

**ANALYZING CAR OWNERSHIP IN QUEBEC CITY: A COMPARISON OF  
TRADITIONAL AND LATENT CLASS ORDERED AND UNORDERED MODELS**

Sabreena Anowar  
PhD Candidate  
Department of Civil Engineering and Applied Mechanics  
McGill University  
Montréal, Québec, H3A 2K6, Canada  
Email: [sabreena.anowar@mail.mcgill.ca](mailto:sabreena.anowar@mail.mcgill.ca)

Shamsunnahar Yasmin  
PhD Candidate  
Department of Civil Engineering and Applied Mechanics  
McGill University  
Montréal, Québec, H3A 2K6, Canada  
Email: [shamsunnahar.yasmin@mail.mcgill.ca](mailto:shamsunnahar.yasmin@mail.mcgill.ca)

Naveen Eluru\*  
Assistant Professor  
Department of Civil Engineering and Applied Mechanics  
McGill University  
Montréal, Québec, H3A 2K6, Canada  
Ph: 1-514-398-6823, Fax: 1-514-398-7361  
Email: [naveen.eluru@mcgill.ca](mailto:naveen.eluru@mcgill.ca)

Luis F. Miranda-Moreno  
Assistant Professor  
Department of Civil Engineering & Applied Mechanics  
McGill University  
Montréal, Québec, H3A 2K6, Canada  
Tel: 1-514-398-6589, Fax: 1-514-398-7361  
Email: [luis.miranda-moreno@mcgill.ca](mailto:luis.miranda-moreno@mcgill.ca)

\*corresponding author

March, 2014

## **Abstract**

Private car ownership plays a vital role in the daily travel decisions of individuals and households. The topic is of great interest to policy makers given the growing focus on global climate change, public health, and sustainable development issues. Not surprisingly, it is one of the most researched transportation topics. The extant literature on car ownership models considers the influence of exogenous variables to remain the same across the entire population. However, it is possible that the influence of exogenous variable effects might vary across the population. To accommodate this potential population heterogeneity in the context of car ownership, the current paper proposes the application of latent class versions of ordered (ordered logit) and unordered response (multinomial logit) models. The models are estimated using the data from Quebec City, Canada. The latent class models offer superior data fit compared to their traditional counterparts while clearly highlighting the presence of segmentation in the population. The validation exercise using the model estimation results further illustrates the strength of these models for examining car ownership decisions. Moreover, the latent class unordered response models perform slightly better than the latent class ordered response models for the metropolitan region examined.

Keywords: Car ownership in the Canadian context, Latent class models, Latent ordered logit, Latent multinomial logit

## 1. Introduction

To many, owning a private car is not only a utilitarian necessity but also a symbol of “power, status, control and freedom” (Yamamoto 2009). Private car ownership plays a vital role in the daily travel decisions of individuals and households influencing a range of long-term and short-term decisions. In the long-term, the vehicle ownership decisions are strongly tied with residential location and residential tenure (Eluru et al. 2010). In terms of short-term decisions, the level of car ownership influences the various aspects of activity travel patterns including activity frequency, activity duration, activity location, and travel mode choice for out-of-home work and non-work pursuits. The combination of the “symbolic perceived utility” (increased social esteem or higher status symbol) along with the tangible utility (increased mobility, higher access to opportunities) has resulted in increased auto-dependency both in the occidental (Caulfield 2012) and the oriental worlds (Wu et al. 1999; Li et al. 2010)<sup>1</sup>.

In Canada, it is reported that 84.4 percent of households owned or leased at least one vehicle in 2007 (Natural Resources Canada 2009). At the provincial level for Quebec, there has been a 17 percent increase in the number of cars over the last decade (Natural Resources Canada 2009). The use of personal vehicles by Canadians for daily trips is increasing while non-motorized travel for short distance utilitarian trips is declining. The concomitant negative consequences of increased auto ownership and the subsequent usage are manifold including: increased oil dependency, acute traffic congestion, increased greenhouse gas (GHG) emissions, deteriorating air quality, and rising health risks (Handy et al. 2005). It has been reported that transport sector accounts for 14 percent of global greenhouse gas (GHG) emissions, and road transportation is the biggest source of these emissions accounting for

---

<sup>1</sup> In recent years, a reversal in vehicle ownership levels in developed countries is being reported; highlighting a possible “peak” in ownership levels (Kuhnimhof et al. 2013; Millard-Ball and Schipper 2011).

about 76 percent share (Wu et al. 1999). In light of the growing attention on global climate change and the ensuing debate on how to reduce GHG emissions, in the past few decades, a considerable number of research efforts have examined household vehicle ownership decisions.

In our study, we extend the literature on vehicle ownership by employing latent class versions of the ordered and unordered response models. Our primary focus is to compare the performance of the latent class ordered and unordered models with their traditional counterparts. Towards this end, we estimate the latent class models using the vehicle ownership database for Quebec City. The models developed examine the influence of household socio-demographics, land-use and built environment variables on car ownership. Subsequently, we employ the model estimates to predict car ownership levels for a hold out validation sample. The exercise provides insights on population heterogeneity in terms of vehicle ownership choice while also providing insights on applicability of latent class ordered and unordered models for examining vehicle ownership. In summary, the current research study contributes to our understanding of car ownership behavior by examining the influence of various potential factors associated with household's decision to procure cars, particularly in the Canadian context.

The remainder of the paper is organized in the following order. Section 2 contains discussion on the relevant earlier research studies on car ownership. In Section 3, model structure and estimation procedure is described. Section 4, describes the main data sources and the sample formation procedure. Empirical results are presented and discussed in Section 5. Model validation outcomes are also included in the same section. Finally, we summarize the major findings of the research in Section 6.

## 2. Earlier research and current study in context

Car ownership and the associated dimensions including fleet size, vehicle type and usage has been one of the most researched transport topics. Historically, models to investigate car ownership and use have been under development since the 1930's (Whelan 2007). As a result of these continued efforts, there is a vast body of literature available on various forms of auto-ownership modeling. An extensive review of the models developed, particularly, for public sector transport planning purposes can be found in de Jong et al. (2004). In the rest of this section, however, we limit our discussions strictly to those studies (in the last two decades) that are relevant in the context of our research, i.e. studies that examine car ownership decision at a disaggregate level where the decision maker is the household. These car ownership studies can be classified into two categories: (1) independent car ownership models, and (2) studies modeling car ownership jointly/endogenously with other decision making processes (e.g. mode choice, residential location, vehicle type).

In our study, we focus our attention on independent car ownership models (see van Acker and Witlox 2010; Eluru et al. 2010 for a review of joint/endogenous modeling approaches). It was observed that among the different discrete choice frameworks, two general decision mechanisms have been extensively used for automobile ownership modeling. These are: the ordered-response mechanisms (ORM) and the unordered response mechanisms (URM). Recognizing the inherent ordinal nature of the car ownership levels, ordered probit (OP) and ordered logit (OL) models from the ORM category have been extensively applied. From the URM category, multinomial logistic regression (MNL) is the most widely employed by the researchers (Bhat and Pulugurta 1998, Potoglou and Susilo 2008).

In terms of explanatory variables, the earlier studies mainly focused on household socio-demographic characteristics (income, number of children, workers, non-workers, adults,

retirees, commuters and, licensed drivers, household size, household head characteristics, family type), residential location (urban/rural location, distance to the central business district (CBD), and population centrality), and built environment variables (such as dwelling type, residential density, population density, employment density, land use mix, transit accessibility, and urban design). The most significant findings from these studies for the different variable groups are briefly summarized here.

In terms of household socio-demographics, high household income, higher number of employed adults, and license holders increased the probability of owning multiple cars (Bhat and Pulugurta 1998; Chu, 2002; Potoglou and Kanaroglou 2006). Residential location variables also influenced car ownership decisions significantly. For instance, Dargay (2002) demonstrated that urban car owners were more sensitive to changes in motoring costs compared to their rural counterparts. This result suggests that car ownership in rural areas is a greater necessity. Schimek (1996) and Bento et al. (2005) demonstrated that households had fewer cars when their locations were close to the centre of the city. From the built environment category, it was found that increased population and residential density had a negative effect on car ownership (Li et al. 2010; Hess and Ong 2002). In addition, both Chu (2002) and Potoglou and Kanaroglou (2006) observed that car ownership decreased when the land-use mix increased.

Another important determinant of car ownership is the transit accessibility measure usually captured as the proximity to transit stations (bus/rail), and transit supply. Increased transit access and high quality of transit service has a significant negative effect on the number of automobiles owned (Potoglou and Kanaroglou 2008; Bento et al. 2005; Kim and Kim 2004). Schimek (1996) and Hess and Ong (2002) illustrated that traditional neighbourhoods with friendly walking and biking environments tended to reduce car ownership.

## 2.1 Current research

Despite the enormity of literature, it is surprising that there are very few studies in the context of Canadian urban regions (Potoglou and Kanaroglou 2008; Roorda et al. 2000). The most recent study, conducted for the city of Hamilton, Ontario, Canada, was based on an internet survey that considered respondents who were employees of either City of Hamilton or McMaster University. The dataset employed in the analysis does not reflect the overall vehicle ownership preference of urban residents in Hamilton. The first objective of our study is to address this limitation. We propose to estimate a vehicle ownership model using data from an entire metropolitan area, specifically the Quebec City region. The second objective of our study is to investigate the potential existence of population heterogeneity in the context of vehicle ownership. Towards this end, we propose the application of the latent class version of the ordered and unordered response models. Specifically, we estimate latent segmentation based ordered logit (LSOL) and latent segmentation based multinomial logit (LSMNL) models. Finally, we also undertake a comparison exercise of the latent class models with their traditional counterparts in the choice context examined.

A number of earlier studies assume that the influence of exogenous variables remain the same for the entire population. To illustrate the importance of varying impact of exogenous variables, let us consider the car ownership decision outcomes of two households (H1 and H2) with the same attributes except for transit accessibility variable; H1 has low accessibility and H2 has high accessibility. Now let us consider the influence of “number of employed adults” variable in these households. H1, with low transit accessibility, is inclined to have higher vehicle ownership with increased number of employed adults. On the other hand, for H2, the household with high transit accessibility, the increasing number of employed adults might not increase vehicle ownership (at least not at the same magnitude as for H1). This is an example of how

transit accessibility moderates the influence of “number of employed adults” in determining vehicle ownership. If instead of estimating a latent segmentation model, we impose population homogeneity on the “number of employed adults” variable, the resulting coefficient would be incorrect. The illustration provided is a case of one variable (transit accessibility) moderating the influence of another variable (number of employed adults). However, in the context of car ownership, it is possible that multiple variables might serve as a moderating influence on a reasonably large set of exogenous variables. The proposed latent class models provide a tractable approach to accommodate such moderations. Of course, the results from the analysis need to be examined carefully by the analyst to ensure that the outputs are not just statistical manifestations but are based on intuition and past evidence from literature.

A common approach employed to relax the homogeneity assumption is to employ mixed versions of the ordered and unordered models (Eluru and Bhat 2007; Bhat 1998; Nobile et al. 1997; Bjorner and Leth-Petersen 2007; Nolan 2010). These approaches, though attractive, are focussed on the error component of the model and usually require extensive simulation for model estimation. The advances in simulation have resulted in the widespread use of these approaches. However, one disadvantage is that they do not capture the heterogeneity corresponding to observed variables (systematic heterogeneity) in the modeling framework. Another alternative for addressing systematic heterogeneity is to introduce interaction effects of various exogenous variables. For instance, in the example described above, it is possible to interact the transit accessibility variable with “number of household workers” variable. While this will definitely improve the model, it might not always be adequate to capture the variability in the data<sup>2</sup>. In such contexts, latent class models offer an alternative approach to accommodating

---

<sup>2</sup> To illustrate the difference between the latent segmentation model and a traditional model with interactions, we explore the influence of transit accessibility variable. Specifically, we estimate the traditional models with transit accessibility interactions and a latent segmentation model with transit



heterogeneity within the systematic component. Recent research in various transportation fields has seen a revival of interest in the latent class models (Eluru et al. 2012; Yasmin et al. 2014; Sobhani et al. 2014; Greene and Hensher 2003; Bhat 1997; Xie et al. 2012). However, the role of systematic heterogeneity in the car ownership context has not been investigated in the existing literature.

### 3. Model structure and estimation

The latent class approach recognizes that households can be probabilistically assigned to different behaviourally similar segments as a function of observed attributes (Bhat, 1997; Srinivasan et al. 2009). Since the segments are unobserved to the analyst, they are termed as latent or endogenous. Within each segment, separate vehicle ownership models predict household choice behavior. The mathematical formulations are provided in the Appendix B.

The model estimation process begins with a model considering two segments. The final number of segments is determined by adding one segment at a time until further addition does not enhance intuitive interpretation and data fit (Tang and Mokhtarian 2009; Eluru et al. 2012). The evaluation of the model fit in terms of the appropriate number of segments is based on the Bayesian Information Criterion (BIC)<sup>3</sup>. Estimation of the model is terminated when the increase in the number of segments results in an increase in the BIC value. Finally, the number of segments corresponding to the lowest value of BIC is considered the appropriate number of segments. The decision regarding the optimal number of classes should be taken considering the significance of the number of parameters and the interpretability as well as parsimony of the model (Beckman and Golias 2008; Bujosa et al. 2010). The model estimates provide the

---

accessibility as a segmentation variable. The estimation results of the traditional models (OL and MNL) and latent segmentation models for OL and MNL are presented in Appendix A.

<sup>3</sup>The BIC for a given empirical model is equal to  $[-2(\text{LL}) + K \ln(Q)]$ , where (LL) is the log likelihood value at convergence, K is the number of parameters, and Q is the number of observations. BIC is found to be the most consistent Information Criterion (IC) for correctly identifying the appropriate number of segments in latent segmentation models (for more details, see Nylund et al. 2007; Roeder et al. 1998).

segment characteristics, the segment specific discrete choice model estimates and number of segments.

#### **4. Data**

The proposed latent segmentation models are estimated using data derived from the Origin-Destination (O-D) surveys of Quebec City for the year 2001. The Quebec City database contained a total of 27,822 household data. After removing inconsistent and missing/miscoded values, we were left with 26,362 usable household records. From this, we randomly sampled 5,218 records for estimation and 1,326 records for model validation purpose.

Car ownership levels in the dataset were classified as no car, one car, two cars, and three or more cars. The dependent variable was truncated at three because the number of households with more than three automobiles was relatively small in the dataset. Table 1 provides a summary of the characteristics of the sample used in this study. The distribution of auto ownership levels in the estimation sample indicate that the number of two or more cars owning households is noticeably higher (42.7%) in Quebec City. From the descriptive analysis, we can observe that about 37 percent of the households have two or more full-time workers, about 9 percent have one or more part-time workers, and about 70 percent have two or more license holders. About three-quarters of the households respectively have no children and no retirees, and more than two-thirds have no students.

#### **5. Empirical analysis**

##### **5.1 Variables considered**

The variables considered in our analysis can be broadly categorized into two categories: (1) household socio-demographic characteristics and (2) land use patterns. The demographic variables that were employed in our analysis included number of children, number of employed

adults (full-time and part-time), presence of executives, number of retirees, number of students, number of transit pass holders, number of household members and number of licensed drivers.

In order to assess the impact of different land use characteristics on car ownership, three indicators were used: residential density, entropy index ( $E_i$ ) representing land use mix, and transit accessibility ( $A_j$ ). For all calculations involving residential density, only residential land use area was used. The entropy index,  $E_i$  is defined as:  $E_i = - \sum_k \frac{[p_k \ln p_k]}{\ln(K)}$ , where:  $p_k$  is the proportion of the developed land in the  $k$ th land use type. In our study, five ( $K = 5$ ) land use types were considered including residential, commercial, industrial, institutional<sup>4</sup> and park facilities. The value of this index varies between zero and one (since the measure was normalized by  $\ln(K)$ , zero (no mix) corresponds to a homogenous area characterized by single land use type and one to a perfectly heterogeneous mix). This index has been used in numerous studies for measuring land use mix (Chu 2002; Kockelman 1997; Potoglou and Kanaroglou 2008; Miranda-Moreno et al. 2011).

The transit accessibility indicator takes into account the number of bus lines in the vicinity of the household, distance (in km) from the household to the closest bus stop of each of these lines ( $d_{ij}$ ), and the average daily headway for each of these lines ( $\bar{h}_i$ ). The formula for transit accessibility is:  $A_j = \sum_{i=1}^n \frac{1}{d_{ij} * \bar{h}_i}$ . This means that as the bus-stop distances and/or headways increase, the transit accessibility of household's decreases (Miranda-Moreno et al. 2011). On the other hand, a stop being closer or a smaller headway would mean a larger contribution to transit accessibility. This variable was used as a proxy for the level-of-service (LOS) measure of the local public transit system.

---

<sup>4</sup>Institutional land use refers to land uses that cater to community's social and educational needs (schools, town hall, police station) while park facilities refer to land used for recreational or entertainment purposes.

The final specification was based on a systematic process of removing statistically insignificant variables (in our analysis we considered 90 percent significance level) and combining variables when their effects were not significantly different. The specification process was also guided by prior research, intuitiveness and parsimony considerations.

## 5.2 Model specification and performance evaluation

In this research, we considered three different model specifications from both ordered and unordered choice mechanism. From the ordered category we estimated: (1) traditional ordered logit (OL) model, (2) latent segmentation based ordered logit model with two segments (LSOL II) and (3) latent segmentation based ordered logit model with three segments (LSOL III). From the unordered category we estimated: (1) traditional multinomial logit (MNL) model, (2) latent segmentation based multinomial logit model with two segments (LSMNL II) and (3) latent segmentation based multinomial logit model with three segments (LSMNL III). The six models were estimated using the car ownership dataset for Quebec City.

Prior to discussing the model results, we compare the performance of the OL, LSOL II and LSOL III models as well as the MNL, LSMNL II and LSMNL III models. These models are not nested within each other. Hence, for evaluating their performance, we employ the Bayesian Information Criterion (BIC) measure. The model with the lowest value of BIC is preferred. The BIC values for the final specifications of the OL, LSOL II and LSOL III, MNL, LSMNL II and LSMNL III models are 7398, 7298, 9063, 7469, 7219 and 10334, respectively. These test statistics clearly prove that the specifications with two segments (LSOL II and LSMNL II) outperform all the other models within their respective regimes. Moreover, if more than two classes are included in the model, the third group represent only a small portion of the total households and thus does not yield any interpretable segment characteristics. Moreover, the LSMNL II has the lowest BIC value indicating that it fits the data better than the LSOL II model.

These results provide strong evidence in favour of our hypothesis that car ownership of households can be better investigated through segmentation of households. From here on, we restrict ourselves to the discussion of only the LSOL II and LSMNL II models. The results for the traditional models are presented in Appendix C.

### 5.3 Behavioral interpretation

Prior to discussing the impact of various coefficients on segmentation and car ownership, it is important to discuss the overall segmentation characteristics. The model estimations can be used to generate information regarding: (1) percentage household share across the two segments and (2) overall car ownership level shares within each segment. These estimates are provided in Table 2. Strikingly, we notice that the various measures computed for the LSOL II and LSMNL II exhibit very similar trends. In fact, the similarity across the ordered and unordered models confirms the presence of segmentation in the sample population.

In the two models, the likelihood of households being assigned to segment 1 is substantially higher than the likelihood of being assigned to segment 2. Further, the car ownership probabilities for households, conditional on their belonging to a particular segment, indicate that the two segments exhibit very distinct car ownership profiles. The households allocated to segment 1 are less likely to own zero cars (only 7% or 8%) whereas the households assigned to segment 2 are less likely to own 3 or more cars (only 2%). We also estimated the mean values of the segmentation variables within each segment to characterize and explain each segment more intuitively (Table 2, see Bhat 1997 for details on computing these means). Based on the differences in the mean values of the segmentation variables, we can observe that the variables transit accessibility and transit pass holders offer the most substantial differences across the two segments. Hence, we employ these two variables to characterize our segments: segment 1 as *transit averse (TA)* and segment 2 as *transit friendly (TF)*.

## 5.4 Estimation results

### 5.4.1 Latent segmentation component

The LSOL II and LSMNL II model estimation results, for the segmentation component and the car ownership components for the two segments for Quebec City are presented in Table 3. In the following discussion, we discuss the variable effects on car ownership for the LSOL II and LSMNL II model simultaneously.

The latent segmentation component determines the probability that a household is assigned to one of the two segments identified. In our empirical analysis, Segment-1 is chosen to be the base and the coefficients presented in the table correspond to the propensity for being a part of the Segment-2. The constant term clearly indicates a larger likelihood for households being part of Segment-1. We found that the segment share is influenced by socio-demographic characteristics of household as well as land-use patterns. The attributes include: transit accessibility, entropy index, number of transit pass holders, number of household members and if any employed member of the household holds an executive position.

For all segmentation variables, both systems offer similar behavior. An increase in transit accessibility is likely to increase the probability that the household is part of Segment-2. With increase in the land use mix, represented by the entropy index, the likelihood of assigning the households to Segment-2 increases. Higher values of the entropy index imply that household members have the option to easily access many activities and amenities by walking or biking in addition to riding transit, thereby minimizing their need to procure and use cars (Cervero and Kockelman 1997; Hess and Ong 2002). Again, the higher the number of transit-pass holders in a household, the higher is the likelihood for assigning the household to Segment-2. Interestingly, households with two or more than two members were also more likely to be part of

Segment-2. As expected, increased presence of executive job holders increases the chance that households would be assigned to Segment-1.

#### *5.4.2 Car ownership component: Segment-1*

Households with more employed adults (both full-time and part-time) and persons with driving license were associated with higher levels of car ownership; an indicator that these households have greater mobility needs (Kim and Kim 2004; Potoglou and Kanaroglou 2008). The effect of full-time working adults is greater than that of part-time working adults. This is expected since full-time working adults have greater time-constraints and daily commitments, hence greater needs for personal vehicles. Gradually increasing alternative specific coefficients of full-time working adults and license holders in the LSMNL II imply that their effect on household's utility is higher as levels of car ownership increases.

Interestingly, number of children was associated with reduced likelihood of owning multiple cars. The result might seem counterintuitive at first glance. However, the negative effect of increased number of children on car ownership could be explained by the increased living expenses (food, clothing, and housing) that might curtail the amount of financial resources available for expenditures on acquiring and maintaining cars (Bhat and Koppelman 1993; Soltani 2005). The negative coefficients are gradually increasing, meaning that households associate greater disutility to multiple vehicle ownership levels with increase in the number of children. Similar to number of children, number of students also had a significant negative impact on car ownership. It is expected because households with more students would have increased budget constraints and hence, would be less inclined to own cars. Moreover, students may share their activities with friends and other household members that might further reduce the need for owning multiple cars (Vovsha et al. 2003). The result of the LSMNL II indicates that the likelihood of owning three or more cars decreases with increase in number of students in

households. We also found that increase in number of retirees was associated with increased likelihood of owning multiple cars. Please note that the variable was significant in the LSMNL II model only. The finding is probably indicating that households with more retired persons are in a financially healthy situation (Matas and Raymond 2008). Further, it is possible that these individuals prefer car mode for participating in activities.

The only land use variable that affected car ownership in this segment was residential density. As expected, the results indicated that as the residential density increases, the likelihood of households owning more cars decreases. The effect was found significant in both latent segmentation models. The signs of the coefficients as well as their magnitudes in the model show the expected trend (gradual increase in the disutility with increasing car ownership levels in the LSMNL II model). Households in denser areas tend to have fewer cars presumably due to lower car level-of-service (LOS) resulting from congestion, parking space constraints leading to escalated parking cost (Bhat and Koppelman 1993) as well as more frequent and easily accessible public transport services (Hess and Ong 2002). The lower speed in the dense residential zones might also be another deterrent to increased car ownership (Karlaftis and Golias 2002).

#### *5.4.3 Car ownership component: Segment-2*

With increase in the number of employed adults (both full-time and part-time) in households, the likelihood of owning multiple cars increases (same as Segment-1). The results of the LSMNL II model show that higher number of workers increases the probability of households owning one or two cars (relative to zero or 3+ cars). This is expected since households in this segment have better transit accessibility and improved land use mix which might be obviating the need for purchasing and using more vehicles.



Similar to Segment-1, number of licensed drivers emerged as another important factor affecting car demand in Segment-2 as well. The number of licensed drivers was used as a surrogate for potential drivers in the household. The increase in potential drivers is more likely to increase the car ownership level of households. It is interesting to note that the contribution of licensed drivers reduces for the 3 or more car ownership category for the MNL system. The result indicates that the increase in the utility for households is not the same for car ownership levels of 3 and higher. Number of children has a negative impact on car ownership decision of households in LS (same as Segment-1). Interestingly, households with higher number of students had higher likelihood of owning more vehicles in both models. With increase in number of retirees, households in Segment-2 have a higher likelihood of purchasing multiple cars. From the LSMNL II estimates, it is found that increased number of retirees increased the probability of households owning two cars. This might be explained by the fact that retirees, who presumably have the time flexibility to take frequent leisure trips, are more likely to be dependent on cars for their mobility needs due to old age.

Overall, we see that the results for the LSOL II and LSMNL II models offer very similar interpretations. The difference in the mathematical framework and the differences in the formulation of the two frameworks can lead to the minor differences we observe. The results clearly underscore the importance of considering population heterogeneity through latent class models in the context of car ownership. Further, we also tend to observe that the additional flexibility of the MNL regime allows the LSMNL II model to better explain the dependent variable.

## 5.5 Validation results

To ensure that the statistical results obtained were not a manifestation of over fitting to data, we evaluate the performance of the estimated models on a hold-out validation sample (1326

household data). This subsample of data was set aside during model estimation. Our validation analysis is conducted for the LSOL II and LSMNL II models.

To undertake the validation exercise, we employ the final parameters of the models to predict alternative probabilities for the households in the hold-out sample. To evaluate the performance, we compute both aggregate and disaggregate measures of fit. At the aggregate level, we compare the predicted<sup>5</sup> and actual auto ownership level shares and compute the root mean square error (RMSE) as well as the mean absolute percentage error (MAPE) of the predicted shares. At the disaggregate-level, we compute the predictive log-likelihood which is computed by calculating the log-likelihood for the predicted probabilities of the sample (Eluru et al. 2008). The results are reported in Table 4.

From Table 4, MAPE statistic shows that the LSMNL II model performs better than the LSOL II model in the overall for Quebec City dataset. The predictive performance from the LSMNL II model is also superior compared to that of the LSOL II model based on the predictive log-likelihood value. Hence, there is enough evidence to suggest that LSMNL II performs slightly better in the empirical analysis compared to its ordered counterpart.

## 6. Summary and conclusions

There has been substantial interest in the transportation and planning literature on examining the factors that influence household car ownership levels. The topic is of great interest to policy makers given the growing focus on global warming, public health, and sustainable development issues. Two alternative model structures: ordered and unordered, have been extensively applied in the empirical studies to examine the underlying choice process for household's auto ownership preferences. These studies assume that the influence of exogenous variables remain

---

<sup>5</sup>The aggregated predicted probabilities of car ownership outcome  $k$  of households belonging to a particular segment  $s$  can be calculated using the following equation:  $\frac{\sum_q P_{qs} \times [P_q(k)|s]}{Q}$  and the overall predicted share is obtained by summing these probabilities over  $s$ .

the same for the entire population, although it is possible that the exogenous variable effects might vary across the population. The current research proposes the use of latent class modeling approach in the context of vehicle ownership. In latent class model, segment membership is probabilistically determined as a function of the socio-demographic and land use attributes of households. The approach accommodates heterogeneity within the systematic component as opposed to heterogeneity within the unobserved component captured in the simulation based mixed model approaches.

In our study, we estimate latent segmentation based ordered logit (LSOL) and latent segmentation based multinomial logit (LSMNL) models of car ownership using the data from Quebec City region of Canada. Using several goodness of fit criteria, we conclude that the model specification with two-segments offered the best data fit. For Quebec City, the probability of belonging to any segment was a function of land use characteristics (transit accessibility and entropy index) and household demographics (number of transit pass holders, presence of more than two household members and executive job position holding by any household member). Based on the differences in the mean values of the segmentation variables, we characterized our segments: segment 1 as *transit averse (TA)* and segment 2 as *transit friendly (TF)*. In Segment-1, higher number of employed adults and license holders increase the propensity for more cars, while increased number of children and students reduce the propensity. For Segment-2, in addition to number of employed adults and license holders, number of retirees was associated with increased car ownership of households. Similar to Segment-1, number of children and students has a negative impact on household's decision to own higher number of cars.

We also assess the relative performance of the LSOL and LSMNL models with their traditional counterparts using several measures of fit for hold-out validation samples. A

consistent result that emerges from all the different fit measures and for all data sets is that the latent models outperform the traditional models. It indicates that from behavioural viewpoint, the latent class choice mechanism is a better representation of household's auto ownership decision process. In summary, our comparative analysis clearly offers evidence in favour of the hypothesis that car ownership can be better examined through segmentation of households. Moreover, between the two latent class models, the unordered choice mechanism appears to perform slightly better than the ordered response mechanism. For a better understanding of the impacts of exogenous factors, we compute the relevant elasticities (presented in Appendix D) for changes in selected variables (number of employed adults, number of children, transit pass holders, transit accessibility and residential density). The elasticity effects indicated that both full-time working adults and part-time working adults increase household car ownership levels. On the other hand, increase in number of children, transit pass holders, transit accessibility and residential density, reduced the probability of multiple vehicle ownership. Between the two land use attributes, residential density was found to have a greater impact on car ownership levels.

To investigate further into the matter, future studies could extend the comparison exercise between the latent class models. From the ordered regime, latent class version of the generalized ordered logit model (LSGOL) as proposed in Yasmin et al., 2014 could be estimated.

### **Acknowledgements**

The corresponding author would like to acknowledge financial support from Natural Sciences and Engineering Research Council (NSERC) of Canada under the Discovery Grants program and for undertaking the research. The authors would like to acknowledge useful feedback from three anonymous reviewers and Editor Professor Patricia Mokhtarian on a previous version of the paper.

## References

- Beckman JD, Goulias KG.: Immigration, residential location, car ownership, and commuting behavior: A multivariate latent class analysis from California. *Transportation* 35 (5), 655-671 (2008)
- Bento AM, Cropper ML, Mobarak AM, Vinha K.: The impact of urban spatial structure on travel demand in the United States. *Rev. Econ. Stat.* 87 (3), 466-478 (2005)
- Bhat CR.: Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transp. Res. A* 32A (7), 495-507 (1998)
- Bhat CR.: Endogenous segmentation mode choice model with an application to intercity travel. *Transp. Sci.* 31 (1), 34-48 (1997)
- Bhat CR, Koppelman FS.: An endogenous switching simultaneous equation system of employment, income, and car ownership. *Transp. Res. A* 27 (6), 447-459 (1993)
- Bhat CR, Pulugurta V.: A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transp. Res. B* 32 (1), 61-75 (1998)
- Bjørner TB, Leth-Petersen S.: A dynamic random effects multinomial logit model of household car ownership. *Natl. Tidsskr.*, 145 (1), 83-100 (2007)
- Brownstone D, Golob TF.: The impact of residential density on vehicle usage and energy consumption. *J. Urban Econ.* 65 (1), 91-98 (2009)
- Bujosa A, Riera A, Hicks RL.: Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environ. Resour. Econ.* 47 (4), 477-93 (2010)
- Caulfield B.: An examination of the factors that impact upon multiple vehicle ownership: The case of Dublin, Ireland. *Transp. Policy* 19 (1), 132-138 (2012)
- Cervero R, Kockelman K.: Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. D* 2 (3), 199-219 (1997)

- Chu Y-L.: Automobile ownership analysis using ordered probit models. *Transp. Res. Rec.* 1805, 60-67 (2002)
- Dargay JM.: Determinants of car ownership in rural and urban areas: a pseudo-panel analysis. *Transp. Res. E* 38 (5), 351-366 (2002)
- de Jong G, Fox J, Daly A, Pieters M, Smit R.: Comparison of car ownership models. *Transp. Rev.* 24 (4), 379-408 (2004)
- Eluru N, Bhat CR.: A joint econometric analysis of seat belt use and crash-related injury severity. *Accid. Anal. Prev.* 39 (5), 1037-1049 (2007)
- Eluru N, Bhat CR, Hensher DA.: A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid. Anal. Prev.* 40 (3), 1033-1054 (2008)
- Eluru N, Bhat CR, Pendyala RM, Konduri KC.: A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation* 37 (4), 603-626 (2010)
- Eluru N, Bagheri M, Miranda-Moreno LF, Fu L.: A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. *Accid. Anal. Prev.* 47 (1), 119-127 (2012)
- Greene WH, Hensher DA.: A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp. Res. B* 37 (8), 681-698 (2003)
- Handy S, Cao X, Mokhtarian P.: Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transp. Res. D* 10 (6), 427-444 (2005)
- Hess DB, Ong P.: Traditional neighborhoods and automobile ownership. *Transp. Res. Rec.* 1805, 35-44 (2002)

- Karlaftis M, Golias J.: Automobile ownership, households without automobiles, and urban traffic parameters: are they related? *Transp. Res. Rec.* 1792, 29-35 (2002)
- Kim HS, Kim E.: Effects of public transit on automobile ownership and use in households of the USA. *Rev. Urban Reg. Devel. Stud.* 16 (3), 245-260 (2004)
- Kockelman K.: Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay area. *Transp. Res. Rec.* 1607, 116-125 (1997)
- Kuhnimhof Y, Zumkeller D, Chlond B.: Who made peak car, and how? a breakdown of trends over four decades in four countries. *Transp. Rev.* 33 (3), 325-342 (2013)
- Li J, Walker J, Srinivasan S, Anderson W.: Modeling private car ownership in China. *Transp. Res. Rec.* 2193, 76-84 (2010)
- Matas A, Raymond JL.: Changes in the structure of car ownership in Spain. *Transp. Res. A* 42 (1), 187-202 (2008)
- Miranda-Moreno L, Bettex L, Zahabi S, Kreider T, Barla P.: Simultaneous modeling of endogenous influence of urban form and public transit accessibility on distance traveled. *Transp. Res. Rec.* 2255, 100-109 (2011)
- Mohammadian A, Miller E.: Empirical investigation of household vehicle type choice decisions. *Transp. Res. Rec.* 1854, 99-106 (2003)
- Canada NR.: 2009 Canadian vehicle survey summary report. Office of Energy Efficiency, Ottawa, Canada (2009)
- Millard-Ball A, Schipper L.: Are we reaching peak travel? trends in passenger transport in eight industrialized countries. *Transp. Rev.* 31 (3), 357-378 (2011)
- Nobile A, Bhat CR, Pas EI.: A random-effects multinomial probit model of car ownership choice. In: Gatsonis C, Hodges J, Kass R, McCulloch R, Rossi P, Singpurwalla N (Eds.) *Case studies in Bayesian statistics*, vol 121. Springer, New York, pp 419-434 (1997)

- Nolan A.: A dynamic analysis of household car ownership. *Transp. Res. A*, 44 (6), 446-455 (2010)
- Nylund KL, Asparouhov T, Muthén BO.: Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struc.Equ. Model.* 14 (4), 535-569 (2007)
- Potoglou D, Kanaroglou PS.: Modelling car ownership in urban areas: a case study of Hamilton, Canada. *J. Transp. Geogr.* 16 (1), 42-54 (2008)
- Roeder K, Lynch KG, Nagin DS.: Modeling uncertainty in latent class membership: a case study in criminology. *J. Amer. Statistical Assoc.* 94 (447), 766-776 (1999)
- Roorda M, Mohammadian A, Miller E.: Toronto area car ownership study: a retrospective interview and its applications. *Transp. Res. Rec.* 1719, 69-76 (2000)
- Schimek P.: Household motor vehicle ownership and use: how much does residential density matter?. *Transp. Res. Rec.* 1552, 120-125 (1996)
- Srinivasan KK, Naidu GM, Sutrala T.: Heterogeneous decision rule model of mode choice incorporating utility maximization and disutility minimization. *Transp. Res. Rec.* 2132, 59-68 (2009)
- Sobhani A, Eluru N, Faghih-Imani A.: A latent segmentation based multiple discrete continuous extreme value model. *Transp. Res. B.* 58, 154-169 (2013)
- Soltani, A.: Exploring the impacts of built environments on vehicle ownership. In: *Proceedings of the Eastern Asia Society for Transportation Studies*, vol. 5, pp. 2151-2163 (2005)
- Tang W, Mokhtarian PL.: Accounting for taste heterogeneity in purchase channel intention modeling: an example from northern California for book purchases. *J. Choice Modelling.* 2 (2), 31-55 (2009)



- van Acker VV, Witlox F.: Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *J. Transp. Geogr.* 18 (1), 65-74 (2010)
- Vovsha P, Petersen E, Donnelly R.: Explicit modeling of joint travel by household members: statistical evidence and applied approach. *Transp. Res. Rec.* 1831, 1-10 (2003)
- Whelan G.: Modelling car ownership in Great Britain. *Transp. Res. A* 41 (3), 205-219 (2007)
- Wu G, Yamamoto T, Kitamura R.: Vehicle ownership model that incorporates the causal structure underlying attitudes toward vehicle ownership. *Transp. Res. Rec.* 1676, 61-67 (1999)
- Xie Y, Zhao K, Huynh N.: Analysis of driver injury severity in rural single-vehicle crashes. *Accid. Anal. Prev.* 47, 36-44 (2012)
- Yamamoto T.: Comparative analysis of household car, motorcycle and bicycle ownership between Osaka metropolitan area, Japan and Kuala Lumpur, Malaysia. *Transportation* 36 (3), 351-366 (2009)
- Yamamoto T, Kitamura R, Kimura S.: Competing-risks-duration model of household vehicle transactions with indicators of changes in explanatory variables. *Transp. Res. Rec.* 1676, 116-123 (1999)
- Yasmin S, Eluru N, Bhat CR, Tay R.: A latent segmentation generalized ordered logit model to examine factors influencing driver injury severity. *Analytic Methods Accid. Res.*, 1, 23-38 (2014)

### **Author Biographies**

**Sabreena Anowar** is currently a PhD candidate in Transportation engineering at the McGill University in Montreal, Canada. Her primary research interests include application of discrete

choice models in the fields of transportation planning and safety. She received her MSc degree in Civil Engineering from the University of Calgary in Alberta and her BSc in Civil Engineering from Bangladesh University of Engineering and Technology (BUET) in Dhaka, Bangladesh.

**Shamsunnahar Yasmin** is currently a PhD Candidate in Transportation engineering at the McGill University in Montreal, Canada. Her primary research interests include road safety, transportation modeling, and traffic data analysis. She received her MSc degree in Civil Engineering from the University of Calgary in Alberta and her BSc in Civil Engineering from Bangladesh University of Engineering and Technology (BUET) in Dhaka, Bangladesh.

**Naveen Eluru** is an Assistant Professor in Transportation at McGill University in Montreal. His primary research areas of interest include transportation planning, socio-demographic and land-use modeling, sustainable urban design, integrated demand supply models, activity time-use analysis and transportation safety. He is currently a member of Transportation Research Board (TRB) committee on Statistical Methods (ABJ80).

**Luis F Miranda-Moreno** is an Assistant Professor in Transportation at McGill University in Montreal. His research interests include the development of crash-risk analysis methods, the integration of emergency technologies for traffic monitoring, the impact of climate on transportation systems, the analysis of short and long-term changes in travel demand, the evaluation of energy efficiency measures and non-motorized transportation.

**TABLE 1** Summary statistics of variables

Variables	Frequency	%
<i>Car Ownership Levels of Households</i>		
0 Car	562	10.8
1 Car	2427	46.5
2 Cars	1886	36.1
≥ 3 Cars	344	6.6
<i>Household Demographics</i>		
Number of Full-time Employed Adults		
0	1510	28.9
1	1795	34.4
≥ 2	1913	36.7
Number of Part-time Employed Adults		
0	4763	91.3
1	437	8.4
≥ 2	18	0.3
Number of License Holders		
0	342	6.6
1	1231	23.6
≥ 2	3645	69.8
Number of Children		
0	3791	72.7
1	638	12.2
≥ 2	789	15.1
Number of Students		
0	4234	81.1
1	755	14.5
≥ 2	229	4.4
Number of Retirees		
0	3722	71.3
1	919	17.6
≥ 2	577	11.1
Sample size, N	5218	100

**TABLE 2** Segment characteristics and mean values of segmentation variables (N = 5218)

	Latent OL		Latent MNL		
	Segment-1	Segment-2	Segment-1	Segment-2	
Household share	0.71	0.29	0.80	0.20	
Car ownership within each segment					
0 Car	0.08	0.18	0.07	0.26	
1 Car	0.41	0.57	0.44	0.53	
2 Cars	0.39	0.22	0.41	0.18	
≥ 3 Cars	0.11	0.02	0.08	0.02	
Mean Values of Demographic and Land Use Variables in Each Segment					
	Overall Market	Segment-1	Segment-2	Segment-1	Segment-2
Transit Accessibility	317.23	275.9	418.5	281.97	454.90
Entropy Index	0.37	0.35	0.43	0.35	0.45
Number of Transit Pass Holders	0.19	0.03	0.60	0.08	0.61
Two Persons	0.38	0.42	0.27	-	-
More than Two Persons	0.41	0.31	0.63	0.37	0.54
Executive Job Holder	0.12	0.12	0.12	0.12	0.09

**TABLE 3** Parameter estimates (N=5218)

Variables	Latent OL				Latent MNL			
	Segment-1		Segment-2		Segment-1		Segment-2	
	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat
<b>Segmentation Component</b>								
Constant	-	-	-4.2017	-6.716	-	-	-3.0933	-11.487
<i>Land Use Variables</i>								
Transit Accessibility	-	-	0.0013	3.796	-	-	0.0011	4.835
Entropy Index	-	-	2.2129	3.792	-	-	1.5455	3.455
<i>Household Demographics</i>								
Number of Transit Pass Holders	-	-	3.0622	7.330	-	-	1.642	9.256
Number of Household Members (Base: Single person)								
Two persons	-	-	0.9768	2.132	-	-	-	-
More than two persons	-	-	2.4409	4.619	-	-	0.7454	3.341
Executive Job Holder	-	-	-0.5191	-1.949	-	-	-0.6368	-2.517
<b>Car Ownership Component</b>								
Constants/Thresholds								
Threshold 1/Constant 1	1.4568	6.782	1.6718	5.379	-16.3443	-2.242	-3.3287	-5.891
Threshold 2/Constant 2	7.3418	30.792	6.1667	15.706	-23.1409	-3.150	-5.8831	-8.523
Threshold 3/Constant 3	11.7227	39.31	10.0706	18.382	-31.3775	-4.191	-4.4137	-4.228
<i>Land Use Variables</i>								
Residential Density	-0.0012	-4.488	-0.0022	-6.181	-	-	-	-
1 Car								
					-0.0032	-2.271	-0.0022	-3.545
2 Cars								
					-0.0048	-3.304	-0.0039	-4.643
≥ 3 Cars								
					-0.0067	-4.314	-0.0039	-4.643
<i>Household Demographics</i>								
Number of Full-time Employed Adults								
	0.8524	11.455	0.9945	7.327				
1 Car								
					7.6091	1.853	1.0878	4.979
2 Cars								
					8.5137	2.068	1.8853	6.858
≥ 3 Cars								
					8.8236	2.136	-	-
Number of Part-time Employed Adults								
	0.5636	3.395	0.6311	3.121				
0 Car								
					-0.6486	-3.678	-0.7045	-2.13
1 Car								
					-0.6486	-3.678	-0.7045	-2.13
2 Cars								
					-	-	-	-
≥ 3 Cars								
					-	-	-	-

Number of License Holders	3.6924	30.201	1.6565	13.74				
1 Car					19.2163	2.588	2.6001	10.636
2 Cars					22.4946	3.019	2.6001	10.636
≥ 3 Cars					25.1929	3.359	2.2289	3.885
Number of Children	-3.7143	-26.59	-1.1172	-7.615				
0 Car					-	-	2.0193	7.069
1 Car					-10.6869	-2.587	-	-
2 Cars					-13.9118	-3.350	-	-
≥ 3 Cars					-16.8293	-3.994	-	-
Number of Students	-0.3415	-2.733	0.2271	1.786				
1 Car					-	-	-0.6363	-4.373
2 Cars					-	-	-	-
≥ 3 Cars					-0.6073	-3.439	-	-
Number of Retirees	-	-	1.0505	6.187				
0 Car					-5.5836	-1.919	-	-
1 Car					-	-	-	-
2 Cars					-	-	1.2481	4.183
≥ 3 Cars					-	-	-	-
Log-likelihood at zero		-7233.68					-7233.68	
Log-likelihood at sample shares		-5964.82					-5964.82	
Log-likelihood at convergence		-3568.07					-3481.08	
Log-likelihood at convergence of traditional OL and MNL		-3647.70					-3619.15	
Log-likelihood at convergence of traditional OL and MNL with interaction terms		-3641.60					-3607.67	
<i>Note: - denotes variables which are not significant. Also, the coefficient estimates across different alternatives are constrained to be same when the effects are not significantly different.</i>								

**TABLE 4** Measures of fit in the validation sample (N = 1326)

<b>DISAGGREGATE MEASURES OF FIT</b>			
Summary Statistic		LSOL II	LSMNL II
Log-likelihood at zero		-1838.23	-1838.23
Log-likelihood at sample shares		-1489.29	-1489.29
Predictive log-likelihood		-877.91	-852.71
Number of observations		1326	1326
Number of parameters estimated		19	30
Predictive adjusted likelihood ratio index		0.398	0.407
<b>AGGREGATE MEASURES OF FIT</b>			
Car Ownership Levels/ Measures of fit	Actual Shares	Predictions	
		LSOL II	LSMNL II
0 car	10.1	10.7	10.9
1 car	46.7	45.8	45.1
2 cars	37.3	36.4	37.3
≥ 3 cars	6.0	6.9	6.7
<b>RMSE</b>	-	0.83	0.88
<b>MAPE</b>	-	6.30	5.75

### Appendix A: Estimation Results of the Traditional Models

**Table A.1** Traditional Ordered Logit Model (OL) Estimates with *Transit Accessibility* Interactions (N = 5218)

Variables	Estimate	t-stat
<i>Thresholds</i>		
Threshold 1	0.2538	2.122
Threshold 2	5.2263	33.044
Threshold 3	9.0811	46.899
<i>Land Use Variables</i>		
Residential Density	-0.0075	-3.909
Transit Accessibility	-0.0085	-6.014
Entropy Index	-1.0490	-5.523
<i>Household Demographics</i>		
Number of Children	-2.4389	-30.272
Number of Full-time Employed Adults	0.6943	12.914
Number of Part-time Employed Adults	0.6115	4.227
Number of Students	-0.2255	-2.963
Executive Job Holder	0.4975	5.018
Number of License Holders	2.5471	36.298
Number of Transit Pass Holders	-1.0615	-14.12
Two-person Household	0.2689	3.439
<i>Interactions</i>		
Number of Part-time Employed Adults*Transit Accessibility	-0.0061	-1.724
Number of Retirees*Transit Accessibility	0.0042	3.557
Log-likelihood at Convergence	-3641.60	
Log-likelihood at Sample Shares	-5964.82	



**Table A.2** Latent Segmentation based Ordered Logit Model (OL) Estimates with only *Transit Accessibility* as the Segmentation Variable (N = 5218)

Variables	Segment-1		Segment-2	
	Estimate	t-stat	Estimate	t-stat
<b>Segmentation Component</b>				
Constant	-	-	-1.6363	-10.172
Transit Accessibility	-	-	0.0134	6.039
<b>Car Ownership Component</b>				
<i>Thresholds</i>				
Threshold 1	0.6958	2.724	0.6583	0.982
Threshold 2	6.0852	26.558	8.276	9.222
Threshold 3	10.8423	37.554	14.1706	6.41
<i>Household Demographics</i>				
Number of Full-time Employed Adults	0.8859	11.468	-	-
Number of Part-time Employed Adults	0.6297	3.883	-	-
Number of Children	-3.3168	-26.74	-2.1336	-7.471
Number of License Holders	3.2494	30.945	3.0808	12.241
Number of Students	-0.1860	-1.835	-0.6084	-2.769
Two-person Household	0.3073	2.443	-	-
Number of Transit Pass Holders	-1.0387	-10.887	-1.716	-6.9
Executive Job Holder	0.7622	5.009	-	-
Two-person Household	-	-	-0.6333	-1.62
<i>Land Use Variables</i>				
Residential Density	-0.0099	-3.566	-0.0131	-2.059
Entropy Index	-1.4137	-5.011	-1.5671	-1.935
Log-likelihood at Convergence	-3568.22			
Log-likelihood at Sample Shares	-5964.82			

**Table A.3** Traditional Multinomial Logit Model (MNL) Estimates with *Transit Accessibility* Interactions (N = 5218)

<b>Variables</b>	<b>Estimate</b>	<b>t-stat</b>
<i>Constants</i>		
Constant 1	-1.8042	-5.916
Constant 2	-5.952	-16.707
Constant 3	-11.7669	-23.932
<i>Land Use Variables</i>		
Residential Density		
1 Car	-0.0062	-1.862
2 Cars	-0.0135	-3.316
≥ 3 Cars	-0.0163	-2.490
Entropy Index		
1 Car	-1.5132	-3.368
2 Cars	-2.2993	-4.618
≥ 3 Cars	-3.6049	-5.648
<i>Household Demographics</i>		
Number of Children		
1 Car	-3.9217	-11.205
2 Cars	-5.7777	-15.711
≥ 3 Cars	-7.7994	-19.737
Number of Part-time Employed Adults		
1 Car	-	-
2 Cars	0.5244	3.969
≥ 3 Cars	0.7539	3.378
Number of Full-time Employed Adults		
1 Car	1.0900	6.735
2 Cars	1.7981	10.131
≥ 3 Cars	2.0468	9.801
Executive Job Holder		
1 Car	-	-
2 Cars	0.3804	3.277
≥ 3 Cars	0.8097	4.318
Number of Retirees		
1 Car	0.3561	1.849
2 Cars	0.3864	1.805
≥ 3 Cars	0.5452	2.124
Number of License Holders		
1 Car	4.2629	18.613
2 Cars	6.2085	25.023
≥ 3 Cars	7.9213	29.222
Number of Transit Pass Holders		
1 Car	-1.4514	-10.137
2 Cars	-2.3630	-14.219
≥ 3 Cars	-3.4770	-14.691
<i>Interactions</i>		
Number of License Holders*Transit Accessibility		
1 Car	-0.0124	-5.636
2 Cars	-0.0163	-6.893
≥ 3 Cars	-0.0173	-6.528
Number of Children*Transit Accessibility		
1 Car	0.0139	2.958
2 Cars	0.0172	3.360
≥ 3 Cars	0.0202	2.966
Number of Retirees*Transit Accessibility		
1 Car	5.5952	2.009
2 Cars	9.4581	2.891
≥ 3 Cars	9.4581	2.891
Log-likelihood at Convergence	-3607.67	
Log-likelihood at Sample Shares	-5964.82	

**Table A.4** Latent Segmentation based Multinomial Logit Model (OL) Estimates with only *Transit Accessibility* as Segmentation Variable (N = 5218)

Variables	Segment-1		Segment-2	
	Estimate	t-stat	Estimate	t-stat
<b>Segmentation Component</b>				
Constant	-	-	-2.4578	-15.551
Transit Accessibility	-	-	0.0124	6.071
<b>Car Ownership Component</b>				
<i>Constants</i>				
Constant 1	-4.8171	-4.571	-2.3356	-3.216
Constant 2	-10.5113	-9.547	-7.3323	-2.714
Constant 3	-17.5067	-14.762	-4.0809	-5.186
<i>Land Use Variables</i>				
Residential Density				
1 Car	-0.0155	-2.192	-	-
2 Cars	-0.0272	-3.551	-	-
≥ 3 Cars	-0.0337	-3.567	-	-
Entropy Index				
1 Car	-2.7973	-2.17	-3.4785	-3.352
2 Cars	-4.0389	-3.066	-	-
≥ 3 Cars	-5.6967	-4.103	-	-
<i>Household Demographics</i>				
Number of Children				
1 Car	-8.964	-6.828	-	-
2 Cars	-11.8039	-8.796	-	-
≥ 3 Cars	-14.0273	-10.36	-	-
Number of Part-time Employed Adults				
1 Car	-	-	-	-
2 Cars	0.6543	3.807	-	-
≥ 3 Cars	1.055	3.906	-	-
Number of Full-time Employed Adults				
1 Car	2.0097	5.179	1.342	4.336
2 Cars	2.9361	7.331	-	-
≥ 3 Cars	3.3985	8.066	-	-
Executive Job Holder				
1 Car	-	-	-	-
2 Cars	0.6729	3.68	-	-
≥ 3 Cars	1.2247	4.977	-	-
Number of Retirees				
0 Car	-	-	1.2385	3.886
1 Car	1.5333	3.11	-	-
2 Cars	1.676	3.336	-	-
≥ 3 Cars	1.9844	3.764	-	-
Number of License Holders				
1 Car	8.627	6.888	1.7672	6.314
2 Cars	11.4739	8.987	1.9901	1.945
≥ 3 Cars	13.4881	10.482	1.095	2.114
Number of Transit Pass Holders				
1 Car	-2.1339	-7.158	-	-
2 Cars	-3.1675	-9.926	-	-
≥ 3 Cars	-4.4106	-11.789	-	-
Log-likelihood at Convergence		-3482.08		
Log-likelihood at Sample Shares		-5964.82		

## Appendix B: Mathematical Formulation of Latent Class Models

Let us consider  $S$  homogenous segments of households (the optimal number of  $S$  is to be determined). We need to determine how to assign the households probabilistically to the segments for the segmentation model. The utility for assigning a household  $q$  (1,2,.. $Q$ ) to segment  $s$  is defined as:

$$U_{qs}^* = \beta_s' z_q + \xi_{qs} \quad (\text{B.1})$$

$z_q$  is a ( $M \times 1$ ) column vector of attributes that influences the propensity of belonging to segment  $s$ ,  $\beta_s'$  is a corresponding ( $M \times 1$ ) column vector of coefficients and  $\xi_{qs}$  is an idiosyncratic random error term assumed to be identically and independently Type 1 Extreme Value distributed across households  $q$  and segment  $s$ . Then the probability that household  $q$  belongs to segment  $s$  is given as:

$$P_{qs} = \frac{\exp(\beta_s' z_q)}{\sum_s \exp(\beta_s' z_q)} \quad (\text{B.2})$$

Within the latent segmentation approach, the probability of household  $q$  choosing auto ownership level  $k$  is given as:

$$P_q(k) = \sum_{s=1}^S (P_q(k) | s)(P_{qs}) \quad (\text{B.3})$$

where  $P_q(k) | s$  represents the probability of household  $q$  choosing auto ownership level  $k$  within the segment  $s$ . Note that the choice construct of car ownership considered to compute  $P_q(k) | s$  may be either the ordered or unordered response mechanism.

Now, if we consider the car ownership levels of households ( $k$ ) to be ordered,

$$y_{qs}^* = \alpha'_s x_q + \varepsilon_{qs}, \quad y_q = k \quad \text{if } \psi_{s_{k-1}} < y_{qs}^* < \psi_{s_k} \quad (\text{B.4})$$

where  $y_{qs}^*$  is the latent propensity of household  $q$  conditional on  $q$  belonging to segment  $s$ .  $y_{qs}^*$  is mapped to the ownership level  $y_q$  by the  $\psi$  thresholds ( $\psi_{s_0} = -\infty$  and  $\psi_{s_k} = \infty$ ) in the usual ordered-response fashion.  $x_q$  is a  $(L \times 1)$  column vector of attributes that influences the propensity associated with car ownership.  $\alpha$  is a corresponding  $(L \times 1)$  column vector of coefficients and  $\varepsilon_{qs}$  is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across households  $q$ . The probability that household  $q$  chooses car ownership level  $k$  is given by:

$$P_q(k)|s = \Lambda(\psi_{s_k} - \alpha'_s x_q) - \Lambda(\psi_{s_{k-1}} - \alpha'_s x_q) \quad (\text{B.5})$$

where  $\Lambda(\cdot)$  represents the standard logistic cumulative distribution function (cdf).

If we consider the car ownership levels ( $k$ ) to be unordered, we employ the usual random utility based multinomial logit (MNL) structure. Equation (6) represents the utility  $U_{qk}$  that household  $q$  associates with car ownership level  $k$  if that household belongs to segment  $s$

$$U_{qk}|s = \alpha'_s x_q + \varepsilon_{qk} \quad (\text{B.6})$$

$x_q$  is a  $(L \times 1)$  column vector of attributes that influences the propensity associated with car ownership.  $\alpha$  is a corresponding  $(L \times 1)$ -column vector of coefficients and  $\varepsilon_{qk}$  is an idiosyncratic random error term assumed to be identically and independently generalized extreme value (GEV) distributed across households  $q$ . Then the probability that household  $q$  chooses car ownership level  $k$  is given as:

$$P_q(k)|s = \frac{\exp(\alpha'_s x_q)}{\sum_k \exp(\alpha'_s x_q)} \quad (\text{B.7})$$

The log-likelihood function for the entire dataset with appropriate  $P_q(k) | s$  for ordered and unordered regimes is provided below:

$$L = \sum_{q=1}^Q \log(P_q(k_q^*)), \quad (\text{B.8})$$

where  $k_q^*$  represents the ownership level chosen by household  $q$ .

### Appendix C: Estimation Results of the Traditional Models

**Table C.1** Traditional Ordered Logit Model (OL) Estimates with All Variables (N = 5218)

Variables	Estimate	t-stat
<i>Thresholds</i>		
Threshold 1	0.2107	1.768
Threshold 2	5.1967	32.64
Threshold 3	9.0296	46.626
<i>Land Use Variables</i>		
Transit Accessibility	-0.0073	-5.599
Entropy Index	-1.0536	-5.555
Residential Density	-0.0073	-3.809
<i>Household Demographics</i>		
Number of Transit Pass Holders	-1.0727	-14.238
Number of Household Members		
Two persons	0.4634	4.063
More than two persons	0.3516	2.171
Number of Children	-2.4805	-31.269
Number of Full-time Employed Adults	0.5816	12.001
Number of Part-time Employed Adults	0.3369	3.205
Number of Students	-0.3068	-4.027
Number of License Holders	2.5195	33.326
Executive Job Holder	0.4887	4.942
Log-likelihood at Convergence	-3647.70	
Log-likelihood at Sample Shares	-5964.82	

**Table C.2** Traditional Multinomial Logit Model (MNL) Estimates with All Variables (N = 5218)

<b>Variables</b>	<b>Estimate</b>	<b>t-stat</b>
<i>Constants</i>		
Constant 1	-1.4137	-5.109
Constant 2	-5.3681	-16.354
Constant 3	-11.1327	-23.854
<i>Land Use Variables</i>		
Residential Density		
1 Car	-0.0060	-1.748
2 Cars	-0.0128	-3.083
≥ 3 Cars	-0.0174	-2.769
Transit Accessibility		
1 Car	-0.0089	-3.738
2 Cars	-0.0159	-5.591
≥ 3 Cars	-0.0159	-5.591
Entropy Index		
1 Car	-1.4165	-3.186
2 Cars	-2.1731	-4.399
≥ 3 Cars	-3.5333	-5.63
<i>Household Demographics</i>		
Number of Children		
1 Car	-3.1663	-13.176
2 Cars	-4.9127	-18.926
≥ 3 Cars	-6.8824	-23.876
Number of Part-time Employed Adults		
1 Car	-	-
2 Cars	0.5231	3.968
≥ 3 Cars	0.7516	3.374
Number of Full-time Employed Adults		
1 Car	1.0842	6.728
2 Cars	1.7922	10.133
≥ 3 Cars	2.0430	9.812
Executive Job Holder		
1 Car	-	-
2 Cars	0.3786	3.272
≥ 3 Cars	0.8035	4.295
Number of Retirees		
1 Car	0.6376	4.621
2 Cars	0.7832	4.934
≥ 3 Cars	0.9313	4.373
Number of License Holders		
1 Car	3.6048	19.696
2 Cars	5.4324	26.646
≥ 3 Cars	7.1236	31.016
Number of Transit Pass Holders		
1 Car	-1.4623	-10.089
2 Cars	-2.3874	-14.251
≥ 3 Cars	-3.5138	-14.812
Log-likelihood at Convergence		-3619.15
Log-likelihood at Sample Shares		-5964.82



### **Appendix D: Elasticity effects**

The exogenous variable coefficients do not directly provide the magnitude of impacts of variables on the probability of car ownership levels. For better understanding the impacts of exogenous factors, we compute the relevant elasticities for changes in selected variables. The calculation results are presented in Table D.1. For the analysis, we selected three socio-demographic variables (number of employed adults, number of children and number of transit pass holders) and two land use attributes (transit accessibility and residential density). Note that the elasticity effects were computed for the OL, LSOL II, MNL and LSMNL II models.

The results illustrate that both full-time working adults and part-time working adults increase household car ownership levels. However, as expected full-time working adults had greater impact on increasing vehicle ownership levels (2 or more) compared to the part-time working adults. The impact of change in number of children demonstrates the likelihood of vehicle fleet size reduction with similar impacts in magnitude in all the models. The reduction in fleet size observed in the elasticity analysis, while counterintuitive, is consistent with the coefficients of that variable in the models and is similar across all models; in particular, with respect to the large percentage increase in zero-car households it should be kept in mind that the base proportion of those households is not very large (10%). It might be useful to investigate this result further in future analysis.

Increase in number of transit pass holders resulted in a decrease in car ownership levels. The decreasing effect was more pronounced for 3 or more car ownership level. We can also see from the table that increase in transit accessibility and residential density reduces the probability of household's owning 2 or more cars. However, between the two attributes, residential density has a greater impact on car ownership levels than transit accessibility. The

computation exercise provides an illustration of the applicability of the proposed framework for policy analysis.

**TABLE D.1** Elasticity effects<sup>6</sup> of important variables

Models	Car ownership levels	Variables Considered					
		Number of Full-time Employed Adults	Number of Part-time Employed Adults	Number of Children	Number of Transit Pass Holders	Transit Accessibility	Residential Density
OL	0 Car	-24.31	-14.72	170.26	60.48	4.12	3.19
	1 Car	-10.35	-5.96	20.79	15.78	0.41	0.39
	2 Cars	13.26	8.09	-63.16	-29.61	-1.41	-1.14
	≥ 3 Cars	39.26	21.35	-79.44	-47.77	-1.91	-1.70
LSOL II	0 Car	-32.96	-22.46	187.82	44.39	1.88	6.86
	1 Car	-14.11	-9.17	19.15	17.30	0.36	0.50
	2 Cars	19.07	13.16	-67.35	-25.17	-0.76	-2.26
	≥ 3 Cars	49.15	29.62	-77.61	-57.40	-1.46	-2.54
MNL	0 Car	-32.58	-1.38	174.08	68.29	4.13	2.11
	1 Car	-14.81	-15.92	11.21	13.89	0.50	0.71
	2 Cars	23.55	16.50	-51.46	-27.30	-1.77	-1.25
	≥ 3 Cars	28.52	24.05	-81.36	-59.85	-0.60	-1.64
LSMNL II	0 Car	-24.20	-2.77	156.31	49.65	3.41	4.31
	1 Car	-16.54	-16.29	23.52	5.41	-0.11	1.47
	2 Cars	25.05	20.23	-62.01	-17.19	-0.73	-2.32
	≥ 3 Cars	18.34	7.74	-82.27	-25.43	-0.86	-4.69

<sup>6</sup> For the ordinal variables (number of employed adults, number of children and number of transit pass holders), the variable was increased by one unit and for the continuous variables (transit accessibility and residential density), the value was increased by 25 percent and the resulting percentage change in probability was calculated. The elasticity effects represent percentage change in the share of the dependent variable for a unit increase (increased by 1 for ordinal variables and 25% for continuous variables) in the independent variable.