**TOMAŽ MAHER**, Ph.D.
E-mail: tomaz.maher@fgg.uni-lj.si
**IRENA STRNAD**, B.Eng.
E-mail: irena.strnad@fgg.uni-lj.si
**MARIJAN ŽURA**, Ph.D.
E-mail: mzura@fgg.uni-lj.si
University of Ljubljana,
Faculty of Civil an Geodetic Engineering
Traffic Technical Institute
Jamova cesta 2, SI-1000 Ljubljana, Slovenia

Transport Engineering
Preliminary Communication
Accepted: Apr. 7, 2010
Approved: May 17, 2011

# ESTIMATION OF EVA MODE CHOICE MODEL PARAMETERS WITH DIFFERENT TYPES OF UTILITY FUNCTIONS

***ABSTRACT***

*This paper presents the estimation of nine types of utility function parameters for the application in EVA mode choice model for the city of Ljubljana, Slovenia. Four different modes (private car, public transport, bike and walking) and five purposes (work, education, shopping, leisure and other) were taken into consideration. This paper presents first the design of the Stated Preference survey, then a brief review of the EVA model, different types of utility functions and the estimation method. The final log-likelihood enables comparison of different types of utility functions. The results show that absolute differences in final log-likelihood among most types of utility functions are not high despite the different shapes, which implies that different functions may best describe different variables.*

***KEY WORDS***

*mode choice, Stated Preference survey, utility functions, maximum likelihood method, generalized cost*

## 1. INTRODUCTION

An up-to-date disaggregated four-step traffic model for passenger transport in the Ljubljana region has already been developed. This model includes four modes – private car, public transport, bike, and walking. The purpose of this study is to upgrade the existing model with additional parameters that affect the mode choice, since this model contains only utility functions of travel time and not of all mode choice affecting factors.

The EVA (EVA - German abbreviation for *Erzeugung*, *Verteilung* and *Aufteilung* meaning *Production*, *Distribution*, and *Mode Choice*) algorithm has been adapted for this purpose as it allows a non-linear specification of the utility function to be contrasted with linear formulations of the variables in discrete choice models that have previously played the most important role in transport modelling.

Our first step was to undertake a Stated Preference survey in order to obtain the necessary data. Many instructions for conducting a Stated Preference survey are given by Ortúzar and Willumsen [1] and in Engineering Statistics handbook [3]. Several conference presentations i.e. those by Vrtic [4], Axhausen, Köll, Bader [5], Axhausen [6] and the article by Vrtic, Fröhlich, Schüssler, Axhausen, Lohse, Schiller, Teichert [7] were taken into consideration when designing the Stated Preference survey.

The Stated Preference survey was carried out with portable computers at different locations around Ljubljana. Different locations were needed to ensure a representative sample and the required sample size for investigation purposes. For each origin-destination purpose, 75 to 100 questionnaires per segment were needed, so the sample size has to be about 1,000 survey respondents as utility functions are to be estimated for ten origin-destination purposes. However, as mode choice does not usually change for trips back from the destination, five trip purposes were used (work, education, shopping, leisure and other). Thus, the sample size of 500 survey respondents was sufficient.

The survey included questions about the usage of different modes in different situations. The questionnaire consisted of ten hypothetical situations in which values of mode parameters changed. For each of those hypothetical situations, each survey respondent had to choose the most suitable mode for them provided they had already experienced such a situation.

The second step was to estimate and calibrate the utility functions for each generalized trip cost parameter by using maximum likelihood method. As some types of utility functions have linear and others have non-linear elasticities, differences in function shapes and log-likelihood were expected. The results show that absolute differences in final log-likelihood among most types of utility functions are not high despite different shapes, different functions would best describe different variables. To decide on the function that fits best, the usage of various types of utility functions for each generalized trip cost parameter would be necessary, which would implicate a rather high number of combinations.

## 2. STATED PREFERENCE (SP) SURVEY

The Stated Preference (SP) survey has become a widely used transport planning tool, in spite of its known limitations. SP data are convenient when an alternative as a whole is described as a constituent of different variables as the analysis of SP survey data derives the relative importance of different variables. Since SP derives the relative importance of variables, their nature is best described with the term generalized cost parameters.

In our Stated Preference survey the travellers were first asked about the employment status and car availability. Then questions about the trip they were making followed to learn the purpose, length, duration, costs and available travel alternatives. If there were other available and acceptable alternatives, similar information for those was collected as well. Each survey respondent was then asked to choose the most suitable mode in ten hypothetical situations with different values of trip attributes. The questionnaire design, the generation of the situations and the survey performance are briefly explained below.

### 2.1 Questionnaire Design

In our Stated Preference survey, four modes were taken into account: private car, public transport, bike and walking. The parameters included are:

- Car: travel time in minutes, walking time from parking to destination in minutes, parking price in euro.
- Public transport: travel time in minutes, comfort, price of public transport in euro, frequency in minutes between two successive arrivals and walking time from origin to start station and from final station to destination in minutes.
- Bike: cycling time in minutes.
- Walking: walking time from origin to destination in minutes.

The fractional factorial design was used to design the hypothetical situations needed. The fractional factorial designs are experimental designs consisting of a carefully chosen fraction of the experimental runs of a full factorial design. Fractional designs are expressed with the notation $l^{k-p}$, where $l$ is the number of levels of each factor investigated, $k$ is the number of factors investigated, and $p$ is the number of generators, i.e. assignments as to which effects or interactions are confounded (cannot be estimated independently of each other).

The study contained seven factors at three levels and one factor at two levels, as shown in *Table 1*.

The levels for car use vary up and down, while levels for public transport usage improve. The reason is the transport policy goal to enlarge the share of public transport users in comparison to private car users. This means that only improved changes in public transport are needed.

*Table 1* shows the parking price variability in percentage. Since few free parking lots in Ljubljana are available, the price would not change in situations, which would mean less realistic results of the parking price affecting the mode choice.

In case of free parking, a new parking price was set to generate situations. In case of trip purposes such as work and education, the price is €5 and for other purposes - €2. The prices were set on the basis of average parking price in Ljubljana with different trip

*Table 1 - Factors with their level*

| Mode&levels | car | levels | public transport | levels |
|---|---|---|---|---|
| factors | parking price | actual<br>+50%<br>-50% | price | actual<br>-25%<br>-50% |
| | travel time | actual<br>+20%<br>-20% | walking | actual<br>-20%<br>-40% |
| | walking | actual<br>+20%<br>-20% | frequency | actual<br>-20%<br>-40% |
| | | | travel time | actual<br>-25%<br>-50% |
| | | | comfort | actual<br>better |

*Table 2 - One factor at 2 levels and seven factors at 3 levels design*

| $L_{18}$ – 2 x $3^{7-5}$ Fractional Factorial (Mixed-Level) Design<br>One factor at 2 levels and seven factors at 3 levels (18 Runs) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Run | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 1 |

*Source: Engineering statistics handbook [3]*

purposes. Prices set as described change according to situations.

Building a fractional factorial design required generation of eighteen hypothetical situations (*Table 2*).

In each of those eighteen situations, factors were at a different level and the traveller had to choose the most suitable mode. Generating eighteen hypothetical situations was a huge practical barrier, since it meant a long questionnaire, which would have caused problems in the travellers' concentration and the possibility to make choices different from those they had ever made in situations experienced earlier. We decided to split the eighteen designed situations between two surveyed persons, each giving answers for nine situations only. Our choice included one control situation, in which generalized trip cost parameters were the same as given for the actual trip. If the traveller choice process complied with his actual mode choice, the situation was taken as a control situation.

The survey forms were made in Microsoft Access program. The form was designed to begin with entering data about the car ownership and the driving license, employment status, purpose, origin and destination of the trip, frequency of trip, start time of the trip and comments.

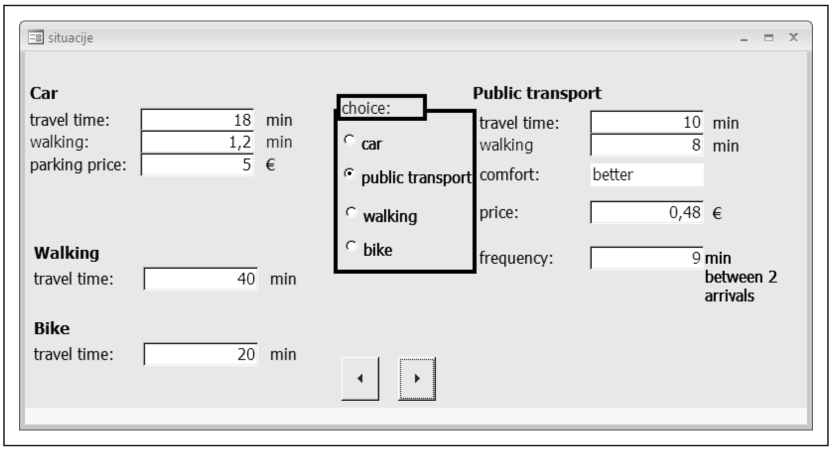


*Figure 1 - Example of the choice process*

The focus of the rest of the survey was on the actual trip surveyed and the available alternatives for the particular trip. On this basis the generation of ten hypothetical situations was made.

An example of the choice process in one situation is shown in *Figure 1*. The modes offered to be chosen are the same as those available for the traveller. If the person surveyed does not mark one alternative as available, choosing the same will not be possible and parameters of this alternative do not appear.

## 3. EVA TRIP DISTRIBUTION AND MODE CHOICE MODEL

Here, a brief review of EVA trip distribution and mode choice model are given, presenting the types of utility functions tested, the model for calculating mode choice probability, and the parameter estimation method as described in Visum 11.0 Basics [2].

Let $T_{ijk}$ be trips from zone i to zone j by mode k. Let $o_i$, $d_j$, $m_k$ be balancing factors used to keep marginal sums of productions, attractions, and mode trips.

Let $W_{ijk}$ be weighted utility of making trip from zone i to zone j by mode k.

Then EVA model generalizes simultaneous trip distribution and mode choice to trilinear model:

$$T_{ijk} = W_{ijk}.o_i.d_j.m_k \quad (i = 1,\ldots,n;\ j = 1,\ldots,n;\ k = 1,\ldots,p) \tag{1}$$

To keep marginal sums of production, attraction balancing factors $o_i$, $d_j$, $m_k$ must be determined in a way that the following constraints will be satisfied:

$$\sum_{j=1}^{n}\sum_{k=1}^{p} W_{ijk}.o_i.d_j.m_k = \sum_{j=1}^{n}\sum_{k=1}^{p} T_{ijk} = P_i \tag{2}$$

$$\sum_{i=1}^{n}\sum_{k=1}^{p} W_{ijk}.o_i.d_j.m_k = \sum_{i=1}^{n}\sum_{k=1}^{p} T_{ijk} = A_j \tag{3}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ijk}.o_i.d_j.m_k = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{ijk} = M_k \tag{4}$$

Weighted utilities $W_{ijk}$ are calculated as product of accessibility of mode k in zone i ($MA_{ik}$) and product of all utilities of making trip from zone i to zone j by mode k taking into consideration only one attribute of generalized cost $c_{aijk}$ (eq. time, parking cost, fare,...).

$$W_{ijk} = MA_{ik}\prod_{a} f_a(c_{aijk}) \tag{5}$$

### 3.1 Utility functions

The main task of the study was to estimate the type and parameters of utility functions to be used in trip distribution and modal choice model.

Let x be the generalized cost parameter and a, b, and c parameters of the utility function. Then $f(x)$ is utility function. The following types of utility functions, some of which with constant, some with linear, and others with non-linear elasticity, have been studied:

EVA1

$$f(x) = (1+x)^{-\varphi(x)},\ \text{where } \varphi(x) = \frac{a}{1+e^{b-cx}} \tag{6}$$

EVA2

$$f(x) = \left[1+\left(\frac{x}{c}\right)^{b}\right]^{-a} \tag{7}$$

Schiller

$$f(x) = \frac{1}{1+\left(\frac{x}{b}\right)^{a}} \tag{8}$$

Logit

$$f(x) = e^{cx} \tag{9}$$

Kirchhoff

$$f(x) = x^{c} \tag{10}$$

BoxCox

$$f(x) = e^{\left(c\frac{x^{b}-1}{b}\right)} \tag{11}$$

Box-Tukey

$$f(x) = e^{(cx^{\alpha})},\ \text{where } \alpha = \begin{cases}\dfrac{(x+1)^{b}}{b}, b > 0\\ \ln(x+1), b = 0\end{cases} \tag{12}$$

Combined

$$f(x) = ax^{b}e^{cx} \tag{13}$$

Code

$$f(x) = \frac{1}{x^{b}+cx^{a}} \tag{14}$$

Parameters a, b, and c of those nine types of utility functions were estimated in the study and evaluated according to their shape and log-likelihood, described below.

### 3.2 Model parameter estimation method

Probability that trips between zone i and j will be realized by mode k can be calculated from

$$P_{ijk} = \frac{W_{ijk}}{\sum_{A_l \in A(ij)} W_{ijl}} \tag{15}$$

where A(ij) is a set of available alternatives between zone i and j, and $W_{ijk}$ previously defined as weighted utilities.

Model parameters a, b, c have been estimated using the Maximum Likelihood method, described by Ortúzar and Willumsen [1]:

Let Q be a set of all situations conducted in experiment, A(q) - the alternatives available in situation q, and $A_j$ - the alternative chosen in situation q. Then we define

$$g_{jq} = \begin{cases}1\ \textit{if}\ A_j \text{ was chosen by } q\\ 0\ \textit{otherwise}\end{cases} \tag{16}$$

The likelihood function, which shows the model probability that each individual chooses the option they have already selected in an actual situation, is

$$L = \prod_{q=1}^{Q} \prod_{A_j \in A(q)} P_{jq}^{g_{jq}} \quad (17)$$

As it is more convenient to use the natural logarithm of L, model parameters can be estimated by solving the following non-linear program:

Find

$a_k, b_k, c_k$

where

$$l = \ln L = \sum_{q=1}^{Q} \sum_{A_j \in A(q)} g_{jq} \ln P_{jq} \quad (18)$$

has the maximum.

## 4. RESULTS

### 4.1 SP survey results

From 2,438 survey forms made, only 1,276 were used for calculating utility functions, as only those travellers with more alternatives and with willingness to use them were taken into account. Besides, data from situations were useful only if at least one of them referred to car or public transport since only parameter values for these two alternatives change in situations. Some basic information about survey performance is shown in the tables below.

*Table 3* shows that the number of surveys for each purpose was sufficient to enable a representative sample. The number of surveys made on the trips by bike (*Table 3* and *Table 4*) was low because of cold weather conditions.

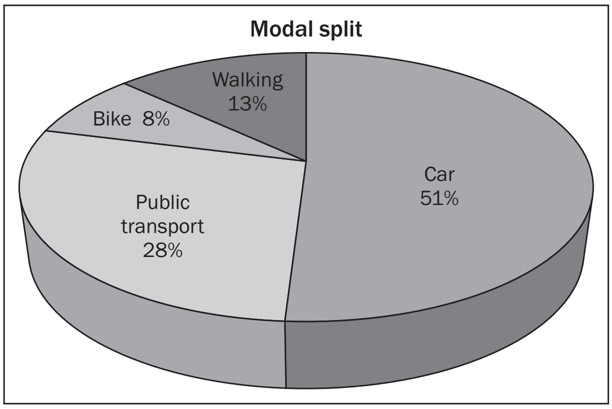


*Chart 1 - Modal split in the surveys*

Last column in *Table 4* and Chart 1 show the modal split in the surveys. This modal split cannot be taken as actual modal split for several reasons:

- The most important factor is the choice of locations, which directly affects the choice in the sample (e.g.: the more surveys made on trains, the greater share of public transport choice in the sample).
- Car users and cyclists are relatively unready to use any alternative.
- Relatively small sample to investigate modal split.
- Only trips in progress were measured.

### 4.2 Estimated utility function parameters

*Table 5* shows the estimated values of parameters a, b, c for different utility functions x.

*Table 3 - Number of surveys made for each mode and each purpose*

| Purpose | Mode | | | | total |
|---|---|---|---|---|---|
| | car | public transport | bike | walking | |
| work | 243 | 99 | 18 | 33 | 393 |
| education | 37 | 83 | 22 | 24 | 166 |
| shopping | 94 | 28 | 13 | 16 | 151 |
| leisure | 153 | 79 | 26 | 47 | 305 |
| other | 118 | 72 | 23 | 48 | 261 |
| total | 1,276 | | | | |

*Table 4 - Number of surveys made for each mode regarding car ownership*

| Mode | car ownership | | |
|---|---|---|---|
| | yes | no | total % |
| Private car | 598 | 47 | 51% |
| Public transport | 182 | 179 | 28% |
| Bike | 59 | 43 | 8% |
| Walking | 94 | 74 | 13% |
| Total % | 73% | 27% | 100% |

*Table 5 - Estimated parameters and final log-likelihood*

| | EVA1 | EVA2 | Schiller | Logit | Kirchhoff | Boxcox | Combined | Code | Box-Tukey |
|---|---|---|---|---|---|---|---|---|---|
| a | 1.4250 | 2420.5750 | 5.9872 | | | | 0.7202 | 3.3555 | |
| b | 4.7334 | 1.9820 | 4.4407 | | | 1.7276 | -0.0687 | 0.1063 | 0.5122 |
| c | 0.8336 | 216.7200 | | -0.3060 | -0.5429 | -0.0431 | -0.4536 | 0.0103 | -0.0555 |
| Final log-likelihood | -598.1250 | -595.8490 | -679.4600 | -601.3290 | -618.5900 | -590.7410 | -586.9050 | -626.3640 | -581.5790 |

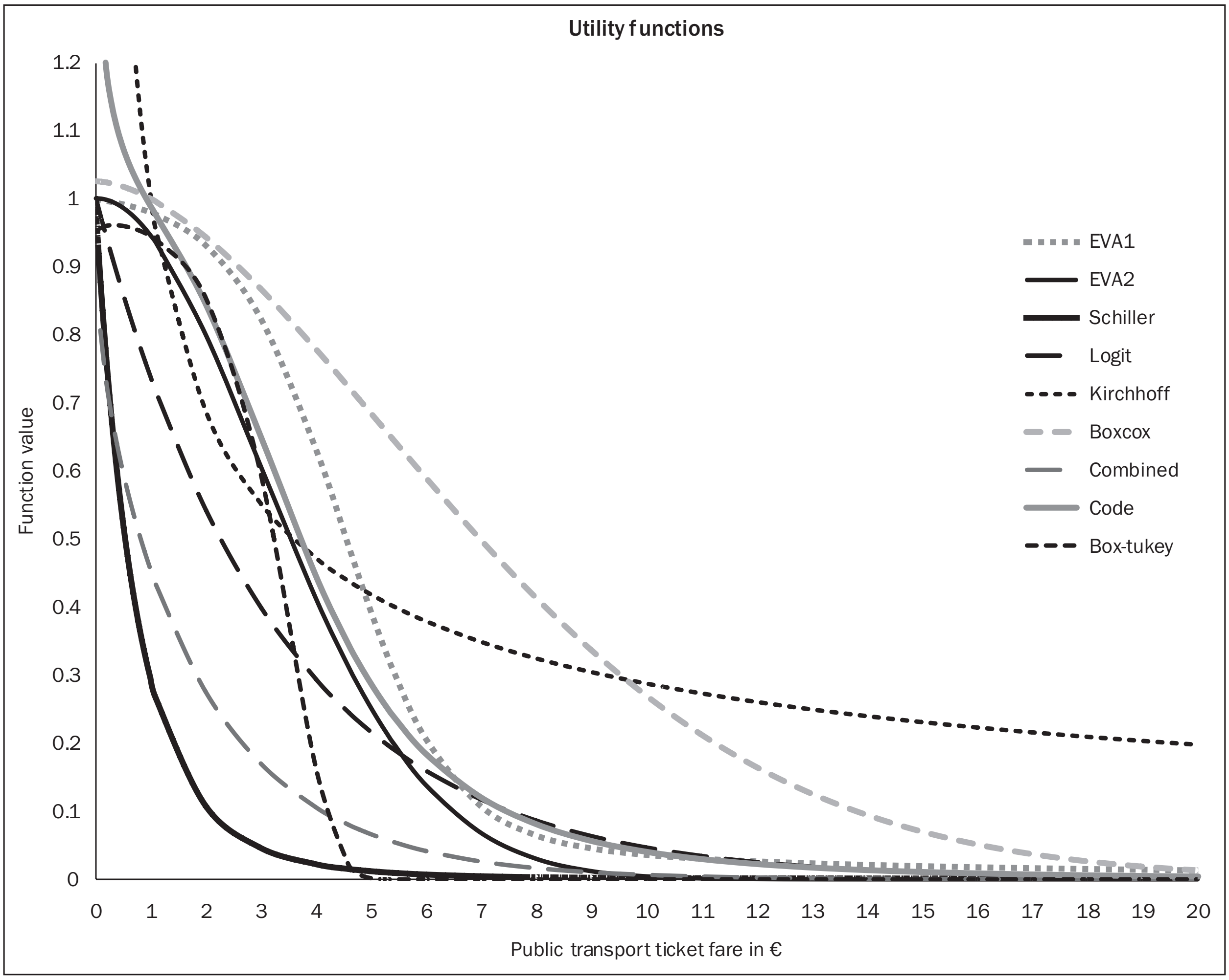


*Chart 2 - Utility functions*

Chart 2 displays only the utility functions for the trip purpose – work, and for one generalized trip cost parameter – the public transport fare.

## 5. DISCUSSION

One can observe that the model which gives maximum final log-likelihood is the model with Box-Tukey utility function, followed by Combined and others. Although final log-likelihood does not differ much from most utility functions, the graph shows different function shapes. While some of utility functions are monotonously falling and convex (Schiller, Logit, Combined, Kirchhoff), others show more believable results.

The shape of Kirchoff utility function appears as a surprise at first sight; its values are high (not falling to zero) even when public transport fares are high. However, since the probability of using each mode is the quotient between the weighted utility of that mode and the sum of weighted utilities of all available modes, the height of functions is not important.

Unlike the above, the lower values of final log-likelihood for three out of four convex functions (Schiller, Logit, Kirchhoff) are no surprise. The utility function Combined with high final log-likelihood is surprisingly convex, which may propose that Combined utility function does not fit best for this particular generalized cost parameter.

The final log-likelihood is the lowest when Schiller utility function is used. The graph shows that Schiller utility function is the lowest of all and therefore, the usage of this utility function for our model is less appropriate. The log-likelihood of other functions does not differ much, except for Code and Kirchhoff utility functions that give lower values. While Kirchhoff gives the highest values when the fare is very low or very high, the shape of Code utility function when the fare is low is not as expected and is therefore questionable.

The graph shows the utility functions for only one generalized cost parameter – public transport fare. In the mode choice model EVA2 the utility functions were used, though final log-likelihood for some utility functions was higher. The reason is that no outstanding results for different generalized cost parameters and different purposes (which gave different values of final log-likelihood) were found. In general, different utility functions for different generalized cost parameters would fit best, varying from purpose to purpose. Using the Fundamental Counting Principle, the number of possibilities when choosing one of nine different utility functions for each ten generalized cost parameters for each purpose is

$$N_{possibilities} = 9^{10} = 3,486,784,401 \tag{19}$$

The total number of 3,486,784,401 possibilities for each purpose would be impossible to explore.

## 6. CONCLUSION

The paper presents the estimation of nine types of utility function parameters for application in EVA mode choice model for the city of Ljubljana, Slovenia. The method used for EVA mode choice model parameter estimation was the Maximum Likelihood method that enables comparison among nine types of utility functions according to final log-likelihood. Since absolute differences in final log-likelihood among most types of utility functions are not high despite the different shapes, different functions would best describe different variables. To decide on the function that fits best, the usage of various types of utility functions for each generalized trip cost parameter would be necessary, which would implicate a rather high number of combinations.

Dr. **TOMAŽ MAHER**
E-mail: tomaz.maher@fgg.uni-lj.si
**IRENA STRNAD**
E-mail: irena.strnad@fgg.uni-lj.si
Dr. **MARIJAN ŽURA**
E-mail: mzura@fgg.uni-lj.si
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo
Prometno tehniški inštitut
Jamova cesta 2, 1000 Ljubljana, Slovenia

***POVZETEK***

***OCENA PARAMETROV EVA MODELA IZBIRE PROMETNEGA SREDSTVA Z RAZLIČNIMI TIPI UPORABNOSTNIH FUNKCIJ***

*V članku je predstavljena ocena parametrov devetih tipov uporabnostnih funkcij za uporabo v EVA modelu izbire prometnega sredstva za mesto Ljubljana. Obravnavana so bila štiri prometna sredstva (osebni avtomobil, javni promet, kolo in peš) in pet namenov potovanj (služba, šola, nakup, prosti čas in ostalo). Najprej je prikazana anketa izražene preference, nato EVA model, različni tipi uporabnostnih funkcij in metoda ocene parametrov. Logaritem verjetja omogoča primerjavo med različnimi tipi uporabnostnih funkcij. Rezultati kažejo majhne razlike v logaritmu verjetja med večino tipov uporabnostnih funkcij, navkljub različnim oblikam funkcij, kar nakazuje, da različne funkcije lahko bolje opišejo različne parametre generaliziranih stroškov.*

***KLJUČNE BESEDE***

*izbira prometnega sredstva, anketa izražene preference, uporabnostne funkcije, metoda največjega verjetja*

## LITERATURE

*Book:*

[1] **Ortúzar, J. de D.**, **Willumsen, L. G.**: *Modeling transport*. Second Edition. John Wiley & Sons, Ltd., West Sussex, 1999.

[2] PTV AG, *VISUM 11.0 Basics*. Karlsruhe, 2009.

*Internet sources:*

[3] http://www.itl.nist.gov/div898/handbook/pri/section3/pri33a.htm (1.9.2009)

*Presentation at a conference:*

[4] **Vrtic, M.**: *Ein hierarchisches („Nested“) Logit-Modell für die Analyse kombinierter Stated- und Revealed-Preference-Daten zur Verkehrsmittelwahl, Vortrag*, 12 DVWG-Workshop Statistik und Verkehr“, Mannheim, 2004.

[5] **Axhausen, K.W.**, **Köll, H.**, **Bader, M.**: *Experiments with SP and CA approaches to mode choice*, European Transport Conference, 1998

[6] **Axhausen, K.W.**: *The Swiss National Model and the Swiss value of travel time savings: Aggregate and disaggregate results*, presentation at the University of Toronto, Toronto, August 2006.

*Article in a Journal:*

[7] **Vrtic, M.**, **Fröhlich, P.**, **Schüssler, N.**, **Axhausen, K.W.**, **Lohse, D.**, **Schiller, C.**, **Teichert, H.**: *Two-dimensionally constrained disaggregate trip generation, distribution and mode choice model: Theory and application for a Swiss national model*, Transportation Research Part, Vol.41, No. 9, 2007, pp. 857-873