The Economically Active People in the Transport Process

Marek Drliciak 1, Jan Celko 1, Michal Cingel 1

1 University of Žilina, Faculty of civil engineering, Department of highway engineering, Univerzitná 8215/1, 010 26 Žilina

marek.drliciak@fstav.uniza.sk

Abstract. Traffic-planning process, the traffic management process requires continuous collection and analysis of traffic data. The direct measurement usually determines the volume or speed of the traffic flows. In comprehensive traffic, assessment is necessary to know the causes of mobility and its characteristics, the group of habitants and purpose of the trips, above all. The article deals with the mobility characteristics of economically active people for work and their application to calculations using discrete functions. The data are based on an extensive transport-sociological survey of the mobility of the inhabitants of the Žilina region. The first part of the article is devoted to the basic survey data. The core of the article is the application of discrete choice models in the process of trips generation and modal split.

1. Introduction

The traffic forecast is based on a combination of multiple factors. The socio data contain the first group of factors (population, age structure or economic activity). As a rule, this data culminates in trend lines based on territorial analysis.

The second essential component is the traffic habits of the population. At present, great emphasis is placed on the value of travel time. The travel time affects the mobility, length of trips, modal split, as well as a choice of route. Several factors affect travel time. The moods of the population are mutable and oftentimes it is difficult to understand the causes and establish connections.

For this reason, it is necessary to approach the calculation of transport relations by using procedures that allow multiple factors in the decision-making process. The discrete choice models are presented as a development and a renovation of the classical theory of choice.

2. Discrete choice model

Discrete choice models statistically relate the choice made by each person to the personal attributes and the attributes of the alternatives available to the person. For example, the choice of which car a person buys is statistically related to the person’s income and age as well as to price, fuel efficiency, size, and other attributes of each available car. The models estimate the probability that a person chooses a particular alternative. The models are often used to forecast how people’s choices will change under changes in demographics and/or attributes of the alternatives [8].
Discrete choice models take many forms, including Binary Logit, Binary Probit, Multinomial Logit, Conditional Logit, Multinomial Probit, Nested Logit, Generalized Extreme Value Models, Mixed Logit, and Exploded Logit [13].

Presented analysis uses a discrete choice model to describe the transport habits of the economically active people. In the first case, the mobility (number of “JOB” trips per day) of the economically active population per household. In the second step, we used a discrete model to calculate the division of transport work.

The most common theoretical basis of discrete choice models is the random utility model. The selection of the utility function form is an important decision. Any cumulative distribution function can be used. Usually, F is a standardized normal cumulative distribution function (the model is called probit in this case) or a cumulative logistic distribution function (called logit in this instance). There are no exact practical rules for this selection (some recommendations can be found at [1]).

The utility is modelled as a random variable in order to reflect this uncertainty [5]. More specifically, the utility that individual n associates with alternative i in the choice set Cn is given by

\[ U_{in} = V_{in} + \varepsilon_{in} \]  

(1)

where \( V_{in} \) is the deterministic (or systematic) part of the utility, and \( \varepsilon_{in} \) is the random term, capturing the uncertainty [2]. The alternative with the highest utility is chosen. Therefore, the probability that alternative i is chosen by decision-maker n from the choice set \( C_n \) is

\[ P(i|C_n) = P[U_{in} \geq U_{jn} \forall j \in C_n] = P[U_{in} = \max_{j \in C_n} U_{jn}] \]  

(2)

Usually, the maximum likelihood estimator is used for estimating the coefficients of unknown discrete choice models. This provides asymptotically efficient and consistent estimates [2].

Multinomial logit model The Logistic Probability Unit, or the Logit Model, was first introduced in the context of binary choice where the logistic distribution is used. Its generalization to more than two alternatives is referred to as the Multinomial Logit Model. The Multinomial Logit (MNL) Model is derived from the assumption that the error terms of the utility functions are independent and identically Gumbel distributed [5].

The probability that a given individual n chooses alternative i within the choice set \( C_n \) is given by

\[ P(i/C) = \frac{e^{V(x,s)}}{\sum_{j \in C} e^{V(x,j,s)}} \]  

(3)

Where:
- \( C \) – finite choice set,
- \( P(i/C) \) – choice probability for alternative \( i \in C \),
- \( x \) – vector of observed characteristics of alternative \( i \),
- \( s \) – vector of observed characteristics of the decision maker and the choice environment.

The scale function \( V_{(i,s)} \) may be represented as the representative utility of alternative i and is normally assumed to be linear in the parameters. The MNL model has significant advantages over the available alternatives in terms of flexibility and computational efficiency and permits a simple behavioural interpretation of the parameters of the scale function [4].

Maximum likelihood estimation is widely used in estimating the parameters of a statistical model. The probability of a decision maker n choosing an alternative that s/he actually chooses is \( \prod_{i} (P_{ni})^{y_{ni}} \), in which \( y_{ni} = 1 \) if n choose i and zero otherwise.

Assume that every decision maker choose alternative independently, the probability of all decision maker choosing their actual choices is
in which $\beta$ is a vector of the parameters to be estimated. The log-likelihood function is

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{ni} \ln P_{ni}$$

the estimated result is then the $\beta$ that maximizes this function, that is the derivative with respect to $\beta$ is zero [4].

$$\frac{dLL(\beta)}{d\beta} = 0$$

The parameter $\beta$ is possible to determinate by more tools. The Biogeme software was used for estimation of the parameters $\beta$ in the research of the authors. Biogeme is an open source freeware designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models. [6] The Biogeme package (biogeme.epfl.ch) is designed to estimate the parameters of various models using maximum likelihood estimation. It is specially designed for discrete choice models.

The used traffic-sociological survey (mobility survey) was conducted in the Žilina self-region. Žilina region with the area of 6 801 km² lies [10] on the north-western and northern part of Slovakia, it has borders with the Czech Republic and with Poland (Figure 1).

![Figure 1. The Žilina self-governing region](image)

The actual data from the Slovak Statistical Office was used for specifying the credible sample. The survey was then performed in every village where the required number of households is greater than 10. Finally was asked 6231 households with 18,382 inhabitants. The survey was conducted through the questionnaires. The questions were formulated to describe the daily activity, transport habits and the causes of mobility. Every interviewee was assigned to a predetermined population group.

The database contains 33 688 trips descriptions. Almost 50% of them contain the activity “JOB” (attraction or production). The economically active people share almost 67% of total inhabitants. The transport mode “car driver” was selected for 42% of work trips.

We focused on the economically active people and their job trips. The discrete choice model is used in the mode choice process, rarely in trip generation calculation. The new formulas of the utility functions were defined based on source data. The application of discrete choice into the trip generation and mode choice process is described in the next part of the article.

3. Trip generation

The total number of trips in the modelled area is calculated as a sum of the person groups mobility. Trip generation for work is one of the main purposes. [11] There are no exact rules for the trip generation process. The model maker could define the specific demand segments such as for example the
economically active people and their job trips. The level of detailed is determining. The high level is linked to a lack of data.

We used a discrete choice model to calculate the number of trips (Home – Job) per household. The number of economically active people (EAP), children (NCh) per household, the type of residence (TR), the education (PS, UNI) and car ownership (NCar) were used in utility functions formulas. The described inputs data defined following utility functions \( V_i^s \):

\[
V_0 = ASC_{0c} + 0 \ast EAP
\]

\[
V_1 = ASC_{1c} + \beta_{1c,EAP} \ast EAP + \beta_{1c,NCh} \ast NCh + \beta_{1c,TR} \ast TR
\]

\[
V_2 = ASC_{2c} + \beta_{2c,EAP} \ast EAP + \beta_{2c,NCh} \ast NCh + \beta_{2c,NCar} \ast NCar + \beta_{2c,TR} \ast TR
\]

\[
V_3 = ASC_{3c} + \beta_{3c,EAP} \ast EAP + \beta_{3c,NCh} \ast NCh + \beta_{3c,NCar} \ast NCar + \beta_{3c,PS} \ast MS + \beta_{3c,UNI} \ast Uni + \beta_{3c,TR} \ast TR
\]

\[\text{Table 1. The result from BIOGEME model – trip generation}\]

| Nbr of parameters | 18 |
|-------------------|----|
| Sample size       | 6118 |
| Excluded data     | 0 |
| Init log likelihood | -8481.35 |
| Final log likelihood | -5530.24 |
| Likelihood ratio test | 5902.224 |
| Rho square        | 0.348 |
| Rho bar square    | 0.346 |
| ASC_0c            | 481089 |
| B1c_byt           | 0.010138 |
| B3c_byt           | 0.033484 |
| ASC_1c            | 3376277 |
| B1c_deti          | 0.010224 |
| B3c_deti          | 0.021559 |
| ASC_2c            | -0.79721 |
| B1c_zam           | -1.25761 |
| B3c_vs            | 0.00112 |
| ASC_3c            | -7.38996 |
| B2c_byt           | 0.020895 |
| B3c_voz           | -0.00111 |
| B0c_zam           | -2.46912 |
| B2c_deti          | 0.019165 |
| B3c_zs            | 0.000282 |
| B2c_voz           | -0.00438 |
| B3c_zam           | 2.817506 |
| B2c_zam           | 0.90923 |

The constant factors of utility functions were evaluated by BIOGEME. 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 easy use.
The application of the discrete choice model shows the possible way to calculate the mobility rate of inhabitants. The results are dependent on the expression of the utility function. The formulas for trip generation could be used in new invest evaluation.

4. Modal split
The mode choice is the process where the means of travelling is determined. The means of travel is referred to the travel mode, which may be by private automobile, public transport, walking, bicycling, or other means. Mode choice is formulated as a discrete choice model with alternatives corresponding to the specific tour or trip modes.[12]

GORR (1997) defines mode choice by assuming individual preferences; i.e. the indifference curves (all modes on one and the same curve are preferred equally) differ between different homogeneous groups. The sum of all characteristics results in a specific attraction of each mode and is crucial for choosing a mode together with individual preferences. Three-dimensional figure 3 demonstrates this context [7] (Figure 3).

![Figure 3. Individual preferences for transport mode choice [7]](image)

The utility for an alternative would consist of a systematic attribute which is a function of relevance to the decision-making process and a constant representing the uncertainty derived from individual behaviour and modeller measurement errors.

\[
V_{foot} = ASC_{foot} + TT_{foot} \cdot \beta_{TT,foot} + DIS_{foot} \cdot \beta_{DIS,foot} + COST_{foot} \cdot \beta_{Cost,foot} \\
V_{bike} = ASC_{bike} + TT_{bike} \cdot \beta_{TT,bike} + DIS_{bike} \cdot \beta_{DIS,bike} + COST_{bike} \cdot \beta_{Cost,bike} \\
V_{mbike} = ASC_{mbike} + TT_{mbike} \cdot \beta_{TT,mbike} + DIS_{mbike} \cdot \beta_{DIS,mbike} + COST_{mbike} \cdot \beta_{Cost,mbike} \\
V_{car,d} = ASC_{car,d} + TT_{car,d} \cdot \beta_{TT,car,d} + DIS_{car,d} \cdot \beta_{DIS,car,d} + COST_{car,d} \cdot \beta_{Cost,car,d} \\
V_{car,p} = ASC_{car,p} + TT_{car,p} \cdot \beta_{TT,car,p} + DIS_{car,p} \cdot \beta_{DIS,car,p} + COST_{car,p} \cdot \beta_{Cost,car,p} \\
V_{PUT} = ASC_{PUT} + TT_{PUT} \cdot \beta_{TT,PUT} + DIS_{PUT} \cdot \beta_{DIS,PUT} + COST_{PUT} \cdot \beta_{Cost,PUT}
\]

The utility functions were defined for the choice set (foot, bike, motorcycle, car-driver, car-passenger and public transport). The factors affecting mode choice were selected distance (DIS), travel time (TT) and cost (COST). The individual attributes of mode (for example travel time) were calculated in the regional transport model as skim matrices.

The parameters $\beta$ were estimated in BIOGEME. The values are shown in Table 2.
Table 2. The result from BIOGEME model – modal split

| Nbr of parameters | 9 | Districts towns | Other towns and villages |
|-------------------|---|-----------------|-------------------------|
| Sample size       | 659 | ASC_B | -0.0888 | -0.979 |
| Excluded data     | 354 | ASC_M | -3.05 | -2.51 |
| Init log likelihood | -1180.77 | ASC_OA | 1.7 | 1.76 |
| Final log likelihood | -688.563 | ASC_Oasp | 0.57 | 0.527 |
| Likelihood ratio test | 984.132 | ASC_P | -0.426 | 0.52 |
| Rho square        | 0.417 | ASC_PUT | 1.29 | 0.682 |
| Rho bar square    | 0.409 | B_COST | 10.4 | -37.6 |
|                   |       | B_DIS | 1.26E-11 | 2.44E-10 |
|                   |       | B_TIME | -2.19 | -4.79 |

The probability that a choice maker chooses alternative i within the choice set of the transport modes is given by (3). The modal split was calculated using equations of utility with estimated coefficients i. The figure 4. presents the comparison of modelled an observed modal split for trips to work.

Figure 4. Comparisons of modelled and observed data – modal split

5. Conclusion
The most challenging task in transportation forecasting process is to identify the influencing factors on a traveller’s choice. Economically active people have the largest share in the transport process. The trip chains with containing activity “JOB” have almost 50% rate.

This paper estimates models for the trip generation and mode choice on the basis of the travel data collected Žilina region. The models used a discrete choice. This procedure is the most used for mode choice process, not for the trip generation process. The results point to a possible application of statistic (freely available) data into the trip generation process. The specific inclusion could easily affect the calculation processes.

The number of calculated trips with purpose “Job” coincides with observed data. The comparison between modelled and observed modal split does not show complete agreement. The result points to the high sensitivity of the utility function parameters. Minimal changes in the assumption of personal choice
of transport mode represent a significant change of the final solution. Adequate attention must be paid to this in the next stage of research.

The input data for the modal split are exported from a parallel supply network model. The geometric simplification, the network parameters (speed, capacity) fundamentally affect data such as the travel time, distances. The use of discrete choice models with supply model data must be conditional by calibration and validation process.

Acknowledgement
The presented results were partially obtained under the grant Vega No. 1/0428/19, Implementation of structural variables in the calculation of the modal split in the transport forecast (Implementácia štrukturálnych veličín pri výpočte deľby prepravnej práce v prognóze dopravy)

References
[1] Pavlyuk, D., Gromule, G., Discrete choice model for a preferred transportation mode, The 10th International Conference “RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION - 2010”, p. 143-151, ISBN 978-9984-818-34-4,
[2] AMAYA LEAL, J., Modelling indifference option in stated preference survey, Graduate Project as a requisite for graduation.
[3] Ortúzar, J. de D. and Willumsen, L.G. (2001) Modelling Transport, 3rd ed. John Wiley & Sons, Chichester.
[4] D. McFadden, K. Train, William B. TZe, An application od diagnostic tests for the independence from irrelevant alternatives propeizt of the multinomial logit model,
[5] https://eml.berkeley.edu/reprints/mcfadden/3_7.pdf
[6] Ben-Akiva, M., Bierlaire, B., Discrete choice methods and their applications to short term travel decisions, Handbook of Transportation Science – 1999, Springer, Boston, MA , ISBN 978-1-4613-7370-4
[7] PandasBiogeme: a short introduction, Michal Bierlaire, December 19,2018 http://transpor.epfl.ch/documents/technicalReports/Bier18.pdf,
[8] GORR, H. (1997) Die Logik der individuellen Verkehrsmittelwahl. Theorie und Realität des Entscheidungsverhaltens im Personenverkehr. Gießen
[9] Train, K. (1986). Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand. MIT Press
[10] Statistical Office of the Slovak Republic (https://slovak.statistics.sk/)
[11] Abhishek L. Hedau & S. S. Sanghai, Development of trip generation model using activity based, International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development (IJCSEIERD) ISSN(P): 2249-6866; ISSN(E): 2249-7978 Vol. 4, Issue 3, Jun 2014, 61-78 © TJPRC Pvt. Ltd.
[12] Travel Forecasting Resource, http://tfresource.org/Category:Mode_choice
[13] Train, K., Discrete Choice Methods with Simulation. Published by Cambridge University Press, Second edition, 2009 , ISBN 978-0-521-76655-5 hardback, ISBN 978-0-521-74738-7 paperback