A Piecewise Linear Multinomial Logit Model of Private Vehicle Ownership Behaviour of Indian Households

Sarojeet Dash¹ · Vinod Vasudevan² · Sanjay Kumar Singh³

Received: 10 February 2015 / Accepted: 22 June 2016 / Published online: 30 June 2016
© Springer International Publishing Switzerland 2016

Abstract Development of vehicle ownership models is challenging in developing countries due to lack of quality data. In the Indian context, although some studies have been conducted about disaggregate modelling of vehicle ownership behaviour, most of them are region specific. This paper reviews the development of a basic multinomial logit vehicle ownership model presented earlier by the same authors using available data sets in India and then presents another multinomial logit model which uses a segmented specification of the systematic utilities of various alternatives. The new model is designed to examine if the various factors affecting vehicle ownership behaviour do so differently for richer households as compared to the rest of the households. It is observed that the various factors indeed affect private vehicle ownership decision making process differently for richer households. The new model is then tested for efficiency using a simulation testing technique. The simulation test shows that the new model is effective in modelling private vehicle ownership in Indian scenario. The paper also presents various inferences drawn from the coefficients of the explanatory variables included in the model.

Keywords Multinomial logit model · Vehicle ownership · Behavioural modelling

✉ Vinod Vasudevan
vinodv@iitk.ac.in

¹ Indian Institute of Technology Kanpur,
Kanpur 208016, Uttar Pradesh, India

² Department of Civil Engineering, Indian Institute of
Technology Kanpur, Kanpur 208016, Uttar Pradesh, India

³ Indian Institute of Management Lucknow,
Lucknow 226013, Uttar Pradesh, India

Introduction

India, being the second most populated country of the world and a steadily growing economy, is a rapidly growing market for motorized vehicles. The total number of registered motor vehicles in India has increased by 157.8 % from 55.0 million in 2001 to 141.8 million in 2011. Out of the 141.8 million motor vehicles that were registered in India in 2011, 101.8 million (71.8 %) were two-wheelers and 19.2 million (13.6 %) were four-wheelers [1]. Between 2001 and 2011, two-wheeler ownership increased from 35.99 to 82.01 two-wheelers per 1000 people, an increase of 127.9 % whereas four-wheeler ownership increased from 6.57 to 15.53 four-wheelers per thousand people, an increase of 136.4 % [1, 2]. Rapid motorization in India and growing inclination of its people towards private transport modes have made the country the fourth largest petroleum consuming nation in the world after the US, China and Japan with a total petroleum consumption of 3,426 thousand barrels per day in the year 2011 [3].

It is necessary to understand the private vehicle ownership behaviour of Indian households in order to devise effective policies to regulate the dependence of Indian households on private motorized vehicles. Dash et al. [4] made such an attempt by developing a multinomial logit (MNL) model to quantify the effect of various socio-economic factors on the private vehicle ownership behaviour of Indian households. This paper presents further research work in continuation to the work presented by Dash et al. [4].

Literature Review

This section presents a brief description of the existing studies in the field of disaggregate vehicle ownership modelling in India. It also describes, in a separate section,

the study presented by Dash et al. [4] based on which the present study is performed.

Existing Studies

Many disaggregate vehicle ownership models have been developed in the past for developed countries [5–10]. But generally such models utilise datasets of dedicated transport surveys with detailed demographic, socio-economic data and vehicle ownership information including type of vehicles owned by the households. Due to unavailability of such survey datasets in India and many other developing countries, such studies can at best be used as guidelines.

Not many studies on disaggregate vehicle ownership modelling has been done for Indian scenario and most of the models which have been developed are for specific cities in India. Table 1 summarizes some of these studies.

Chamon et al. [11] did make a disaggregate car ownership model for entire country of India, but they did not consider the effect of various other explanatory variables that have the potential to affect vehicle ownership behaviour. The only explanatory variable included in their

model was total expenditure of households, which was used as a proxy for income.

Dash et al. [4] presented an all-India disaggregate vehicle ownership model that included various socio-economic explanatory variables, all of which were found to be significant in the study. This study is summarized in detail in the next section.

The Basic Model

This section briefly describes the relevant methods used and the MNL model presented by Dash et al. [4]. The model is referred as “Basic Model” henceforth.

The study first presented the importance and advantages of disaggregate vehicle ownership models over aggregate ones and then pointed out that a disaggregate model of vehicle ownership behaviour of households with multiple explanatory variables, developed for India as a whole, was missing in the existing literature. The study aimed at filling this gap.

Income of the households is one of the main explanatory variables used in the literature of disaggregate vehicle

Table 1 Summary of disaggregate vehicle ownership studies conducted in Indian Context

SI no.	Study	Study area	Model	Modelled variable	Important explanatory variables	Description/relevant findings
1	Chamon et al. [11]	India	Binary probit model	Car ownership	Total expenditure	Total expenditure of households used as proxy for income was the only explanatory variable used in the model to estimate the probability of a household owning at least one car
2	Kumar and Rao [12]	Mumbai	Multinomial logit model	Car ownership	Travel time, travel cost, car loan payment option, servicing cost of car per annum	Stated preference study. Attributes like travel time, travel cost, income were used. Reported unwillingness of households towards disclosing their income
3	Srinivasan et al. [13]	Chennai	Ordered probit model	Four wheeler ownership, two-ownership	Income, number of working members	Households with predominantly young workers aged less than 45 are more likely to purchase two-wheelers as compared to those with old workers. Presence of school-aged children is positively related with increase in four-wheelers
4	Padmini and Dhingra [14]	Pune	Multinomial logit model	Four-wheeler ownership, two-ownership	Income, household size, type and cost of parking, possession of driving license	Revealed preference study. Income was represented by an ordinal variable
5	Banerjee [15]	Surat	Multinomial logit model	Combined model for car and two-wheeler ownership	Income, household size	Study did not include households with no private vehicles. Income (ordinal variable) and household size were used as explanatory variables. They found that Income is more significant than household size in explaining car ownership
6	Dash et al. [4]	India	Multinomial logit model	Combined model for car and two-wheeler ownership	Per capita regular expenditure, household size, etc	[Please refer Sect. “The Basic Model”]

ownership modelling. The study presented the inadequacy of income data in developing countries and the reasons why income data is considered disadvantageous and inaccurate as compared to expenditure data, especially in developing countries. The dataset of Consumer Expenditure Survey, which is annually conducted by the National Sample Survey Office (NSSO), was selected for the analysis in the study. The dataset includes data for various socio-economic variables as well as private vehicle ownership information [16]. It is to be noted that the dataset does not provide information about the number of vehicles owned by the households. The survey only recorded whether or not a household owned a vehicle and if yes, was it a two-wheeler or a four-wheeler. In the absence of a dataset of dedicated transport survey containing number and type of vehicles for households from all over the country, NSSO dataset has been selected for the purpose of this study. The sample set of the NSSO survey is spread throughout the country. NSSO has divided the geographical stretch of India into various geographical divisions which are subdivided into states, districts, rural/urban strata, hamlet groups/sub-blocks. In each sub-area, multiple households belonging to various expenditure levels are selected for the survey. The dataset is therefore very elaborate, geographically well distributed and can be considered to be representative of the entire population of the country.

Next, the study selected MNL model as the appropriate type of model for disaggregate vehicle ownership modelling based on comparison studies between various type of models in the existing literature. The behavioural model's alternatives were designed to represent a joint decision scenario about the private motorized vehicle ownership decision process. Each household was thus assumed to have a choice set comprising of the following four alternatives only:

1. The household chooses to own no motorized private vehicles (base alternative), "None".
2. The household chooses to own only two-wheelers, "2W only".
3. The household chooses to own both two-wheelers and cars, "2W & C".
4. The household chooses to own only cars, "C only".

The same choice set has been used for the model developed in this study. Such a joint decision process is a better representation of the actual decision scenario in Indian conditions because of the fact that the use of two-wheelers is very common in India. Unlike many developed countries, in India, two-wheelers are used for a myriad of activities: from work trips to shopping and other non-business trips. In other words, two-wheelers are either sufficient or supplementary in satisfying the travel needs of the Indian household and never superfluous.

Next, the NSSO consumer expenditure dataset for the year 2009–2010 was filtered off the entries that could lead to errors. Finally, the filtered dataset had 89,503 households out of which, 63,472 (70.9 %) chose None; 20,879 (23.3 %) chose 2W only; 3526 (3.9 %) chose 2W & C and 1626 (1.8 %) chose C only.

Then, the NSSO consumer expenditure data was used to design a proxy variable to represent economic standard of households. This variable was named as Annual Per Capita Regular Expenditure (APCRE). Regular expenditure was defined as total expenditure minus that on bedding, education, medical (institutional)¹ and durable goods. Expenditures in bedding, medical (institutional) and durable goods categories were excluded because they were recorded for a 365 days reference period in the survey but the frequency of expenditure for items belonging to these categories is generally lower than once in a year, therefore including this data could give erroneous information about the spending behaviour of the households. Expenditure in the education category were excluded because it proved to be detrimental to the goodness-of-fit of the model, which was explained to be due to the reason that some households may be paying the educational expenses from their savings or from loans.

Finally, the following variables were used as explanatory variables in the MNL model:

1. APCRE.
2. Household size.
3. Rural/urban location of residence (represented by a dummy variable "Rural/Urban" with value = 0 for rural and 1 for urban).
4. If the household has a regular salaried member (represented by a dummy variable "Regular Salary" with value = 0 for no and 1 for yes).
5. Whether children (age <18) are present in the household (represented by a dummy variable "Children Present" with value = 0 for no and 1 for yes).
6. Whether youngsters (18 ≤ age ≤ 35) are present in the household (represented by a dummy variable "Youngsters Present" with value = 0 for no and 1 for yes).
7. Whether seniors (age >60) are present in the household (represented by a dummy variable "Seniors Present" with value = 0 for no and 1 for yes).

All explanatory variables were found to be statistically significant at 5 % level of significance with expected signs. The relative values of the coefficients were also mostly intelligible. The goodness-of-fit measures like Likelihood

¹ Institutional medical expenses are defined as expenses incurred by a household on medical treatment as in-patient of a medical institution.

ratio Index and ρ^2 indicated that the model represented the dataset pretty well.

Besides these goodness-of-fit measures, another measure called “Average Probability of Correct Prediction (APCP)” was used to assess the goodness of fit of the model. It was calculated by summing for each household, the choice probability (predicted by the model) of the alternative that a household had actually chosen, and then dividing the sum by the total number of households in the dataset. This value could vary between zero (for a completely imperfect model) and one (for a perfect model). For the model, APCP was found to be 0.5868, which is considered acceptable.

The paper presented that the probability of a household choosing to own a private vehicle increases with per capita regular expenditure and household size. It also showed that households in rural areas are more inclined to own cars if they can afford it. Households with regular salaried members were found to be more inclined to own private vehicles and it was reasoned to be so because of easier availability of vehicle loans to such households. Presence of children and senior members in a household were found to incline the households to own both two-wheelers and cars. And lastly, it was observed that presence of youngsters in a household strongly increases its likelihood to own two-wheelers.

The Segmented MNL Model

This study attempts to further explore the vehicle ownership behaviour of Indian households. The basic suspicion that drives this study is that for richer households, increase in economic status may not be that important a factor in the decision making process of which type of private vehicle to purchase and own. Other factors could be more influential on their decisions. To examine the validity of this conjecture, the households were classified into two categories: High Expenditure group and Modest Expenditure group and a MNL model was estimated that could take two different set of coefficients depending on the class to which the households belonged. The definition that was chosen for this classification is as follows:

- (i) High Expenditure group: All the households were arranged in ascending order of APCRE. The households that were amongst the top 10 % i.e., the top 8950 households out of the 89,503 households in the dataset were included in this group. Households with APCRE value greater than or equal to Rs. 29,080.38 came in this category.
- (ii) Modest Expenditure group: rest 90 % of the households that were not included in the High Expenditure group were included in this category.

The 10 % value was chosen on arbitrary basis for the purpose of this study. This is known as a priori market segmentation technique. The MNL model was designed so that the model could take two different set of coefficient values for the two segments: high and modest expenditure groups. In a simple MNL model, the utilities of various alternatives are assumed to have a linear form. That is, the systematic component of the utility of an alternative i for a household j is given as:

$$V_{ij} = \beta_{0i} + \beta_{1i}X_{1j} + \beta_{2i}X_{2j} + \cdots + \beta_{ni}X_{nj}$$

This can be converted into a segmented form by introducing a dummy variable D which is defined as follows:

$$D = \begin{cases} 1 & \text{if } X_{1j} \geq a \\ 0 & \text{if } X_{1j} < a \end{cases}$$

This dummy variable basically serves to classify the data points into two classes, one class of data points that have X_{1j} value greater than equal to a and other that have less than a (X_{1j} being APCRE in the MNL model and $a = 29,080.38$).

The systematic utility function for the segmented model can be written as follows:

$$V_{ij} = \beta_{0i} + \beta_{1i}X_{1j} + \beta_{2i}X_{2j} + \cdots + \beta_{ni}X_{nj} + D \times (\beta_{0i}^{\sim} + \beta_{1i}^{\sim}X_{1j} + \beta_{2i}^{\sim}X_{2j} + \cdots + \beta_{ni}^{\sim}X_{nj})$$

This utility function would now take two different forms for the two segments considered:

$$V_{ij} = \begin{cases} \beta_{0i} + \beta_{1i}X_{1j} + \beta_{2i}X_{2j} + \cdots + \beta_{ni}X_{nj} & \text{if } X_{1j} < a \\ (\beta_{0i} + \beta_{0i}^{\sim}) + (\beta_{1i} + \beta_{1i}^{\sim})X_{1j} + (\beta_{2i} + \beta_{2i}^{\sim})X_{2j} + \cdots + (\beta_{ni} + \beta_{ni}^{\sim})X_{nj} & \text{if } X_{1j} \geq a \end{cases}$$

Thus, the $\beta_{ki}^{\sim}$ values are the differences in the two set of beta values. If most of these $\beta_{ki}^{\sim}$'s come out to be statistically insignificant, then it would imply that there is no need to do this segmentation and the effect of the explanatory variables is not very different between the two expenditure groups.

Estimation of Segmented MNL model

A segmented MNL model with the same seven explanatory variables used in the basic model was estimated, but the $(\beta_{ki} + \beta_{ki}^{\sim})$ values for two of the explanatory variables viz., “Youngsters Present” and “Seniors Present” were found to have unexpected signs. This discrepancy could be explained as follows.

Some of the households which had Youngsters Present value as 1 were actually single member households comprising of one youngster only (referred to as “single-youngster households” henceforth). Single-youngster households were observed to prefer to choose None

alternative. 90.4 % of single-youngster households chose None and 9.0 % chose 2W only alternatives, whereas out of the households which had Youngsters Present value as 1, only 69.7 % chose None and 24.7 % chose 2W only. This means youngsters living with family are more inclined to own vehicles than those living single. The above percentages also indicate that youngsters prefer to own two-wheelers.

Some of the households which had Seniors Present value as 1 were actually households with all senior members. Such households are henceforth referred to as “Seniors-only households”. Seniors-only households were also observed to prefer the None alternative. 91.6 % of Seniors-only households chose the None alternative whereas only 66.4 % of the households which had Seniors Present value as 1 chose None. This may be because aged people have lesser travel needs.

To take these factors into consideration, another segmented MNL model was estimated which used the same variables as the basic model except for the following changes. “Youngster Present” was replaced by “Youngster Present in family” which was 1 if at least one youngster was present in the household and it was not a single-youngster household and 0 otherwise. “Seniors Present” was replaced by “Seniors Present in family” which was 1 if at least one senior member was present in the household and it was not an seniors-only household and 0 otherwise. Other than these changes, two extra dummy variables were added into the set of explanatory variables. First one was “Single Youngster” which was 1 if the household was a single-youngster household and 0 otherwise and the second was “Seniors only” which was 1 if the household was an seniors-only household and 0 otherwise. The estimated coefficients and related statistics of the segmented MNL model described above are shown in Table 2.

As can be observed from Table 2, all coefficients of the model are highly significant except the ones marked by “a”. The fact that most of the coefficients of $D \times X_{kj}$ ’s are significant indicates that the segmentation was necessary. Moreover, if the goodness of fit measures like adjusted rho square and average probability of correct prediction of the basic model ($\bar{p}^2 = 0.49361$, $APCP = 0.5868$) are compared to those for the segmented model, it is clear that later model fits the dataset much better. Likelihood ratio index was also calculated for the segmented MNL model with respect to the base model as follows:

$$\begin{aligned} \text{Likelihood ratio index} &= -2[L(\beta_{\text{Base}}) - L(\beta_{\text{Segmented}})] \\ &= 2[62807.4 - 54197.1] = 17220.6 \end{aligned}$$

Table 3 shows the two set of coefficient values that the segmented model takes for High and Modest Expenditure groups.

From Table 3, it can be observed that as suspected, the coefficients of APCRE are much lower for the High Expenditure group, which shows that indeed for households in the High Expenditure group, increase in economic standard is a less important factor in the decision making process about their private vehicle ownership. The coefficient for 2W & C alternative is highest for Modest Expenditure group, which means that for this group as APCRE increases, households become more and more inclined to own both two-wheelers and cars together. For the High Expenditure group, the coefficient for C only alternative is the highest, which means that for this group as APCRE increases, households become more and more inclined to own only cars. It should be noted that the two set of coefficients are comparable because they were estimated by one MNL model which has only one scale factor for both set of coefficients. If two separate models were estimated for each segment, then they could have different scale factors in which case the coefficients would not have been directly comparable to one another.

The coefficients of household size are higher for the High Expenditure group (than that for the Modest Expenditure group), which means that household size is a more important factor in private vehicle ownership decision for households in this category. Moreover, for both the categories, the coefficient for 2W & C is highest, followed by C only and 2W only. This means that increase in household size increases the inclination of households to choose 2W & C the most, followed by C only, then 2W only, then None which makes perfect sense.

Coefficients of Rural/Urban variable suggest that for Modest Expenditure group, the rural households are more inclined to own all private vehicles, especially cars, as compared to urban households. The absence of public transport in rural areas could be a possible reason for this observation. But in the High Expenditure group, urban households prefer to choose 2W only more than rural households. This could be because the households in urban areas face traffic congestion and parking problems due to which they prefer two-wheelers over cars. However, rural households especially the ones in the High Expenditure group are inclined to make longer distance travels to avail various facilities available only in cities and cars are more preferable for long trips. This could be the reason for them not to prefer two-wheelers.

Presence of a regular salaried member in a household was expected to have a positive effect on the utilities of owning private vehicles because of easier availability of loan facility for them. The coefficients of Regular Salary variable in Table 3 show that presence of regular salary earning member inclines the households to own private vehicles in the Modest Expenditure group as expected. But the coefficients for 2W & C and C only for the High

Table 2 Estimated beta values and related statistics for the Segmented MNL model

Variable	2W only	2W & C	C only
Alternative specific constant	-6.037 (-111.7)	-12.142 (-83.5)	-10.343 (-57.9)
APCRE	2.054×10^{-04} (98.5)	3.515×10^{-04} (73.6)	2.790×10^{-04} (44.4)
Household size	0.293 (54.4)	0.466 (47.7)	0.365 (24.7)
Rural/urban (rural = 0, urban = 1)	-0.179 (-8.7)	-0.569 (-11.2)	-0.892 (-12.2)
Regular salary (no = 0, yes = 1)	0.316 (14.5)	-0.049 (-0.9) ^a	0.234 (3.3)
Children present (no = 0, yes = 1)	0.279 (10.4)	0.706 (10.4)	0.560 (6.1)
Youngsters present in family (no = 0, yes = 1)	0.291 (10.9)	0.496 (7.0)	0.027 (0.3) ^a
Seniors present in family (no = 0, yes = 1)	0.242 (10.8)	0.535 (10.1)	0.265 (3.6)
Single youngster (no = 0, yes = 1)	-3.266 (-11.6)	-4.075 (-4.2)	-3.916 (-3.8)
Seniors only (no = 0, yes = 1)	-1.802 (-12.6)	-1.745 (-4.4)	-1.215 (-3.2)
D × alternative specific constant	3.642 (25.6)	7.022 (31.6)	5.004 (19.2)
D × APCRE	-1.914×10^{-04} (-68.2)	-3.078×10^{-04} (-58.6)	-2.322×10^{-04} (-34.8)
D × household size	0.376 (9.5)	0.627 (13.8)	0.594 (11.1)
D × rural/urban	0.351 (4.8)	0.271 (2.6)	0.173 (1.4) ^a
D × regular salary	-0.178 (-2.9)	-0.253 (-2.8)	-0.449 (-3.8)
D × children present	-0.377 (-4.5)	-0.684 (-5.8)	-0.552 (-3.6)
D × youngsters present in family	-0.139 (-1.8) ^a	-0.777 (-6.8)	-0.384 (-2.7)
D × seniors present in family	-0.622 (-7.0)	-0.644 (-5.8)	-0.127 (-0.9) ^a
D × single youngster	2.299 (7.7)	-0.216 (-0.2) ^a	1.037 (1.0) ^a
D × seniors only	0.867 (4.5)	0.542 (1.2) ^a	0.894 (2.1)
Number of observations = 89503			
Log-likelihood with all zero coefficients $L(0) = -124077.5$			
Log-likelihood with alternative specific constants only $L(c) = -70124.3$			
Log-likelihood of model $L(\beta) = -54197.1$			
Likelihood ratio Index = $-2[L(0)-L(\beta)] = 139760.9$			
$\rho^2 = 0.56320$			
$\bar{\rho}^2 = 0.56272$			
Average Probability of Correct Prediction (APCP) = 0.6573			

t-statistic in parenthesis

^a Not significant at 5 % level of significance ($\alpha = 0.05$)**Table 3** Two set of coefficient values of the Segmented MNL model

Variable	Modest expenditure group			High expenditure group		
	2W only	2W & C	C only	2W only	2W & C	C only
Constant	-6.037	-12.142	-10.343	-2.395	-5.120	-5.339
APCRE	2.054×10^{-04}	3.515×10^{-04}	2.790×10^{-04}	1.396×10^{-05}	4.365×10^{-05}	4.678×10^{-05}
Household size	0.293	0.466	0.365	0.668	1.093	0.960
Rural/urban	-0.179	-0.569	-0.892	0.172	-0.298	-0.719 ^b
Regular salary	0.316	-0.049 ^a	0.234	0.138	-0.301	-0.215
Children present	0.279	0.706	0.560	-0.098	0.022	0.008
Youngsters present in family	0.291	0.496	0.027 ^a	0.152 ^b	-0.281	-0.356
Seniors present in family	0.242	0.535	0.265	-0.380	-0.109	0.139 ^b
Single youngster	-3.266	-4.075	-3.916	-0.967	-4.292 ^b	-2.879 ^b
Seniors only	-1.802	-1.745	-1.215	-0.934	-1.203 ^b	-0.321

^a Not significant at 5 % level of significance ($\alpha = 0.05$)^b Difference not significant at 5 % level of significance

Expenditure group are negative. These apparently illogical coefficients could be explained as follows: businessmen are included as non-regular salaried people. Perhaps in High Expenditure group, a greater fraction of the households without a regular salaried member are families of businessmen that may be more inclined to own cars than other type of households.

It was expected that presence of children stimulates households to own private vehicles because households with children are likely to have more non-business trips. The fact that most of the coefficients for this variable are positive is in accordance with this expectation. For both the High and Modest Expenditure groups, for Children Present variable the coefficients of 2W & C is the highest followed by C only and 2W only. This is in accordance to the logic that households associate a sense of safety with cars. Additionally, the negative coefficient for 2W only alternative of High Expenditure group suggests that households belonging to this category which have children are strongly averse to owning two-wheelers only.

Youngsters belong to the target age group for many of the two-wheeler products that are launched in India. The presence of youngsters is expected to create a positive bias towards ownership of two-wheelers. The coefficients in Table 3 suggest that for the Modest Expenditure group, the presence of youngsters in a family increases the utility of 2W & C the most followed by 2W only and it has no significant effect on the utility of C only. This indicates that if youngsters are in a family, the family becomes inclined to own two-wheelers irrespective of whether it owns a car or not. For the High Expenditure group, the presence of youngsters in a family seems to increase the utility of 2W only alternative but has a negative impact on the utilities of 2W & C and C only. This can be explained by the fact that on an average, the household size of households in the High Expenditure group is less, so the effect of presence of a youngster is more. Moreover, a significant share of the families with youngsters in High Expenditure group could be comprised of youngster couples i.e., all household members are youngsters. Such households can be expected to prefer to own only two-wheelers.

For the Modest Expenditure group, the presence of seniors in a family has positive effect on all private vehicle ownership alternatives with the highest positive impact on 2W & C alternative. This represents the fact that for aged people, using public transit facilities is inconvenient. The positive impact is highest for 2W & C alternative; this may be because families with seniors like to own two-wheelers for general purpose use and cars for making trips with the senior members. For the households in the High Expenditure group, the presence of seniors in family has negative impact on the utilities of 2W only and 2W & C alternatives

and positive impact on C only alternative. This can also be explained by the fact that household size of households in the High Expenditure group is less, so the effect of presence of seniors is more. Moreover, the families with seniors in High Expenditure group could be comprised of households consisting of senior individuals and live-in servants.

As expected, for Modest Expenditure group, the set of three negative coefficients indicates that single-youngster households are averse to own private vehicles. This is most probably because they can manage to travel by public transit facilities and therefore do not need to own private vehicles. Another reason could be that most single youngsters in this expenditure group are in the initial stages of their careers and therefore have more saving tendency. In High Expenditure group also, single-youngster households are averse to own private vehicles but they are not that averse to choose 2W only alternative. This is very logical because this indicates that single youngsters in the High Expenditure group either choose to own no private vehicles or they choose to own two-wheelers only. Single-youngster households in both the expenditure groups are most averse to own both two-wheelers and cars simultaneously.

The coefficients of “Seniors only” variable suggest that households comprising of seniors only are averse to owning private vehicles but they are least averse to choose C only alternative, as per expectation. Comparison of the two set of coefficients for this variable suggests that households in the High Expenditure group comprising of seniors only are relatively more likely to choose C only alternative as compared to households in the Modest Expenditure group comprising of seniors only as can be logically expected.

The model was also validated by preparing McFadden’s prediction success table [17]. This has been presented in Table 4. As can be observed, the overall prediction success rate of the model is 66 %. It should be noted that for preparation of McFadden’s prediction success table, the predicted choices for each household is based on the choice probabilities of individuals as per the model.

In reality, the choice of a household can be different from other households with exactly same values of all the explanatory variables included in the model. In other words, the observed choice of a household can be different from the alternative with highest probability of selection according to the model. Especially, in cases where probabilities of selection of two alternatives are very close for a household, assuming that the alternative with highest probability will be selected by the household is arguable. To take this into account, a simulation test was additionally performed as described in the next section.

Table 4 McFadden prediction success table

Observed choices	Predicted choices				Row total	Observed share (%)
	None	2W only	2W & C	C only		
None	50481.2	11167.4	1164.2	659.2	63472	70.92
2W only	11284.6	7557.1	1439.5	597.8	20879	23.33
2W & C	1067.6	1534.9	671.6	252.0	3526	3.94
C only	633.0	618.7	253.4	120.9	1626	1.82
Column total	63466	20878	3529	1630	89503	
Predicted share (%)	70.91	23.33	3.94	1.82		
% Correctly predicted	80 %	36 %	19 %	7 %		

Overall prediction success rate = 66 %

The Simulation Test

The process of simulation used to test the models in this study was inspired by the simulation tests designed by Dubedi et al. [18]. The simulation test is described in the following paragraphs. This testing method gives an indication if the model represents the dataset well.

In each iteration a set of predicted choices is generated by a stochastic simulation process according to the choice probabilities of various alternatives calculated by the model. This can be done by using a simple random number generating function of any standard data-processing software (MATLAB Version 7.10, R2010a was used for this study) which generates a random number with uniform probability distribution. The process can easily be understood by an example as follows:

- (i) say the dataset has N number of households
- (ii) the model to be tested has generated four probability values, one for each alternative, for each of these households. For n th household, these values are say $P_{\text{None},n}$, $P_{2W,n}$, $P_{2W\&C,n}$ and $P_{C,n}$
- (iii) a value of a random number X , which is uniformly distributed between 0 and 1, is generated, say x (Note: $0 < x < 1$).
- (iv) Now, the predicted choice is generated as per the following conditions:
 - (a) If $0 < x < P_{\text{None},n}$, the household is predicted to choose the first alternative, None
 - (b) If $P_{\text{None},n} < x < (P_{\text{None},n} + P_{2W,n})$, the household is predicted to choose the second alternative, 2W only
 - (c) If $(P_{\text{None},n} + P_{2W,n}) < x < (P_{\text{None},n} + P_{2W,n} + P_{2W\&C,n})$, the household is predicted to choose the third alternative, 2W & C
 - (d) If $(P_{\text{None},n} + P_{2W,n} + P_{2W\&C,n}) < x < (P_{\text{None},n} + P_{2W,n} + P_{2W\&C,n} + P_{C,n} = 1)$, the household is predicted to choose the fourth alternative, C only

- (v) In this way, for each household, the probability of selection of each alternative is according to that calculated by the model being tested.
- (vi) The above process, viz. steps iii and iv, is repeated for each of the N households. Thus, a set of N predicted choices is generated in each iteration.
- (vii) Then, for each alternative, the predicted percentage share of households is calculated from the predicted choices.

These steps are iterated a desired number of times, say 10,000. The next step is to plot frequency histograms of the predicted percentage shares for each of the alternatives. Fitting an appropriate distribution function to each of the histograms, 95 % confidence intervals can be calculated for the predicted shares of each of the alternatives. If the observed shares of the alternatives calculated from the dataset lies within the corresponding confidence intervals, then the null hypothesis that the model effectively represents the dataset cannot be rejected, otherwise it is rejected and the model is proved to be ineffectual.

Performance Check by Simulation

The simulation test described in Sect. “The Simulation Test” was conducted on the Segmented MNL Model. Figure 1a–d show the frequency histograms of the simulated predicted shares for the four alternatives. Table 5 presents the 95 % confidence intervals calculated for the predicted shares of the four alternatives and the corresponding observed shares.

From Table 5, it is clear that the observed shares lie well within the calculated 95 % confidence intervals and therefore the null hypothesis that the model accurately represents the dataset cannot be rejected. This proves that the segmented MNL model is a robust representation of the dataset.

Fig. 1 Frequency histograms of predicted percentage shares calculated by simulation for the four alternatives for the piecewise linear model

Table 5 95 % confidence intervals for the predicted percentage shares and observed shares

	None	2W only	2W & C	C only
95 % Confidence interval	[70.66, 71.16]	[23.07, 23.56]	[3.82, 4.06]	[1.74, 1.91]
Observed shares	70.92	23.33	3.94	1.82

Percentage values presented in this table are rounded up to two decimal places and may not exactly add up to 100

Conclusions

This paper presents a segmented MNL model to examine the differences in the way various factors affect the vehicle ownership behaviour of high expenditure households and moderate expenditure households.

The NSSO consumer expenditure dataset for the year 2009–2010 was cleaved into two segments: High Expenditure group and Modest Expenditure group based on a threshold APCRE value chosen conveniently. All the coefficients of the model were comprehensible and could be logically explained. The significance of the differences in coefficients for the two segments indicates that the households in the two expenditure groups do behave significantly differently and a segmented functional form of the utility functions was indeed necessary.

In other words, it was proved that the explanatory variables did affect the utility functions differently for the two segments. For example, increase in APCRE or economic standard is less important and household size is

more important a consideration for the households in the High Expenditure group compared to those in the Modest Expenditure group during their private vehicle ownership choice decision process.

In addition to the inferences about the effect of the various explanatory variables included in the model which was presented by Dash et al. [4], this study showed how the effects of the variables are different amongst the high and modest expenditure groups. Furthermore, a strong inclination of single-youngster households and seniors-only households to choose to own no private vehicles has been revealed.

This study also presents a simulation test technique to assess the efficiency of the developed model. The segmented MNL model presented in this paper passed the simulation test as the observed percentage share of all the four alternatives were found to lie within the corresponding 95 % confidence intervals calculated from the frequency distribution of the simulated predicted percentage shares of the respective alternatives.

References

1. MORTH (2012) Road transport year book 2009–2010 and 2010–2011. Transport Research Wing, Ministry of Road Transport and Highways (July 2012). New Delhi
2. World Bank (2013) World development indicators, india, population total. The World Bank website, <http://databank.worldbank.org/ddp/home.do>. Accessed 30 Mar 2013
3. USEIA (2012) The United States Energy Information Administration website. <http://205.254.135.7/cfapps/ipdbproject/iedindex3.cfm?tid=5&pid=5&aid=2&cid=IN,&syid=1980&eyid=2011&unit=TBDP>. Accessed 29 Jun 2012
4. Dash S, Vasudevan V, Singh SK (2013) A disaggregate vehicle ownership behaviour model of Indian households. *Transp Res Rec* 2394:55–62
5. Nolan A (2010) A dynamic analysis of household car ownership. *Transp Res Part A Policy Pract* 44:446–455
6. Potoglou D, Kanaroglou PS (2008) Modelling car ownership in urban areas: a case study of Hamilton, Canada. *J Transp Geogr* 16:42–54
7. Whelan G (2007) Modelling car ownership in Great Britain. *Transp Res Part A Policy Pract* 41:205–219
8. Dargay JM (2002) Determinants of car ownership in rural and urban areas: a pseudo-panel analysis. *Transp Res Part E Logist Transp Rev* 38:351–366
9. Chu YL (2002) Automobile ownership analysis using ordered probit models. *Transp Res Rec* 1805:60–67
10. Karlaftis M, Golias J (2002) Automobile ownership, households without automobiles and urban traffic parameters: are they related? *Transp Res Rec* 1792:29–35
11. Chamon M, Mauro P, Okawa Y (2008) Mass car ownership in the emerging market giants. *Econ Policy* 23:243–296
12. Kumar M, Rao KVK (2006) A stated preference study for a car ownership model in the context of developing countries. *Transp Plan Technol* 29:409–425
13. Srinivasan KK, Bhargavi PVL, Ramadurai G, Muthuram V, Srinivasan S (2007) Determinants of changes in mobility and travel patterns in developing countries: case study of Chennai, India. *Transp Res Rec* 2038:42–52
14. Padmini G, Dhingra SL (2010) Development of behavioural models of travel for metropolitan areas. In: *Proceedings of the 12th World Conference on Transport Research*, Lisbon, Portugal
15. Banerjee I (2011) *Automobility in India: a study of car acquisition and ownership trends in the city of Surat*. Ph.D. Thesis, University of California, Berkeley, USA
16. NSS (2011) Level and pattern of consumer expenditure 2009–2010, NSS 66th round, report no. 538, National Sample Survey Office, National Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India (December 2011). New Delhi
17. McFadden D (1978) The theory and practice of disaggregate demand forecasting for various modes of urban transport. In: *Proceedings of the seminar on Emerging Transport Planning Methods*, December 1976, Florida
18. Dubedi A, Chakroborty P, Kundu D, Reddy KH (2012) Modelling automobile driver's toll lane choice behaviour at a toll plaza. *J Transp Eng* 138:1350–1357