MODELLING HOUSEHOLD CAR OWNERSHIP IN THE SELECTED REGIONS OF THE CZECH REPUBLIC

Martina Čampulová¹, Roman Čampula²

¹ Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00, Brno, Czech Republic
² CDV - Transport Research Centre, Líšeňská 33a, 636 00, Brno, Czech Republic

Link to this article: https://doi.org/10.11118/actaun202068030605
Received: 11. 5. 2020, Accepted: 12. 6. 2020

To cite this article: ČAMPULOVÁ MARTINA, ČAMPULA ROMAN. 2020. Modelling Household Car Ownership in the Selected Regions of the Czech Republic. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 68(3): 605–612.

Abstract

Car ownership models are essential for forecasting transport demand as well as for addressing the resulting impacts of traffic emissions and appropriate measures. The paper presents the multinomial logit model for the selected regions of the Czech Republic analysing the relationship between household car ownership and household characteristics such as household income, number of household members and number of children. The model was applied to data collected by questionnaire survey conducted in two Czech cities Litoměřice and Písek in the years 2017 and 2019. In total, three car ownership categories were specified: one car, two cars, and three or more cars. The results showed that household income and the number of household members owning a driving licence has a positive impact on owning two and more cars. Besides that, it was also found out that the number of cars owned by a household is influenced by the number of retired pensioners and a possibility to use a company car.

Keywords: car ownership, multinomial logit model, household characteristics, Czech Republic, Litoměřice, Písek

INTRODUCTION

Travelling by car providing a relatively huge level of comfort belongs to the most popular modes of transport. Car owners use cars to travel to work, shopping and leisure time activities. In some areas car ownership positively affects possibilities to obtain employment opportunities (Gurley and Bruce, 2005). Car ownership is thus an essential element of today’s life, and significantly influences travel behaviour (Pucher and Renne, 2003). In recent years there has been a significant increase in ownership of passenger cars (Ministry of Transportation, 2018), which negatively affects the traffic situation in cities, results in traffic congestion on urban roads and contributes to air pollution.

Transport is considered as Europe’s most significant source of CO₂, and one of the most critical sources of aerosol pollution (Transport & Environment, 2018; Al-Thani and Koç, 2018; Verma et al., 2010). Transport emissions also contain high contents of trace metals, organic and inorganic carbon, and inorganic anions (Coufalík et al., 2019; Agarwal et al., 2011; Maricq, 2007). Therefore the transport sector is a significant source of greenhouse gas emissions and urban air pollution. Several studies have confirmed that transport emissions have a negative impact on human health (Cheung et al., 2010; Geller et al., 2006; Hu et al., 2009; Verma et al., 2010). Car ownership thus has a negative impact on both the environment and human health.
Car ownership models are essential for forecasting transport demand and addressing the resulting impacts of traffic emissions and appropriate measures. Besides that, car ownership models are useful also in other areas such as forecasting of needs for fuels and tax revenues or investment decision making processes.

Modelling of car ownership is a common problem in the scientific environment, and a number of models have been proposed for that purpose. One of the first approaches for car ownership modelling can be found in (Beesley and Kain, 1964; Kain and Beesley, 1965). Early reviews of car ownership models are presented, e.g. in (Bates et al., 1981; Button et al., 1982; Allanson, 1982) and relatively recent comparison of car ownership models is given in (de Jong et al., 2004).

In general, the models for car ownership modelling can be divided into disaggregate and aggregate models, however, disaggregate models give better results from a statistical point of view (Potoglou and Kanaroglou, 2008; Potoglou and Susilo, 2008). Disaggregate models are further classified into unordered and ordered-response models. Unordered-response models include multinomial logit regression model while ordered-response models are represented by ordinal logit and ordinal probit models.

Multinomial logit model can be supposed to be one of the most widely used models for modelling vehicle ownership based on household characteristics as used by many studies (Zegras, 2010; Oyedepo et al., 2017; Hsu and Lin, 2007; Wedagama, 2009).

An empirical study presented in (Potoglou and Susilo, 2008) showed that unordered-response model outperforms both ordered-response models based on several measures of fit. Therefore to determine the influence of household characteristics on car ownership, the multinomial logit model is preferred in this study.

A number of research studies reported a significant relationship between the likelihood of household's car ownership and household income (Dargay, 2002; Dargay, 2001; Nolan, 2010; Oakil et al., 2014; Whelan, 2001; Maltha et al., 2017), household size and structure (Nolan, 2010; Maltha et al., 2017), number of full-time working members (Potoglou and Kanaroglou, 2008), the level of urbanisation (Jong et al., 2004; Potoglou and Kanaroglou, 2008; Oakil et al., 2016), accessibility of public transport (Zegras, 2010), number of licence holders (Whelan, 2001) and the provision of company vehicles (Whelan, 2001).

This paper aims to develop the multinomial logit model for the selected regions of the Czech Republic analysing the relationship between household car ownership and household characteristics including household income, household size, the access to a company car, the number of driving licence holders, the number of children and size of household home-town. The car ownership levels were restricted to one, two, three and more. The data were collected by questionnaire survey conducted in two Czech cities Litoměřice and Písek in the years 2017 and 2019. We expect that the results of this paper are helpful for transportation planning and solving transportation problems in Czech crowded cities with comparable size to Litoměřice and Písek.

MATERIALS AND METHODS

Data

The multinomial logit model was applied on data obtained from traffic behaviour surveys carried out in two Czech cities Litoměřice and Písek as part of the development of sustainable urban mobility plans. Both assessed cities are currently working on this plan, and traffic surveys are one of the primary data inputs. Due to the possibility of comparing the results and subsequent processing, great emphasis is placed on adhering to the same methodology of data collection. In both cases, a household survey was conducted. The sample was compiled using a random stratified address sample from a database representing an anonymous list of all households in the area.

Litoměřice is a town in the Ústí nad Labem Region in the north of the Czech Republic with a population of 24,000 inhabitants. The city management is active in the field of sustainable transport planning and testing of the use of alternative fuels. The survey was conducted in September–December 2017 on a sample of 900 households.

Písek is a town in the South Bohemian Region in the south of the Czech Republic with a population of approximately 30,000 inhabitants. The city management is active in joint transport planning, urban greenery and the use of public space. The survey was conducted in May–June 2019 on a sample of 337 households.

The surveys collected data on households, household cars, individual household members older than five years and all trips made by these members on one pre-selected day (prospective interviewing). Information on the traffic behaviour of the population was collected for one regular working day (typically Tuesday–Thursday).

The survey included, among others:

- household data:
  - household data (e.g. number of members),
  - data on the means of transport available at home (passenger cars, commercial vehicles, bicycles, etc.),
  - data on all household members (age, sex, education, economic activity, type of driving license, etc.),
  - data on the record date,
• travel data:
  • the starting and ending point of the trip,
  • start and end time of the trip,
  • used means of transport,
  • the purpose of the journey.

Model

As mentioned in the introduction, this paper analyses the relationship between the number of cars owned by a household and household characteristics in the selected regions of the Czech Republic using multinomial logistic regression (MNL) (Agresti, 2007). MNL used for modelling nominal response variables with more than two possible outcome categories is a generalisation of logistic regression for binary response variables. The principle of MNL is to estimate a function predicting the probabilities of the different possible outcomes categories of the nominal dependent variable based on regressors (independent variables) which may be both quantitative and qualitative. In the following paragraph, a short description of the model is given.

Suppose the car ownership variable \( Y \) with \( J \) outcome categories, the vector \( \mathbf{X} = (X_1, ..., X_J)' \) of regressors \( X_i, ..., X_J \) and let \( \{\pi_1, ..., \pi_J\} \) denote the probabilities \( \{\Pr(Y = 1), ..., \Pr(Y = J)\} \) of \( J \) possible response categories such that \( \sum \pi_j = 1 \). Given \( n \) different observations, the number of outcomes of the \( J \) types follows a multinomial distribution. To model the probabilities \( \pi_1, ..., \pi_J \), one category is chosen as a baseline, or reference category and the model equations are constructed as follows:

\[
\log \left( \frac{\pi_j}{\pi_1} \right) = \mathbf{X}' \mathbf{\beta}_j, \quad j = 1, ..., J - 1, \tag{1}
\]

where \( \mathbf{\beta}_j = (\beta_{1j}, ..., \beta_{Jj})' \) is a vector of unknown parameters corresponding to baseline-category logit \( \log(\pi_j/ \pi_1) \). Each model equation thus gives a log-odds that the response falls in non-baseline category \( j \) instead of baseline category \( 1 \) and the log-odds are modelled as a linear function of the regressors. Thus, an individual vector of unknown parameters for each comparison of the \( J - 1 \) non-baseline categories with the baseline is specified. It can be derived that the equations (1) can be written in the form of equations:

\[
\pi_j = \frac{e^{\mathbf{X}_j \mathbf{\beta}_j}}{1 + \sum_{j=1}^{J-1} e^{\mathbf{X}_j \mathbf{\beta}_{j'}}}, \quad j = 1, ..., J - 1, \tag{2}
\]

from which all \( J - 1 \) vectors of unknown parameters \( \mathbf{\beta}_1, ..., \mathbf{\beta}_{J-1} \) are estimated simultaneously (most often by using maximum likelihood method) which ensures smaller standard errors of the parameters than separate estimation.

RESULTS

The numbers of cars owned by households were categorised into three categories: one car (82.7% households), two cars (15.7% households), three and more cars (approximately 1.7% households). Household characteristics expected to explain the number of cars owned by the household and considered in the model included: household income, number of household members, number of children, number of economically active members (members with an economic status employee, employer, self-employed or members with its source of livelihood), number of retired non-working pensioners, number of members owning driving license B and the possibility of a household to use a company car. As shortly reviewed in the introduction, several research studies found a relationship between household’s cars ownership and the level of urbanisation as well as the accessibility of public transport. In the Czech Republic, there is a well developed urban public transport system that operates in all larger cities, and that is interconnected with suburban and regional transport within integrated transport systems. The larger the city, the more frequent the departures of private buses and trams and the denser network of urban or integrated transport stops. For this reason, a household characteristic expressing a household residence-city size was considered in the research. Specifically, the household residence-city size was categorised into three categories:

- less than 1,000 inhabitants (category 1),
- 1,000–10,000 inhabitants (category 2) and
- more than 10,000 inhabitants (category 3).

Household income was classified into six categories:

- less than 9,200 CZK,
- 9,201–15,000 CZK,
- 15,001–22,000 CZK,
- 22,001–30,000 CZK,
- 30,001–50,000 CZK,
- 50,001 CZK and more.

The median net monthly income per household member is in the Czech Republic approximately 14,700 CZK (Czech Statistical Office, 2019). It means that the household’s income less than 9200 CZK is relatively small compared with the median. For comparison, the household’s income from other categories with the median household income, the number of household members must be considered. Interviewed households with income corresponding to the second category are mainly single-member, households with income corresponding to third category have mostly one or two members, households with income 22,001–50,000 CZK are mostly two-member and households with income 50,001 CZK and more have mostly two, three or four members. Thus, considering the number of household members, we can conclude that approximately 32% of interviewed households have
income relatively high compared to the median, approximately 43% households have income comparable or slightly smaller than the median, and approximately 25% of interviewed households have income more or less low compared to the median.

Levels of individual household characteristics against car ownership levels are illustrated in Fig. 1. As can be seen, each household income category is represented by its mean value. We can also see that the number of households owning two and more cars increases with increasing household income. Three and more cars are owned by households with at least three members, and the proportion of two owned cars is growing with the growing number of household members. A relatively large number of households, whose members were found to be mainly non-working pensioners, had no economically active members. The most significant proportion of two cars is owned by a household with three and more active members. As can be further seen, most households are childless. The most significant proportion of two cars is owned by a household with one child, and the largest proportion of three and more cars is owned by households with 1 or 2 children. The graphs in Fig. 1 also show that the largest proportion of two and more cars, which is growing with the growing number of driving licenses holders, corresponds to households living in residence with less than 1,000 inhabitants. We can also see that three and more cars are owned only by households that have no possibility to use a company car.

To explain the relationship between the number of cars owned by the household and household characteristics, the multinomial logit model with both numeric and categorical variables was constructed. To determine, which variables were found to have a statistically significant effect on household car ownership levels at the 95% confidence level, together with standard deviations of the estimates are given in Tab. I. All estimations were performed using R version 3.6.1.

Based on this classifications, the numeric variables included:
• $I$ – household income with numeric values corresponding to means of above-mentioned household income intervals that is the values 8,000; 12,000; 18,500; 26,000; 40,000; 60,000 (units are CZK).
• $HP$ – the number of household members.
• $EA$ – the number of economically active persons.
• $DL$ – the number of persons owning a driving license.

The categorical variables included:
• $Ch_x$ – the number of children represented by a set of dummy variables $Ch_1$, $Ch_2$ and $Ch_3$ for 1, 2 and 3 children, respectively.
• $CS_y$ – household residence-city size represented by a set of dummy variables defined by the three, above mentioned categories as follows: $CS_1$ for category 1 (less than 1 ths. inhabitants), $CS_2$ for category 2 (1–10 ths. inhabitants) and $CS_3$ for category 3 (more than 10 ths. inhabitants).
• $CC$ – dummy variable expressing that the household has a possibility to use a company car.
• $RP_x$ – the number of retired pensioners represented by a set of dummy variables $RP_1$, $RP_2$ and $RP_3$ for 1, 2 and 3 retired pensioners, respectively.

The one car ownership category was selected as the reference model category. The unknown parameters were estimated using the maximum likelihood method, and the best submodel was found using a backward elimination process. The estimates of unknown parameters that were found to have a statistically significant effect on the household car ownership levels at the 95% confidence level, together with standard deviations of the estimates are given in Tab. I. All estimations were performed using R version 3.6.1.

---

**Table I: Estimates of unknown parameters of MNL (1) corresponding to statistically significant regressors**

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Variable</th>
<th>Estimate</th>
<th>St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3+</td>
</tr>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>-5.656</td>
<td>-12.401</td>
</tr>
<tr>
<td>Household income</td>
<td>$I$</td>
<td>$4.776 \times 10^3$</td>
<td>$8.738 \times 10^3$</td>
</tr>
<tr>
<td>Access to company cars</td>
<td>$CC$</td>
<td>-2.269</td>
<td>-</td>
</tr>
<tr>
<td>No. of persons with driving license</td>
<td>$DL$</td>
<td>1.630</td>
<td>3.570</td>
</tr>
<tr>
<td>No. of economically active persons</td>
<td>$EA$</td>
<td>-1.488</td>
<td>-</td>
</tr>
<tr>
<td>Number of retired pensioners = 2</td>
<td>$RP_2$</td>
<td>-1.089</td>
<td>-2.221</td>
</tr>
</tbody>
</table>
As already mentioned, the parameters of MNL model (1) were estimated for two alternatives (two, three and more cars), whereas the one-car ownership level was considered as a reference alternative for model estimation. Therefore, the parameters of the MNL model given in Tab. I express the propensity of households to own two or three and more cars as against owning one car.

As can be seen from Tab. I, household income $I$ has a positive influence on both car ownership categories. It means that keeping all other variables constant, a unit increase in household income increases the probability of a household to own two and more cars as opposed to one car. Although that this confirms the general
expectation that wealthier households have higher disposable income for improving their travel comfort, note that the decision to invest the income in buying a car does not depend only on net income, but also on other household expenses such as a mortgage or rent. However, information on necessary household expenses was not included in the survey. Note also that the purchase of a car itself may not be decisive, because a lot of used cars for prices that are relatively favourable even for lower-income households is available on the market. However, each household must consider the additional costs associated with the operation and maintenance of the purchased car, which may be affected by the necessary expenses mentioned above.

Positive estimates of coefficients corresponding to the variable \( DL \) imply that the probability of households to own two and more cars increases with an increasing number of members owning a driving license. This suggests that with an increasing number of members owning a driving license, the motivation to own a larger number of cars increases as well.

The negative estimates of parameters corresponding to the dummy variable \( RP \), indicating that the households with two retired pensioners are less likely to own two and more cars as against one car. This may be due to the fact that the presence of retired pensioners in a household impacts disposable income. Specifically, with a higher number of retired pensioners, the budget to spend on transportation is smaller. Besides that, retired pensioners are supposed to have less need for travelling.

Both variables \( CC \) and \( EA \) significantly influence only one car ownership alternative. Based on the corresponding estimates we can conclude that a household with a possibility to use a company car is less likely to own two cars as against one car, while the alternative of three and more cars is not significantly influenced by the access to the company car. This result suggests that households with the possibility to use a business car have less need to invest their own income into a purchase of an additional car. The results in Tab. I also show that with the increasing number of economically active person household is less likely to own three and more cars as against to one car.

Although the households with a higher number of economically active persons have higher income, this unexpected result suggests that the number of economically active household members does not increase the household’s need of the third car. This finding can also result from the fact that not every economically active person owns a driving licence. E.g. from the total number of five households with four economically active persons, there are two households where only two members own a driving licence.

Further, as can be seen from Fig. 1c), the proportion of households with three and four economically active persons is relatively small, and only one household with three and more economically active persons owns three and more cars. It means that the results can be influenced by a relatively small sample of data corresponding to households owning three and more cars and three and more economically active persons.

CONCLUSION

The paper presents the results of the MNL model describing the relationship between the number of cars owned by the household and household characteristics in the selected regions of the Czech Republic. The data for the analysis were collected from traffic behaviour surveys carried out in two Czech cities Litoměřice and Písek as part of the development of sustainable urban mobility plans. In total, three car ownership categories were specified: one car, two cars, three and more cars. The results confirmed that household income and the number of household members with a driving licence has a positive impact on owning two and more cars. On the other hand, the presence of two retired pensioners in a household decreases the probability of a household to own two and more cars as against one car. It was also found out that the households with a possibility to use a company car are less likely to own two cars as against one car, and the households with an increasing number of economically active persons are less likely to own three and more cars as against one car.

We are aware that the presented model is limited by the absence of several household characteristics that were found to influence household car ownership based on previously published findings and a relatively small sample of data corresponding to households owning three and more cars. It is recommended that in future research other characteristics such as the time needed for parking or age of the head of the household are included in the analysis to identify more complex relationships between residential locations and car ownership. However, we believe that the presented research will be useful for forecasting transport demand in the area of interest.

Acknowledgements

This article was created with the financial support of the Ministry of Transport within the programme of long term conceptual development of research institutions on the research infrastructure acquired from the Operational Programme Research and Development for Innovation (CZ.1.05/2.1.00/03.0064).
REFERENCES


Contact information
Martina Čampulová: martina.campulova@mendelu.cz
Roman Čampula: roman.campula@cdv.cz

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0) International License.