteristics of vehicle type categories shown at table 2. In fact, this analysis had started with plentiful variables and hypotheses in expectation of being verified rigorously from a various angle, however some interesting variables, particularly existing vehicles' type variables and several life cycle dummy variables, were not successfully explained statistically by this model. Probably, the failure of explaining effects of the existing vehicle type variables on the household purchasing decision of vehicles by type can be interpreted in that the household vehicle acquisition is more influenced by the other attributes such as residence location environment, household demographics, and primary driver's characteristics rather than existing vehicles' type. By the same token, the life cycle dummy variables, for instance, two adults and no children, two adults and the youngest children under 0-5 and two retired adults and no children, are too specified and detailed to certify the vehicle type choice and more general household demographics variables such as household size, the number of workers, and the number of drivers, are much more appropriate to explain the choice. The annual total cost of gas variable was also used for this modeling, however, the results were not only significant statistically but also counter-intuitive. This might be caused from the correlation with annual vehicle miles traveled variable which was already applied to the model.

Three vehicle attributes, vehicle miles traveled, vehicle age and miles per gallon, are all significant in this model. The annual vehicle miles traveled variable is positively associated with pickup trucks, SUVs, and van, while being negatively associated with sports cars and subcompact cars. This indicates that most of routine travels in households are made from pickup trucks, SUVs, and Vans, and the sports car and subcompact cars are considered for leisure or short trip. As expected, households with drive a lot tend to choose large vehicles. The fact that vehicle age variable shows similar results with vehicle miles traveled describes that households become more sensitive when acquiring vehicles for a main transportation role. The probability of buying subcompacts and compacts increases when mile per gallon is better while pickup trucks and SUVs are negatively associated.

Many findings related with household demographics are consistent with expectations and with the large body of literature that speaks to the types of vehicles that households acquire in the context of their socio-demographic characteristics. With regard to other explanatory variables, the final model showed that household characteristics significantly affect vehicle-type choices. Households are more likely to choose subcompact cars when vehicle availability per household member is higher. This means that households having high volume of vehicles in comparison with household size tend to use those vehicles not for family purpose but for individual purpose. Households with more members tend to prefer vans, and avoid sports cars. The effect of annual household income indicates that high income households are likely to own sports cars, full-size sedans, and SUVs. This reiterates the fact that SUVs belong to higher income households. Such households have a lower baseline preference for pickup trucks and minivans. Proportionally higher level of workers to household size makes it least inclined towards buying vans. This is probably because each worker in the household might necessarily need a separate vehicle for work. The effect of number of workers in the household is not significantly different for pickup trucks and SUVs which have higher coefficient value and this indicates that these vehicles are mainly chosen for commute cars. Households with more number of drivers tend to have higher propensity of choosing sports cars and subcompact cars and lower probability of choosing vans indicating that the more licensed drivers in households, the larger volume of individual purpose types of vehicles. Household having children under 8 years old shows an increasing probability of buying a van while showing the lowest likelihood of choosing subcompact cars. Other household-level variables tested included housing ownership type as well as several specified household composition types, however, none of the above significantly improved the model.

Only one resident location variable, population density which is derived from the number of people per square mile, was used for this model. To magnify the effect of this variable, the number of people per square mile is divided by the number of household vehicles. As expected, higher local population density increases the likelihood of subcompact cars purchase, while having the opposite effect on pickup trucks.

Lastly, three variables for primary driver characteristics are applied. Households with higher aged primary drivers apt to choose either full-size sedan or vans while younger primary drivers are likely to choose sports car, and this finding is consistent with expectations as one would expect households with such members to prefer the larger vehicles (Choo and Mokhtarian, 2004). A higher number of male drivers is associated with an inclination to choose pickup trucks.

The value of coefficient for the primary driver's education level points out primary driver with higher level of education level tend to choose SUVs and sports cars similar with the results of household income variable.

Table 3. Estimation results of households' vehicle-type choice

Variable	Vehicle type							
	Sports (Base)	**Sub compact**	**Compact**	**Mid-size Sedan**	**Full-size Sedan**	**Pickup truck**	**SUV**	**Van**
CONSTANTS	-	-10.120 (-10.14)	-3.721 (-4.311)	-5.236 (-5.969)	-4.207 (-4.715)	-1.254 (-1.285)	-2.258 (-2.313)	-5.182 (-5.093)
VEHICLE ATTRIBUTES								
Vehicle miles traveled	-	-	0.2062 (2.839)	0.9991 (12.464)	0.9991 (12.464)	2.1874 (21.499)	2.2158 (21.976)	1.7712 (16.928)
Vehicle age	-	-0.0307 (-2.215)	-0.0307 (-2.215)	-0.0307 (-2.215)	-0.0307 (-2.215)	-0.1455 (-8.798)	-0.1537 (-9.106)	-0.1207 (-6.719)
Miles per gallon	-	0.3340 (-12.347)	0.1074 (-4.445)	-0.1464 (-5.756)	-0.2914 (-10.38)	-0.8318 (-25.517)	-0.8522 (-26.347)	-0.5756 (-17.909)
HOUSEHOLD CHARACTERISTICS								
Vehicle availability per household member	-	0.5399 (2.935)	0.2066 (1.751)	0.0478 (0.439)	0.1352 (1.302)	-	0.1100 (1.298)	0.0867 (0.77)
Household size	-	0.3375 (3.435)	0.3375 (3.435)	0.2046 (1.986)	0.3583 (3.229)	0.2871 (2.605)	0.3375 (3.435)	0.7430 (6.81)
Household income	-	-0.0420 (-2.182)	-0.0292 (-2.522)	-0.0321 (-3.111)	-	-0.0700 (-6.18)	-	-0.0668 (-5.178)
Household driver count	-	-	-0.1252 (-1.273)	-0.1252 (-1.273)	-0.3574 (-2.762)	-0.4555 (-3.474)	-0.3326 (-2.81)	-0.6818 (-5.173)
Ratio of workers to household size	-	0.3062 (0.946)	0.0821 (0.428)	0.1675 (0.967)	-	0.6545 (3.032)	0.5179 (2.506)	-0.4066 (-1.783)
Children under 8-year-old	-	-0.2953 (-0.913)	-0.0931 (-0.355)	0.1598 (0.619)	-	0.6760 (2.365)	0.4190 (1.505)	0.9545 (3.315)
RESIDENCE LOCATION ATTRIBUTES								
Ratio of people per square mile to vehicles	-	0.2654 (5.096)	0.1516 (4.242)	0.1249 (3.771)	0.0337 (1.056)	-0.2269 (-5.74)	0.0337 (1.056)	-
PRIMARY DRIVER CHARACTERISTICS								
Primary driver's age	-	0.0142 (1.797)	0.0223 (3.512)	0.0353 (5.593)	0.0618 (9.187)	0.0259 (3.698)	0.0241 (3.441)	0.0621 (8.346)
Male primary driver	-	-0.4485 (-1.983)	-0.7125 (-3.977)	-0.6772 (-3.788)	-0.3748 (-1.941)	1.4824 (7.255)	-0.7520 (-3.874)	-0.8186 (-4.102)
Primary driver's education level	-	-0.0723 (-2.634)	-0.0723 (-2.634)	-0.0208 (-0.842)	-0.0581 (-2.135)	-0.1718 (-6.928)	-	-0.0150 (-0.532)
Log-likelihood at zero							-11956.79	
Log-likelihood at constants							-11089.26	
Log-likelihood at convergence							-7815.235	
Rho-Squared w.r.t Zero							0.346	
Rho-Squared w.r.t Constants							0.295	
χ_c^2							96042.162	

The number in parentheses indicates the t-value of that coefficient (at a level of α=0.05 a critical t-value = 1.96).

CHAPTER 5 CONCLUSIONS AND FUTURE WORK

This paper utilized the 2001 National Household Survey (NHTS) data to study the patterns of household vehicle ownership by type in the United States. Recently purchased eight vehicle types; subcompact cars, compact cars, mid-size sedans, full-size sedans, pickup trucks, SUVs and vans are distinguished based on what factors and how such factors affect household decision to acquire a certain type of vehicle in terms of four different angles; household demographics, residence location, primary drivers' characteristics and vehicle attributes.

Prior to developing a model for estimating vehicle type choice, manipulation and refinement of the original 2001 NHTS data was performed ahead. Even though this dataset proved to be a rich source for such an analysis, some variables are required to be recoded for better results. For instance, considering that households more than 50% in U.S. own at least one automobile (56.8% from 2001 NHTS), the automobile/car/station wagon class on vehicle type variable provided by original 2001 NHTS data is highly recommended to be more specified and divided into some categories. For this reason, automobile class was rebuilt with new 4 categories (subcompact, compact cars, mid-size sedans, full-size sedans and sports cars) by referencing the Wikipedia automobile classification scheme with vehicle make code, model code and year variables in original vehicle file. Besides, throughout several steps of modification of data a total of 7,500 observations with 33 variables were used for the analysis.

A disaggregate discrete model (specifically, a multinomial logit model) for the type of recently acquired vehicle to estimate the effect of diverse explanatory variables on the probability of choosing each vehicle type provided insights into the recent trends in the vehicle ownership patterns. The final model (with sports cars as base case) includes 13 significant alternative-specific variables representing household demographics, vehicle attributes, residence location and primary drivers' characteristics together with the eight alternative-specific constants.

To sum up the results of the estimation, subcompact cars are most frequently selected from households with higher vehicle availability per household member, with higher number of licensed drivers, or residing in high population density area while being least selected from households with drive a lot or having children under 8 years old. The model does not suggest such a distinct preference for compact cars, however majority of the value of coefficients are either located between the one of subcompact cars and mid-size sedans or very similar with those two types, and this outcome is certainly reasonable intuitively. Mid-size sedans also take a natural attitude among cars as this type represents most popular and universal position among them. Full-size sedans are being chosen by households with higher income or aged primary driver. Households with drive a lot or having high number of workers are more likely to choose pickup trucks while household with lower vehicle availability per household member, lower income, lower population density or more male primary driver shun choosing pickup trucks. SUVs are being acquired by households with drive a lot, with higher income or having higher education level of primary. Lastly, Vans are being commonly chosen by households with more members, having more children aged under 8, or having older primary drivers while being least chosen by households with lower income, lower number of licensed drivers, lower ratio of workers to household size, or younger primary drivers.

This model is not only statistically significant but also consistent with expectation and the results of the previous leading studies, however, more developments are still necessary. First, although this model was developed for the sake of the analysis of temporal trends of vehicle ownership, using data collected in the year 2001 seems to be behind the fashion. Such study with recent dataset can illuminate much about the market share trends and inform policy in a better way. Second, unfortunately some interesting variables such as history dependency in vehicle acquisition and specified household member composition initially considered were not successfully embodied through this model due to a lack of statistical significance. It might be explained that those variables are either too minute to apply or not very influential, however, further analysis with other methods is recommended for better output.

REFERENCES

Adjemian, M. K., Lin, C. Y. C., & Williams, J. (2010). Estimating spatial interdependence in automobile type choice with survey data. Transportation Research Part a-Policy and Practice, 44(9), 661-675.

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

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

Cao, X. Y., Mokhtarian, P. L., & Handy, S. L. (2006). Neighborhood design and vehicle type choice: Evidence from Northern California. Transportation Research Part D-Transport and Environment, 11(2), 133-145.

Chiou, Y. C., Wen, C. H., Tsai, S. H., & Wang, W. Y. (2009). Integrated modeling of car/motorcycle ownership, type and usage for estimating energy consumption and emissions. Transportation Research Part a-Policy and Practice, 43(7), 665-684.

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

Choo, S., Mokhtarian, P. L., University of California Davis. Institute of Transportation Studies., & DaimlerChrysler. (2002). The relationship of vehicle type choice to personality, lifestyle, attitudinal, and demographic variables. Davis, Calif.: Institute of Transportation Studies University of California at Davis.

DeJong, G. (1996). A disaggregate model system of vehicle holding duration, type choice and use. Transportation Research Part B-Methodological, 30(4), 263-276.

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

Feng, Y., Fullerton, D., Gan, L., & National Bureau of Economic Research. (2005). Vehicle choices, miles driven, and pollution policies NBER working paper series working paper 11553 Retrieved from http://papers.nber.org/papers/W11553

Golob, T. F., Bunch, D. S., & Brownstone, D. (1997). A vehicle use forecasting model based on revealed and stated vehicle type choice and utilisation data. Journal of Transport Economics and Policy, 31(1), 69-&.

Golob, T. F., Kim, S. Y., & Ren, W. P. (1996). How households use different types of vehicles: A structural driver allocation and usage model. Transportation Research Part a-Policy and Practice, 30(2), 103-118.

Lave, C. A., & Train, K. (1979). A Disaggregate Model of Auto-Type Choice. Transportation Research Part a-Policy and Practice, 13(1), 1-9.

Mannering, F., & Winston, C. (1985). A Dynamic Empirical-Analysis of Household Vehicle Ownership and Utilization. Rand Journal of Economics, 16(2), 215-236.

Mannering, F. L., & Train, K. (1985). Recent Directions in Automobile Demand Modeling. Transportation Research Part B-Methodological, 19(4), 265-274.

Mohammadian, A., & Miller, E. J. (2003). empirical investigation of household vehicle type choice decisions. Traveler Behavior and Values 2003(1854), 99-106.

Musti, S. (2009). Evolution of the household vehicle fleet anticipating fleet composition, plug-in hybrid electric vehicle (PHEV) adoption and greenhouse gas (GHG) emissions in Austin, Texas Retrieved from http://repositories.lib.utexas.edu/bitstream/handle/2152/ETD-UT-2009-12-681/MUSTI-THESIS.pdf

Potoglou, D. (2008). Vehicle-type choice and neighbourhood characteristics: An empirical study of Hamilton, Canada. Transportation Research Part D-Transport and Environment, 13(3), 177-186.

Silva-Risso, J., & Ionova, I. (2008). Practice prize winner - A nested logit model of product and transaction-type choice for planning automakers' pricing and promotions. Marketing Science, 27(4), 545-566.

Spissu, E., Pinjari, A. R., Pendyala, R. M., & Bhat, C. R. (2009). A copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. Transportation, 36(4), 403-422.