# Estimation of modal split parameters – a case study

Luka Novačko[1*], Karlo Babojelić[1], Nikola God[2], Lidija Babić[1]

Faculty of Transport and Traffic Sciences, Zagreb, Croatia[1]
Growth Strategies Ltd., Zagreb, Croatia[2]
E-mail: luka.novacko@fpz.unizg.hr, karlo.babojelic@fpz.unizg.hr, god.nikola@gmail.com, lidija.babicka@gmail.com

***Abstract:*** *Choosing a mode or type of transportation is certainly one of the most important steps in a classic transportation model, given that it has a significant impact on the traffic planning and transport policy of a city. The choice of transport mode (e.g. passenger car, public transport, bicycle, walking) depends on the availability of transport modes and the generalized cost of transportation by transport mode from origin to destination. The choice of mode of transportation is also significantly influenced by the trip purpose. The most commonly used model for determining the modal split is the logit model. Parameters of utiliy function for each mode were estimated based on a household travel survey in Slavonski Brod, Croatia using the Biogeme program. The generalized cost and impedance towards using each transport mode is expressed by the utility function. This paper presents a methodology for determining the utility function for each transport mode using the Biogeme program.*

***KEYWORDS****: MODE CHOICE, LOGIT MODEL, UTILITY FUNCTION, BIOGEME*

## *1. Introduction*

The mode choice of transport is the third step in a four-step modeling process and is one of the most important steps. It has a significant impact on the traffic planning and traffic policy of the city. The transport mode choice (eg car, public transport, bicycle, walking) depends on the availability of transport modes (especially passenger car) and the travel cost for each mode from origin to destination. Factors influencing the mode choice can be divided into three groups of characteristics (1) passenger characteristics (car ownership; possession of a driver's license; household structure (youth, couple with children, pensioners, singles, etc.); income; (2) characteristics of travel: trip purpose - trip to work is usually easier to do by public transport due to its regularity, part of the day when traveling - eg night travel is more difficult to achieve by public transport, whether traveling alone or with other passengers (3) features of the transport system: travel time components: in-vehicle time, waiting time and walking time to each transport mode, monetary cost components (transport ticket, tolls, fuel costs and other operating costs), availability and parking costs, reliability of travel time and regularity of service; comfort and convenience; driving safety and protection while driving; driving skills requirements; opportunities to carry out other activities while driving (telephoning, reading, etc.) [1].

The basic problem of discrete mode choice analysis is the modeling of choices between a set of different alternatives. Utility maximization is taken as the solution. The decision maker selects the alternatives with the greatest utility over the time period. The operational model consists of a parameterized utility function consisting of independent variables and unknown parameters estimated from the sample. The probability of selecting an alternative is defined as the probability that has the greatest utility among all possible alternative solutions [2].

The choice of alternative can be observed because of a multi-stage decision-making process:

1. definition of the choice problem,
2. generation of alternatives,
3. evaluation of attributes of the alternatives,
4. choice of alternative,
5. implementation.

Choice theory is a set of procedures consisting of the following elements: (i) decision maker, (ii) alternatives, (iii) attributes of alternatives, (iv) decision rule. Not every choice behavior of individuals is described by such a decision-making process. An individual may choose a mode of transportation according to a habit, intuition, or imitation of someone else who is considered as an expert or leader. This behavior is presented as a decision-making process in which the decision maker generates only one alternative.

The modal distribution is determined using discrete choice models, ie Logit models. The most widely used discrete choice models are multinomial (MNL) and nested (NL) logit model. Multinomial logit models imply a larger set of alternatives in the final data set (eg alternatives can be car - driver, carpooling, taxi, bicycle, walking, bus, tram, train), while nested models consist of a group of similar alternatives grouped in a nest, in which each alternative belongs to only one nest. A multinomial logit model was used in this study:

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

$P_{ni}(i)$ = probability of choosing an alternative *i*,

$\mu$ = calibration parameter,

$V_{in}$= utility of mode i.

The aim of this paper is to estimate the parameters of the utility function for each mode of transport using the Biogeme program based on a household survey. Generalized cost and resistance to a mode of transport are expressed using the utility function. This paper presents a methodology for determining the utility function for each mode of transport using the Biogeme program.

## *2. Modal split in Croatia by National Transport Model*

The National Transport Model for the Republic of Croatia was completed in June 2016, and the model represents the second phase of the development of the Transport Development Strategy of the Republic of Croatia 2017. – 2030. [3]. The reference year for the analysis of the current state of the transport sector in the National Transport Model is 2013, since it is the only year for which all data were available. As part of the research for the needs of developing the National Transport Model in the Republic of Croatia, a travel behavior survey was conducted.

Figure 1. presents the proportion of all trips taken by different modes of transport. The most used transport mode was car with about 51% of all trips (40.8% as a driver and 10.4% as a passenger). Walking was the second most used travelling mode with the proportion of 30% of all trips. The bus was the most often used public transport mode with a share of 7.1%, while overall around 12% of all trips were made by public transport (bus, tram, train, and ferry). About 5% of all trips were made by bicycle [3].

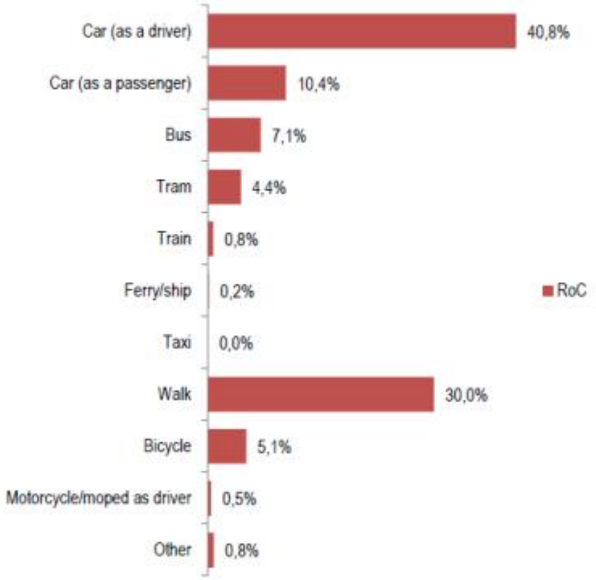


***Figure 1.*** *Proportion of all trips taken by different mode of transport at the national level (Source: [4])*

According to National Transport Model Croatia is divided into two NUTS regions: Continental Croatia and Adriatic Croatia. Figure 2 shows a map of NUTS-2 regions in Croatia.

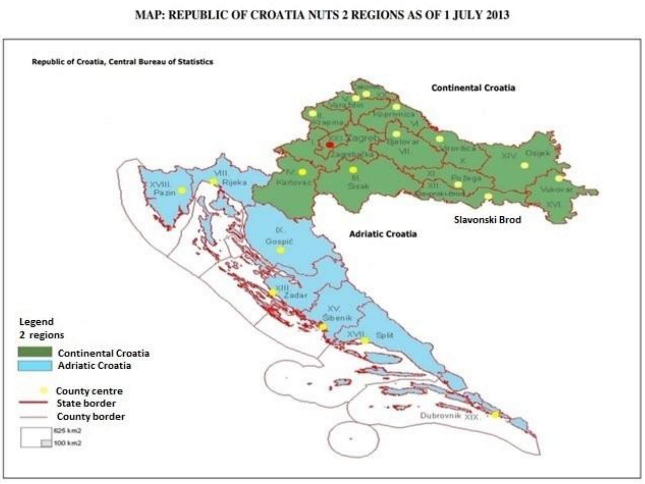


***Figure 2.*** *Map of Republic of Croatia NUTS-2 regions*

The proportion of all trips taken by different transport modes at the regional level reveals a significant difference between Continental and Adriatic Croatia. The ranking order of the mode shares was the same for both regions, though citizens from Adriatic Croatia made more trips by car and by walking while citizens from Continental Croatia used public transport and bicycle more often (Figure 3.).

Other results indicate that about 5% of all trips were made by bicycle. Compared to Continental Croatia, citizens from Adriatic Croatia made almost 40% more car trips, 60% more walking trips, 32% fewer public transport trips, and 65% fewer bicycle trips.

## *3. Case study – Slavonski Brod, Croatia*

The household survey was conducted as part of the development of the Sustainable Urban Mobility Plan of the Urban Area of the city of Slavonski Brod (Croatia). Slavonski Brod is in Brod-Posavina County, which is part of Continental Croatia according to the NUTS-2 division. The survey was conducted on a random sample of households, by direct interview, on representative days, in October and November 2019. A total of 5% of the total number of households in the study area were surveyed. The survey was conducted in the household where trained interviewers asked questions related to general household data and data on all trips of each household member that occurred the previous day (origin and destination of trips, travel time, trip purpose, mode choice). A total of 752 household surveys were collected in 36 residential zones.

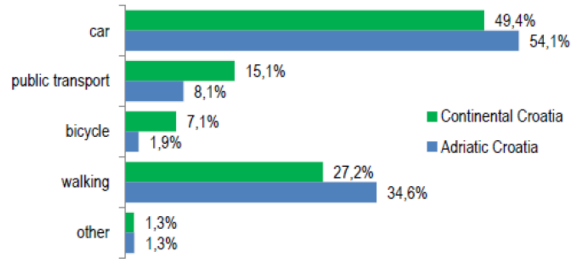


***Figure 3.*** *Proportion of all trips taken by different modes of transport, at the regional (NUTS-2) level (Source:[4])*

The parameters of the utility function for each transport mode were determined by the PandasBiogeme program code. Four types of transport modes were considered: car, bicycle, walking and public transport. The package Biogeme [5] is designed to estimate the parameters of various models using maximum likelihood estimation (MLE). It is particularly designed for discrete choice models. Biogeme is available in three versions.

- BisonBiogeme is designed to estimate the parameters of a list of predetermined discrete choice models such as logit, binary probit, nested logit, cross-nested logit, multivariate extreme value models, discrete and continuous mixtures of multivariate extreme value models, models with nonlinear utility functions, models designed for panel data, and heteroscedastic models. It is based on a formal and simple language for model specification.
- PythonBiogeme is designed for general purpose parametric models. The specification of the model and of the likelihood function is based on an extension of the Python programming language. A series of discrete choice models are precoded for an easy use. The package is written in C++ and is standalone.
- PandasBiogeme is a Python package, that must be imported in a Python code. It relies on the Pandas package for the data manipulation. This is the standard mode of operations of more and more data scientists. The syntax for model specification is almost the same as PythonBiogeme.

The data table from household survey must contain information about trips, travel time for each transport mode and the respondents choice of the transport mode. Each column corresponds to a specific variable, and the row corresponds to a value.

***Table 1. Data table for Biogeme***

| Trip_ID | TT_CAR | TT_BIKE | TT_PUT | TT_PED | CHOICE |
|---|---|---|---|---|---|
| 1 | 11 | 24 | 68 | 106 | 1 |
| 2 | 8 | 16 | 40 | 61 | 1 |
| 3 | 8 | 4 | 29 | 61 | 1 |
| 4 | 8 | 16 | 28 | 58 | 3 |
| 5 | 8 | 16 | 20 | 59 | 3 |
| 6 | 3 | 94 | 19 | 2 | 4 |
| 7 | 3 | 18 | 20 | 2 | 4 |
| 8 | 7 | 82 | 31 | 69 | 1 |
| 9 | 7 | 17 | 37 | 68 | 1 |
| 10 | 6 | 14 | 32 | 55 | 1 |

The variables defined in the program code were:

- Trip_ID = trip ID,
- TT_CAR = travel time by car [min],
- TT_BIKE = travel time by bicycle [min],
- TT_PUT = travel time by public transport - bus [min],
- TT_PED = travel time by walking [min],
- CHOICE = choice of transport mode.

The parameters that need to be evaluated to define the logit function are:

- ASC_CAR (alternative specific constant_car),
- ASC_BIKE (alternative specific constant_bike),
- ASC_PUT (alternative specific constant_put),
- ASC_PED (alternative specific constant_ped).

The utility function is defined for each transport mode where the travel time of each transport means is considered:

- V1 = ASC_CAR + B_TIME * TT_CAR_SCALED,
- V2 = ASC_BIKE + B_TIME * TT_BIKE_SCALED,
- V3 = ASC_ PUT + B_TIME * TT_ PUT_SCALED,
- V4 = ASC_ PED + B_TIME * TT_ PED_SCALED,

Each utility function must be associated with the number or identifier of each alternative using numbering as in the data table (table 1).

V = {1: V1, 2: V2, 3: V3, 4: V4}

After defining the utility function, it is necessary to assign each utility function to a specific transport mode using numbering as in the data table. The following numbering was used:

- CAR = 1,
- BIKE (cycling) = 2,
- PUT (public transport) = 3,
- PED (pedestrian/walking) = 4.

av = {1: CAR_AV_SP, 2: BIKE_AV, 3: PUT_AV_SP, 4: PED_AV_SP}

Next, the choice model is defined. The function bioLogLogit provides the logarithm of the choice probability of the logit model. It takes three arguments:

1. the dictionary describing the utility functions (V),

2. the dictionary describing the availability conditions (av),

3. the alternative for which the probability must be calculated (CHOICE).

logprob = bioLogLogit (V, av, CHOICE)

The results of the estimated parameters for each mode of transport are shown in Table 2. It is concluded that the p-value of the analyzed parameters is less than 0.05, which means that the results are statistically significant.

***Table 2. Estimation parameters from Biogeme***

| Name | Value | Std err | t-test | p-value | Rob. Std err | Rob. t-test | Rob. p-value |
|---|---|---|---|---|---|---|---|
| ASC_BIKE | -0,29 | 0,04 | -6,74 | 1.55e-11 | 0,04 | -6,81 | 9.82e-12 |
| ASC_CAR | 1,41 | 0,03 | 43,80 | 0,00 | 0,03 | 45,50 | 0,00 |
| ASC_PED | 0,65 | 0,04 | 18,40 | 0,00 | 0,03 | 18,70 | 0,00 |
| ASC_PUT | -1,77 | 0,08 | -23,20 | 0,00 | 0,07 | -24,20 | 0,00 |
| B_TIME | -5.06e-05 | 2.33e-05 | -2,18 | 0,0295 | 2.63e-05 | -1,92 | 0,05 |

The values of the utility function parameters obtained by the computer program PandasBiogeme were used for development of a multimodal transport model of the urban area of Slavonski Brod. The multimodal transport model was developed in the computer program PTV VISUM [6].

Data validation was made by comparing the data obtained by the multimodal transport model and the data collected by the household survey. Figure 4 shows a comparison of the modal distribution obtained by the household travel survey in Slavonski Brod, the multimodal transport model of Slavonski Brod and National Transport Model (Continental Croatia (NUTS-2 region)).

It is concluded that the difference of the parameters obtained by Slavonski Brod model from the household survey for transport modes is as follows: for public transport (PuT) the difference is 0.2%, for bicycle (BIKE) 0.3%, for walking (PED) 0.8% and for car (CAR) 0.1%.

Comparing the results from the National Transport Model (Continental Croatia) and the results from the Slavonski Brod model, the modal distribution for public transport (PuT) differs by 11.7%, for bicycles (BIKE) 3.8%, walking (PED) 0.3% and cars (CAR)
9.4%.

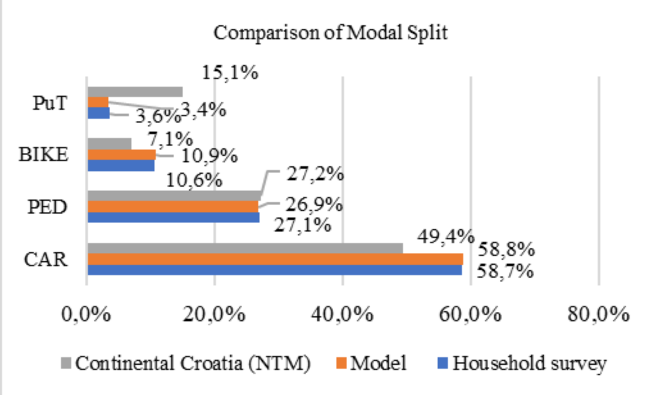


***Figure 4. Comparison of Modal Split of Household Survey, Model and NTM***

## 4. Conclusion

The modal distribution is the third step in the classic four-step model. The choice of transport mode is influenced by certain factors that can be divided into passenger characteristics, travel characteristics and transport system characteristics.

The application of discrete choice models determines the modal distribution. The most widely used discrete choice models are the multinomial logit model (MNL) and the Nested logit (NL) model.

The aim of this paper was to estimate the parameters of the utility functions for each observed transport mode using the PandasBiogeme program. The estimation of the parameters is based on a household survey conducted in the city of Slavonski Brod (Croatia). Slavonski Brod is part of Continental Croatia according to the NUTS-2 division.

The estimated utility functions parameters were implemented in the multimodal transport model. The data were analyzed and compared with the conducted household survey in Slavonski Brod and the modal distribution obtained by the National Transport Model.

The difference between the estimated modal distribution of the household survey and the multimodal transport model is less than 1% for all modes of transport. While the difference between the estimated parameters of the Slavonski Brod model and the National Transport Model is a bit larger, but not significantly. However, it must be considered that in the National Transport Model all cities of Republic of Croatia were analyzed in 2013.

## 5. References


1. J. de D. Ortuzar, L. G. Willumsen. *Modelling Transport* ($4^{th}$ Edition, Wiley, 2011)
2. M. Ben Akiva, S. Lerman.: *Discrete Choice Analysis – Theory and Application to Travel Demand* (MIT Press, Cambridge, 1985)
3. Transport Development Strategy of the Republic of Croatia (2017 - 2030), Ministry of the Sea, Transport and Infrastructure
4. National Transport Model for the Republic of Croatia, 2016.
5. M. Bierlaire, PandasBiogeme: a short introduction. Technical report TRANSP-OR 181219. Transport and Mobility Laboratory, ENAC, EPFL, (2018)
6. PTV Visum Manual, 2020.