VEHICLE TYPE CHOICE AND DEMOGRAPHIC RELATIONSHIPS: AN APPLICATION TO THE WINDSOR REGION, ONTARIO, CANADA

Georgiana Madar and Hanna Maoh
University of Windsor, Ontario

Introduction

Car usage is very prevalent in North America, especially in areas that have strong historical ties to the automotive manufacturing industry, such as the Windsor–Essex region in Ontario, Canada. Despite this local affiliation with the domestic automotive industry in Windsor, acquisitions of vehicles for personal use also include imported brands. Investigating the factors that influence vehicle type choice is an important exercise when attempting to study auto ownership and usage in urban areas. In the context of Windsor, Ontario, investigating this process can help determine whether there is an apparent difference in the choice between domestic or imported brands especially when considering the acquisition of vehicles of different sizes.

To this end, the overall objective of this study was to develop a statistical model for explaining the socio-economic factors that influence the choice decision of domestic versus imported vehicles. However, to devise a sensible model, domestic and imported vehicles were further bifurcated into cars and trucks. Cars are used to resemble smaller vehicles such as compact cars, sedans, regular vehicles, etc., while trucks are used to represent larger-size personal-use vehicles, including Sport Utility Vehicles (SUVs), minivans, and pickup trucks. Consequently, the choice decision is modeled among four exclusive alternatives: (1) domestic cars, (2) imported cars, (3) domestic trucks, and (4) imported trucks.

Vehicle ownership data for this paper were acquired from R. L. Polk and Co for the Windsor-Essex region. The data included individual records of all registered vehicles in the region. Each record was geo-referenced to the census tract level and had information about the model year, make and model type of the vehicle. Furthermore, the records included information about the fuel type, Gross Vehicle Weight (GVW) and vehicle category (i.e. domestic cars and trucks versus imported cars and trucks). Fuel economy information was added to records of the Polk dataset. Furthermore, several variables from the 2006 Canadian census database were used to examine if certain demographic variables influence the choice of domestic and imported vehicles.

A number of discrete choice models are formulated, specified and estimated to explain the choice decision process. The latter is achieved by testing hypotheses regarding the influence that different factors pertaining to different age groups, genders, ethnicities, level of education, and profession has on the choice decision outcomes. Furthermore, the analysis also tested hypotheses regarding the influence of some vehicle characteristics, namely, the model year and fuel efficiency on the choice decision process.

The remainder of this paper is organized into a number of sections. The next section provides an overview of the most recent literature on the topic. This will be followed by a section devoted for describing the modeling approach and postulated hypotheses. In section four results from three different discrete models: a multinomial logit model (MNL), a nested logit model (NL), and a mixed logit model (MXL) will be compared and discussed. Finally, the last section provides a conclusion to our study.

Literature Review

In recent years, a number of studies have been conducted to examine how various characteristics of decision makers influence vehicle type choice. For instance, Choo and Mokhtarian (2004) modeled vehicle type choice among nine different vehicle classes as a function of several demographic factors for the San Francisco Bay Area. They found that individual's attitude toward travel, personality, lifestyle, and mobility can explain vehicle type choice. A more recent study by Spissu et al. (2009) made use of land-use and demographic data along with vehicle characteristics information to study vehicle type choice, as well as the amount of vehicle utilization for the same study area. The results showed that land-use attributes and population density exert no significance on vehicle type choice. The study also concluded that unobservable factors could cause individuals to be more inclined to buy certain types of vehicles, as opposed to others.

In a similar vein, household characteristics and vehicle attributes were used jointly in the study of Roorda et al. (2000) to examine any relationships with the occurrence of vehicle transactions, vehicle type choice, the length of time that a vehicle is owned, and the degree of consumer loyalty to certain vehicle types or brands. The latter study analyzed user behaviour over time using a longitudinal survey. Uni-variate analysis was used to establish the nature of the relationship between household characteristics and vehicle purchasing patterns.

Other efforts such as the work of Potoglou and Kanaroglou (2006) modeled the number of vehicles owned by a household using personal characteristics, such as family structure, socio-economic factors, and dwelling attributes. A multinomial logit model, which was applied to data collected for Hamilton, Ontario, found that the decision of how many vehicles a household will own will be influenced by the stage of the household life cycle, socio-economic factors, mixed-density and mixed land use around residential areas.

Likewise, Paleti et al. (2012) modeled the number of vehicles acquired by a household using a multinomial probit model with spatial effects. The authors argue that spatial effects can be attributed to similar lifestyle choices that lead certain household with similar vehicle choice to reside in proximity to each other. The results from Paleti et al. (2012) reinforced some of the previous findings regarding the influence of demographic factors, vehicle attribute and trans-portation accessibility on vehicle ownership. However, the authors also found that spatial effects play an important role in explaining vehicle ownership.

Vehicle type choice has also been analyzed in past research by considering vehicle characteristics, instead of user attributes. For instance, Train and Winston (2004) examined the choice behavior of product line characteristics and dealerships by analyzing the factors giving rise to market shares of domestic and imported brands. It was determined that vehicle attributes which describe vehicle quality do have an effect on the market share of foreign and domestic vehicles. A study by Mohammadian and Miller (2003) also examined vehicle type choice with respect to distinct classes of vehicle types and was able to establish statistical relationship between individual characteristics, household attributes, and vehicle characteristics and vehicle type choice.

The statistical analysis in this study builds on these previous efforts by developing a model for the Windsor-Essex region. However, several model formulations will be tested and compared to enrich the statistical analysis.

Modeling Approach

Although the Polk dataset included the entire population (i.e. 199,436 individual vehicles) of registered vehicles in the region for the year 2010, only a subset of this data was used in our analysis. More specifically, vehicles with GVW greater than 2 were not included in the analysis. Typically, GVW greater than 2 pertains to trucks that weight over 10,000 lb. This resulted in a total of 193,752 vehicles that were classified as into four classes: car domestic (76,297 vehicles), car imported (25,407 vehicles), truck domestic (84,982 vehicles) and truck imported (7,065 vehicles). Furthermore, a 5% random sample from the population of 193,752 vehicles was selected to perform the modeling to reduce the computation time required to perform the model estimation. The drawn records were validated and were found to be a superior representative sample of the entire population.

Fuel efficiency data for each vehicle type was incorporated into the Polk data before drawing the 5% random sample. Data records from the Fuel Economy Guide were retrieved from the US Department of Energy (USDE, 2012). The fuel economy information was obtained for each vehicle make and model, by year. However, the retrieved information covered the period 1983–2010. Consequently, data points pertaining to vehicles made prior to 1983 in the 5% sample were dropped. This had no effect on the model, since the omitted data points comprised less than 1% of the entire dataset. The final dataset that was used in the model included 9,281 individual observations.

Variables from the 2006 Canadian census were used to describe the age, gender, ethnicity, citizenship, level of education, and profession of the people residing in the census tracts for which the registered vehicles belong. The census variables were added to the Polk dataset based on the census tract id associated with each Polk record. Although the Polk records are disaggregated, the appended census data, in general, represent zonal averages for each Polk record. NLogit 4.0 of the Limdep 9.0 statistical software was used to estimate the choice models. Here, the probability of choosing a particular vehicle type is modeled as a function of various demographic factors in the census tract for which a registered vehicle belongs. Also, vehicle characteristics such as fuel efficiency and size are introduced and controlled for in the model. Following Ben-Akiva and Lerman (1985), the probability of choosing specific vehicle type t which is found in census tract c is given as follows:

$$P_{t/c} = \Pr(\varepsilon \leq V_{t/c} - V_{q/c}) \text{ for all alternatives } q \neq t$$

where $V_{t/c}$ is the utility function associated with alternative t in census tract c and ε is an unobserved random term. Different assumptions regarding the probability density function of ε will lead to different model formulation. For instance, if ε follows the Gumble distribution, then the probability of choosing alternative t can be modeled via the well-known multinomial logit model. In this paper and as noted earlier, the choice decision is modeled among four exclusive alternatives: {domestic car (*CD*), imported car (*CI*), domestic truck (*TD*), and imported truck (*TI*)}.

Following Ben-Akiva and Lerman (1985), a utility function was specified for each of the four alternatives in the estimated choice model. The specification was based on our a-priori expectation with respect to various hypotheses that were inspired by the literature and real world observations. The validity of the tested hypotheses was based on the statistical significance and expected sign of the estimated parameters in the model. The final specifications of the four utility functions pertaining to the four alternatives are listed below. The definitions of the variables entering the specification of the utility functions are also defined in Table 1.

$$U(CD) = ASC_1 + \beta_A \times POP65 + \beta_C \times VOLDLINC + \beta_G \times ETH\_EURO + \beta_E \times CAN\_CIT + \beta_K \times AVG\_UHWY$$

$$U(CI) = ASC_2 + \beta_I \times OLD\_HED + \beta_C \times VOLDLINC + \beta_G \times ETH\_EURO + \beta_F \times ED\_MED + \beta_D \times VNEWHINC + \beta_K \times AVG\_UHWY + \beta_J \times AVG\_UCIT + \beta_H \times YOLDLINC + \beta_Z \times NEWCFE$$

$$U(TD) = ASC_3 + \beta_A \times POP65 + \beta_B \times M3544PR + \beta_E \times CAN\_CIT + \beta_D \times VNEWHINC$$

$$U(TI) = \beta_I \times OLD\_HED + \beta_B \times M3544PR + \beta_G \times ETH\_EURO + \beta_F \times ED\_MED + \beta_D \times VNEWHINC + \beta_J \times AVG\_UCIT + \beta_H \times YOLDLINC + \beta_Z \times NEWCFE$$

The above utility specifications are based on a number of hypotheses that will be tested. Firstly, we hypothesize that older people in the study area might be more prone to choose domestic vehicles, both cars and larger automobiles, and avoid purchasing imported brands. Many older people in Windsor were part of the automotive work force for several decades and might be loyal to the domestic brands.

However, this might not be the case of older people with higher levels of education. The latter sub-group of the older population belonged to the non-automotive work force and as such would be more prone to choose imported brands, both cars and larger vehicles. Also, it is likely that highly educated people retired from higher-earning jobs. Consequently, they will be able to afford the more costly imported vehicles. Furthermore, we hypothesize that men aged 35 to 44 would be more inclined to choose large vehicles, both domestic and imported brands. This is intuitive since vehicles such as pickup trucks are marketed toward and largely purchased by middle-aged men. Also, men in this age class are likely to have families with young kids in which an SUV or a minivan becomes a necessity.

Table 1: Description of Variables Used in Utility Functions

Variable	Description
ASC_1, ASC_2, ASC_3	Alternative specific constants
POP65	Population 65 years of age and older, including males and females, for each zone
ETH_EURO	Proportion of the population in each zone that is of European ethnicity
CAN_CIT	Proportion of the population in each zone that has Canadian citizenship
AVG_UHWY	Average highway fuel efficiency (mpg)
ED_MED	Proportion of the population in each zone with medical degrees
AVG_UCIT	Average city fuel efficiency (mpg)
M3544PR	Proportion of men aged 35–44 in each zone
VOLDLINC	Interaction term between very old vehicles (model year earlier than 1990) and low income (proportion of the population with annual income below $25,000)
OLD_HED	Interaction term between the number of males and females over 65 years of age in each zone and people with higher levels of education (Master's, PhD, medical degrees)
VNEWHINC	Interaction term between very new vehicles (model year later than 2006) and high income (proportion of the population with annual income above $60,000)
YOLDLINC	Interaction term between the VOLDLINC term (interaction between old vehicles and low income) and the number of males and females in each area aged 15–24
CNEWHINC	Interaction term between new vehicles (model year later than 2006) and average city fuel efficiency

We also postulate that drivers who have older vehicles and low annual incomes are more likely to choose domestic brands, both cars and larger vehicles. This is because domestic brands will be more affordable when compared to imported vehicles. Furthermore, we assume that young people with low income such as students might be more inclined to own older imported brands, both small and large vehicle types. This could be attributed to the fact that brands like Honda and Toyota produce vehicles that have been known to be more efficient and last longer, even if the model year is not recent. It is also possible that younger drivers will be driving their parents' vehicles, which can belong to the more expensive categories.

Ethnicity and citizenship status are also expected to play a role in the type of vehicle choice. Here, we hypothesize that people of European ethnicity are more likely to own small cars as well as imported brands. As such, three alternatives small domestic and imported cars and imported trucks might likely be chosen by this demographic group. Typically, small vehicles are prevalent in Europe so the expectation that people from European origins in our study area will be inclined to own such vehicles. Also, Europeans potentially prefer imported brands manufactured in Europe and Asia including small and large vehicle types.

On the other hand, we postulate that Canadian citizens will be more likely to purchase small and large domestic vehicles, while avoiding imported brands, other things being equal. This is intuitively correct, since these citizens would be more loyal to the local automotive industry that they are part of. We also hypothesize that people with medical degrees are more likely to choose imported brands, both small and large automobiles. Medical professionals normally earn higher-than-average incomes and will be able to afford purchasing expensive imported brands. In a more general case, drivers with new vehicles and high income are more likely to choose large vehicles and imported brands.

With respect to vehicle characteristics, we hypothesize that the choice of smaller cars will be associated with higher or better city fuel efficiency. This can be attributed to the fact that traveling within a city is more convenient in a small vehicle and as such city fuel efficiency becomes important. On the other hand, we hypothesize that better (higher) average highway fuel efficiency will be associated with imported brands, for both small and large vehicles. This can be attributed to the reputation of some imported brands as longer-lasting vehicles. Typically, drivers who spend more time driving long distances will be more prone to gravitate towards reliable and fuel-efficient vehicle. Many of the imported brands such as Honda, Acura, Nissan and Toyota are rated high in terms of their reliability and fuel efficiency for highway driving.

These utility specifications were used in three different models. First, a Multinomial Logit model was used. This is the simplest model, having a set of four choices, domestic car (CD), imported car (CI), domestic truck (TD), and imported truck (TI), as shown in Figure 1. The same four choices and utility specifications were used to run a full information Nested Logit model. To determine the appropriate nesting structure for this model, several options were tested, as shown in Figure 1. The last structure shown in Figure 1-e was chosen as it provided the best fit and inclusive value (IV) parameter was within the bound of 0 and 1 (i.e. in line with discrete choice theory). For comparison purposes, a third model was also estimated following a Mixed Logit approach, with 100 Halton sequence draws (Greene, 2002). The 100 Halton draws provided was chosen to achieve statistical stability in the estimated parameters.

Estimation Results

The results from the Multinomial Logit (MNL), Nested Logit (NL), and Mixed Logit (MXL) models are presented in Table 2, with all of the parameter coefficients and their corresponding t-statistics. It can be seen that the MNL had an adjusted ρ^2 value of 0.40. On the other hand, the model pertaining to nested structure #4 in Figure 1 had an adjusted ρ^2 value of 0.60, which is a considerable improvement over the MNL ρ^2 value. The estimated Inclusive Values (IV) for the NL model are less than one for each sub-nest, which is consistent with discrete choice theory. Finally, the results from the MXL model is an improvement over the MNL as far as the achieved adjusted ρ^2 value (0.51), but not as good as the fit produced by the Nested Logit model.

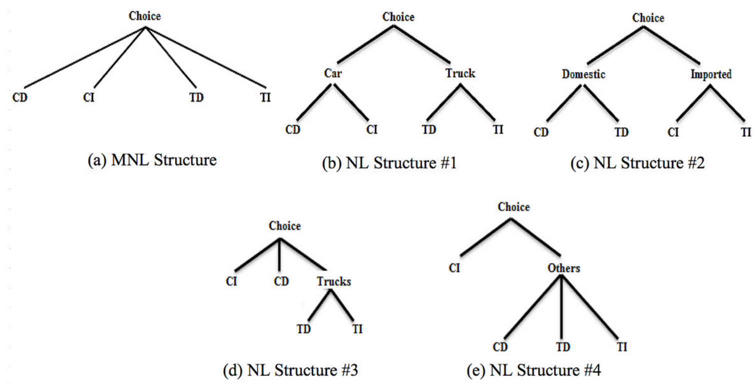

Figure 1. Discrete Choice Model Structures

The results from the three models are fairly consistent across most of the estimated parameters. Cross tabulation analyses were also conducted to determine the percentage of vehicle type choices that were predicted correctly. The overall percentage explained right for the MNL, NL and MXL were 62.7, 63.7 and 62.7%. Since the NL model offers the highest predictive power of the three models, our discussion of the factors explaining the choice probability of vehicle types will be based on the NL model.

It is observed that imported brands, especially larger-type vehicles are not very prevalent in the study area. The market share for imported brands is quite low when compared with domestic automobiles. This can be attributed to the fact that some imported brands are not as readily available in the Windsor region since the auto manufacturing industry focuses on domestic brands. Also, imported vehicles can be more expensive when compared to domestic vehicles.

The two alternative specific constants ASC_1 (for domestic cars) and ASC_2 (for imported cars) are negative and highly significant. This indicates that, other things being equal, the probability for choosing domestic and imported cars tend to decrease relative to choosing imported trucks. However, the positive and significant constant, ASC_3, corresponding to the utility of the "domestic truck" alternative that, suggest that the probability of choosing domestic trucks is higher than the probability of choosing imported trucks, all else being equal.

All of the estimated parameters in the NL model are statistically significant and positive in sign with the exception of the parameter associated with the interaction term NEWCFE (new vehicles × average city fuel efficiency). The achieved negative parameter for the NEWCFE term suggest that people opting for newer vehicles will be less affected by the difference in fuel efficiency simply because most new cars have high and similar fuel efficiency in general.

Looking at each alternative individually, it can be seen that the probability for choosing domestic cars will increase in census tracts with higher percentage of elderly population aged 65 and older. The same could be said about census tracts with higher proportion of people of European ethnicity and higher proportions of Canadian citizens. The propensity of owning domestic cars also increases for when the average highway fuel efficiency for those cars increases. Unsurprisingly, the results suggest that the choice of older domestic cars (namely, model year earlier than 1990) will increase in census tracts with higher percentage of people with low income, namely people earning less than $25,000 per year.

Table 2. Estimation Results

	MNL Model		NL Model		MXL Model	
Param	**Beta**	**t-stat**	**Beta**	**t-stat**	**Beta**	**t-stat**
ASC_1	-17.08	-19.96	-13.82	-14.12	-15.89	-16.31
β_A	0.001	3.21	0.001	2.836	0.001	3.613
β_C	6.815	9.609	7.077	9.553	6.797	9.560
β_G	2.27	3.579	2.24	3.342	2.287	3.543
β_E	8.328	5.756	7.95	4.991	6.704	3.880
β_K	0.61	45.36	0.670	44.20	0.616	45.266
ASC_2	-19.97	-38.45	-19.005	-20.68	-20.13	-38.36
β_I	0.006	4.085	0.007	3.772	0.008	4.381
β_F	18.005	2.18	28.82	2.874	22.14	2.414
β_D	5.06	14.39	5.82	15.194	5.502	14.881
β_J	0.220	25.545	0.45	24.691	0.238	15.860
β_H	0.001	1.950	0.003	3.345	0.002	2.096
β_Z	-0.041	-11.76	-0.044	-10.80	-0.012	-1.374
ASC_3	3.35	4.839	8.502	9.960	4.646	5.520
β_B	2.56	2.114	2.398	1.942	2.461	2.023
***IV*(CI)**			0.711	18.965		
***IV*(OTHERS)**			0.895	14.334		
β_J×VNE					-0.154	-3.677
Standard Deviation (**β_J×VNE**)					0.033	2.831
L(0)	-6293.092		-6293.092		-6293.092	
L(β)	-10478.493		-15349.005		-10478.493	
ρ^2 Adj	0.40		0.59		0.51	

The choice propensity of imported cars increases in census tracts with higher proportions of people holding higher levels of education (namely, Master's, PhD or medical degrees) and who are older than 65. The propensity of choosing older imported cars increases in census tracts with people of low income and in tracts with higher proportions of people of European ethnicity. Intuitively, census tracts with higher proportions of people of any age who have medical degrees are associated with higher propensity of owning new imported cars (namely, model year later than 2006). The same could be said about census tracts with higher proportion of people earning high income of over $60,000 per year. Also, the propensity of owning imported cars increases for vehicles with higher average highway and city fuel efficiencies. Furthermore, the probability of owning older imported vehicles tends to increase in census tracts with higher proportions of young people between the ages of 15 and 24 and who have low income.

When it comes to truck class (i.e. SUVs, minivans and pickup trucks), the probability for owning domestic trucks tends to increase in census tracts with higher proportion of people over the age of 65. Also, the probability of owning domestic trucks increases in census tracts with higher proportion of men aged between 35 and 44. Similarly, census tracts with higher percentage of Canadian citizens tend to show a stronger affiliation with domestic larger vehicles. Also, the probability of owning new domestic vehicles increases in areas with high income people.

As in the case of imported cars, the probability of owning imported larger vehicles (i.e. SUVs, minivans and pickup trucks) will increase in census tracts with higher proportions of people holding higher levels of education and who are older than 65. Imported large vehicles have a higher tendency to be owned in areas with higher percentages of men aged between 35 and 44, more people of European ethnicity and higher proportions of people of any age who have medical degrees. Higher share of new imported trucks are also observed in areas with high income people. Also, higher average city fuel efficiency tends to increase the preference of imported larger vehicles. Finally, areas with higher proportions of young people with low income appear to be affiliated with older imported large vehicles.

Conclusion

This paper investigated the demographic and vehicle specific characteristics that influence the choice decision of domestic and imported vehicle in the Windsor region. The results reinforce some of the previous findings from the literature. We found that age, education, income, ethnicity, gender and profession all influence the type of chosen vehicle. Furthermore, fuel economy appears to play a very important role when choosing between imported and domestic vehicles. The information from the current models provide knowledge that helps understand the type of factor that influence vehicle purchasing patterns in the area especially when it comes to domestic vehicles. The models of the sort developed in this paper can also be used to test transportation policies in the region. However, before such models can be used in such capacity, more work is needed to extend the capabilities of the models we developed in this paper. For instance, more descriptive variables and interaction terms should be tested. The aim would be to predict domestic and imported cars and trucks at the census tract level under different scenarios.

Bibliography

Ben-Akiva, M. and Lerman S. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge Massachusetts.

Choo, S., Mokhtarian, P., 2004. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. Transportation Research Part A 38: 201–222

Greene, W. 2002. *LIMDEP 8.0–NLOGIT 3.0: Reference Guide*. Econometric Software, Inc., New York.

Mohammadian, A., Miller, E., An Empirical Investigation of Household Vehicle Type Choice Decisions. 82nd Annual Transportation Research Board Meeting. January 2003

Paleti, R., Bhat, C., Pendyala, R., Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects. 92nd Annual Meeting of the Transportation Research Board. August 2012

Potoglou, D., Kanaroglou. P.S., Modeling Car Ownership in Urban Areas: A Case Study of Hamilton, Canada. Centre for Spatial Analysis Working Paper Series. McMaster University. March 2006

Roorda, M., Mohammadian, A., Miller, E. 2000. Toronto Area Car Ownership Study: A Retrospective Interview and Its Applications. Transportation Research Record series

Spissu, E., Pinjari, A., Pendyala, R., 2009, A Copula-Based Joint Multinomial Discrete Continuous Model of Vehicle Type Choice and Miles of Travel. http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Copula_VMT_6March09.pdf

Train, K., Winston, C., Vehicle Choice Behaviour and the Declining Market Share of US Automakers. July 2004

US Department of Energy. 2012. Downloadable Fuel Economy Data. http://www.fueleconomy.gov/feg/download.shtml. Accessed November 2012.