EVA mode choice model parameters estimation

Irena Strnad, Marijan Žura

Abstract—In this paper we present estimation of nine types of utility functions parameters for application into EVA mode choice model of city of Ljubljana, Slovenia. First we present design of stated preference survey, then we briefly review EVA mode choice model, present different types of utility functions and Maximum likelihood method as the estimation method. Probabilities of choosing each of four considered modes (private car, public transport, bike, walking) can be calculated by using estimated mode choice model parameters. A practical example of mode choice probabilities for an actual trip is shown at the end. Final log-likelihood enables comparison among different types of utility functions. Results show that absolute differences in final log-likelihood among most types of utility functions are not high in spite of differences in function shapes, which implies that different functions may best describe different variables.

Keywords—Maximum likelihood method, Mode choice model, Stated preference survey, Utility function.

I. INTRODUCTION

Traditional four-step transportation forecasting model consists of trip generation, trip distribution, mode choice and network assignment. An up-to-date disaggregated four-step traffic model for passenger transport of Ljubljana region had already been developed. However, the existent traffic model contains only utility functions of travel time and not of all mode choice affecting trip factors, what indicates the need to upgrade the model. Four modes are included in the traffic model, namely private car, public transport, bike and walking.

First step was to perform a stated preference survey in order to obtain the needed database. Many directions for designing a stated preference survey are described in [1] and [2].

Stated preference survey was performed with portable computers on several locations all over Ljubljana. Several locations were needed to ensure a representative sample and the required sample size for all investigated trip purposes (work, education, shopping, leisure and other). Since the required sample size is 75 to 100 per segment, sample size of 500 is sufficient. References [3] and [4] were taken into consideration when defining relevant trip purposes.

The survey includes questions about the usage of different modes in different situations. The questionnaire consists of ten hypothetical situations in which values of trip factors change. In each situation each traveler was asked to choose the most suitable mode for him if the situation actually occurred.

The second step was to estimate nine types of utility functions for each generalized cost parameter. These were estimated by using Maximum likelihood method.

The last step was to calculate probabilities of choosing each mode for an example trip by using estimated EVA 2 utility functions.

II. STATED PREFERENCE SURVEY

In stated preference survey travelers were first asked about some attributes of current trip, e.g. duration, costs, and about available alternatives. If there were other alternatives available, travelers were asked about similar information for them as well. Each interviewee was then asked to choose the most suitable mode in ten hypothetical situations with different values of trip factors. The questionnaire design, and the generation of the situations will be briefly explained bellow.

A. Questionary Design

In stated preference survey, four modes were taken into account, namely private car, public transport, bike and walking. The following attributes were included for trips made by private car: travel time in minutes, walking time from start station and from final station to destination in minutes. Comfort, price of public transport in euro, frequency in minutes, and about start and destination in minutes. For trips made by bike and on foot, travelers were only asked about travel time.

Fractional factorial design was used to design hypothetical situations needed. Fractional factorial designs are experimental designs consisting of a carefully chosen fraction of the experimental runs of a full factorial design.

The study contains seven factors on three levels and one factor on two levels, as shown in Table I. Factors for car change up and down, whereas factors for public transport only changes for the better. The reason is a transport policy goal to enlarge the share of public transport users in comparison to private car users. That means that only changes for the better in public transport are needed. Some helpful tips for choosing relevant trip factors and their levels were adopted from [5] and [6].

| TABLE I |
| FACTORS WITH THEIR LEVELS |
Table I shows that parking price variability was in percentage. Since some free parking places in Ljubljana are available, the price would not change in situations, which would mean less realistic results of the parking price affect on mode choice. In case of free parking, a new parking price was taken into account to generate situations. For purposes work and education that price was 5€ and for other purposes it was 2€. The prices were chosen on basis of parking duration for different purposes and an average price of parking in Ljubljana. Prices, set like described, than change in situations.

According to number of factors and their levels a Mixed – Level $L_{18} = 2^3 \times 3^5$ Fractional Factorial Design should be built. Generation of eighteen hypothetical situations would therefore be needed. In each of those eighteen situations factors would be on a different level and the traveler would have to choose the most appropriate mode. Generating eighteen hypothetical situations would mean a huge practical barrier, since it would mean a long questionnaire which causes problems in travelers’ concentration and possible choices different to choices made, if the situation actually occurred. We decided to split eighteen designed situations between two interviewed persons, each answering nine situations. Our choice included one control situation, in which factors are the same as given for the actual trip. This situation was a control, if the traveler choice process was in compliance with his actual mode choice.

Survey was performed in two steps. First step was entering requested data of current trip e.g. travel time, costs etc. and listing available alternatives. If there were other alternatives available, entering similar information for alternatives was needed too. On this basis, generation of ten hypothetical situations was made and choosing of most appropriate mode among available in each situation was requested as second step.

An example of choice process in one situation is shown on Fig. 1.

### III. EVA Trip Distribution and Mode Choice Model

In this chapter we briefly review EVA trip distribution and mode choice model, described in [7], present types of utility functions tested, model for calculating mode choice probability and parameter estimation method described in [8].

EVA model generalizes simultaneous trip distribution and mode choice to trilinear model described by (1).

$$T_{jk} = W_{jk} \cdot o_{ij} \cdot d_{ij} \cdot m_{k}$$

(1)
In (1) notation $T_{ijk}$ presents trips from zone $i$ to zone $j$ by mode $k$, $o_i$, $d_j$, and $m_k$ are the balancing factors used to keep marginal sums of productions, attractions, and mode trips, and $W_{ijk}$ are weighted utilities of making trip from zone $i$ to zone $j$ by mode $k$. Weighted utilities are calculated as a product of accessibility of mode $k$ in zone $i$ and product of all utilities of making trip from zone $i$ to zone $j$ by mode $k$ considering each generalized cost attribute for mode $k$ (e.g. time, parking cost, fare,…).

To keep marginal sums of productions and attractions balancing factors $o_i$, $d_j$, and $m_k$ must be determined so, that the following constraints are satisfied:

\[
\sum_{j=1}^{n} \sum_{k=1}^{k} W_{ijk} \cdot o_i \cdot d_j \cdot m_k = \sum_{i=1}^{n} \sum_{k=1}^{k} T_{ijk} = P_i, \tag{2}
\]

\[
\sum_{j=1}^{n} \sum_{k=1}^{k} W_{ijk} \cdot o_i \cdot d_j \cdot m_k = \sum_{j=1}^{n} \sum_{k=1}^{k} T_{ijk} = A_j, \tag{3}
\]

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

### A. 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 mode choice model.

In (5) – (13) $x$ is a generalized cost parameter, $f(x)$ is utility function and $a$, $b$, and $c$ are parameters of the utility function. The following types of utility functions, some with constant, linear and nonlinear elasticity have been studied:

**EVA 1**

\[
f(x) = (1 + x)^{\varphi(x)}, \text{ where } \varphi(x) = \frac{a}{1 + e^{b-cx}}, \tag{5}
\]

**EVA 2**

\[
f(x) = \left[1 + \left(\frac{x}{c}\right)^{b-a}\right], \tag{6}
\]

**Schiller**

\[
f(x) = \frac{1}{1 + \left(\frac{x}{b}\right)}, \tag{7}
\]

**Logit**

\[
f(x) = e^{cx}, \tag{8}
\]

**Kirchhoff**

\[
f(x) = x^2, \tag{9}
\]

**BoxCox**

\[
f(x) = e^{\left(\frac{c^{b-1}}{b}\right)}, \tag{10}
\]

**Box-Tukey**

\[
f(x) = e^{(cx)}, \text{ where } \alpha = \frac{((x+1)^b-1)}{b}, b > 0, \tag{11}
\]

**Combined**

\[
f(x) = ax^b e^{cx}, \tag{12}
\]

**Code**

\[
f(x) = \frac{1}{x + cx^a}. \tag{13}
\]

### B. Maximum likelihood method

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

\[
P_{ijk} = \frac{W_{ijk}}{\sum_{A \in A(ij)} W_{ij}}, \tag{14}
\]

where $A(ij)$ is set of all alternatives between zone $i$ and zone $j$.

Model parameters $a$, $b$, and $c$ have been estimated by using Maximum likelihood method, described in [1].

Likelihood function, which shows the model probability that each individual chooses the option they actually selected is

\[
L = \prod_{q=1}^{Q} \prod_{A \in A(q)} P_{jq}^{g_{jq}}. \tag{15}
\]

Expression $g_{jq}$ appears in likelihood function, which is defined by (16). In both expressions $Q$ stands for a set of all situations conducted in experiment, $A(q)$ alternatives available in situation $q$ and $A_j$ the alternative chosen in situation $q$.

\[
g_{jq} = \begin{cases} 1 & \text{if } A_j \text{ is chosen in } q, \\ 0 & \text{otherwise}. \end{cases} \tag{16}
\]

As it is more convenient to use natural logarithm of $L$, model parameters can be estimated by finding such parameters $a$, $b$, and $c$ where $l$ in (17) has maximum.

\[
l = \ln L = \sum_{q=1}^{Q} \sum_{A \in A(q)} g_{jq} \ln P_{jq}. \tag{17}
\]
IV. ESTIMATED UTILITY FUNCTIONS

Table II shows estimated values of parameters \(a\), \(b\), and \(c\) for nine different types of utility functions for purpose work for one generalized cost parameter – public transport ticket price by using Maximum likelihood method. Parameters were estimated by using optimization software Evolver included in Palisade DecisionTools Suite 5.5 with program Microsoft Excel. Evolver is a genetic algorithm optimization tool, and as the problem is nonlinear, it is needed to find the best solution. Final log-likelihood in Table II was estimated for trip purpose work.

With estimated utility functions parameters values, graphs of utility functions for generalized cost parameter public transport ticket price for purpose work can be drawn and are shown on Fig. 2.

<table>
<thead>
<tr>
<th>Utility function type</th>
<th>parameter (a)</th>
<th>parameter (b)</th>
<th>parameter (c)</th>
<th>Final log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVA 1</td>
<td>1.425</td>
<td>4.733</td>
<td>0.834</td>
<td>-598.126</td>
</tr>
<tr>
<td>EVA 2</td>
<td>2420.575</td>
<td>1.982</td>
<td>216.720</td>
<td>-595.849</td>
</tr>
<tr>
<td>Schiller</td>
<td>5.987</td>
<td>4.441</td>
<td>/</td>
<td>-679.460</td>
</tr>
<tr>
<td>Logit</td>
<td>/</td>
<td>/</td>
<td>-0.306</td>
<td>-601.329</td>
</tr>
<tr>
<td>Kirchhoff</td>
<td>/</td>
<td>/</td>
<td>-0.543</td>
<td>-618.590</td>
</tr>
<tr>
<td>Boxcox</td>
<td>/</td>
<td>1.728</td>
<td>-0.043</td>
<td>-590.741</td>
</tr>
<tr>
<td>Combined</td>
<td>0.720</td>
<td>-0.069</td>
<td>-0.454</td>
<td>-586.905</td>
</tr>
<tr>
<td>Code</td>
<td>3.356</td>
<td>0.106</td>
<td>0.010</td>
<td>-626.364</td>
</tr>
<tr>
<td>Box-tukey</td>
<td>/</td>
<td>0.512</td>
<td>-0.056</td>
<td>-581.579</td>
</tr>
</tbody>
</table>

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

Shape of Kirchhoff utility function is unbelievable at first sight, since its values are high (not close to zero) even when higher public transport ticket prices are high. But since the probability of using each mode is quotient between weighted utility of that mode and the sum of weighted utilities of all available modes, the height of function graphs is not
important.

Lower values of final log-likelihood for three out of four convex functions (Schiller, Logit, Kirchhoff) are not surprising, but Combined utility function with high final log-likelihood is surprisingly convex, which may propose that Combined utility function is not the best fit for this particular generalized cost parameter.

Final log-likelihood is the lowest, when Schiller utility function is used. Graph shows that Schiller utility function is the lowest of them all and usage of this utility function for our model is less appropriate. Log-likelihood among other functions does not differ much, except Code and Kirchhoff utility functions give lower values. Whereas Kirchhoff utility function has the highest values when ticket price is very low or very high, the shape of Code utility function when ticket price is low is not as expected and is therefore questionable.

We chose to use EVA 2 utility functions in final mode choice model, even though final log-likelihood for some utility functions is higher for the particular trip purpose. The reason for choosing this type of utility function is that no outstanding results for different generalized cost parameters and different purposes (which give us different values of final log-likelihood) were found.

In general, different utility functions for different generalized cost parameters would fit best, differently from purpose to purpose. Due to the Fundamental Counting Principle, number of possibilities when choosing one of nine different utility functions for each ten generalized cost parameters for each purpose is

\[ N_{\text{possibilities}} = 9^{10} = 3486784401. \]  

(19)

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

V. ESTIMATED EVA 2 UTILITY FUNCTION PARAMETERS

Table III shows estimated values of parameters \( a \), \( b \), and \( c \) of EVA 2 utility function for purpose work.

<table>
<thead>
<tr>
<th>TABLE III</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVA 2 UTILITY FUNCTION PARAMETERS – PURPOSE WORK</td>
</tr>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>Travel time – private car</td>
</tr>
<tr>
<td>Travel time – public transport</td>
</tr>
<tr>
<td>Walking time - public transport</td>
</tr>
<tr>
<td>Public transport ticket price</td>
</tr>
<tr>
<td>Travel time - bike</td>
</tr>
<tr>
<td>Travel time - walking</td>
</tr>
</tbody>
</table>

On Fig. 3 graphs of probabilities of choosing each mode are shown for an example trip with factors in Table IV, where public transport ticket price changes.

| TABLE IV |
| EXAMPLE TRIP FACTORS |
| Factor | Unit | Factor value |
| Travel time – private car | min | 15 |
| parking price | € | 1.6 |
| Travel time – public transport | min | 30 |
| Walking time needed- public transport | min | 2 |
| Travel time - bike | min | 25 |
| Travel time- walking | min | 45 |

Fig. 3 shows that the most preferable mode for the example trip is private car, as the probability of choosing it is the highest for all public transport ticket prices. Relative difference in probability of choosing private car in comparison to public transport is higher when public transport ticket price is higher.

For low ticket prices usage of public transport is preferable to bike and walking, whereas for higher public ticket prices usage of bike and even walking is preferable.

Probability of choosing public transport is the highest when public transport ticket price is zero, and it is falling with increasing public ticket price, whereas probability of choosing other modes increases when increasing public transport ticket price, which is the expected result.
VI. CONCLUSION

In order to estimate utility functions Maximum likelihood method was used, which enables comparison among nine types of utility functions according to final log-likelihood. For the best fit, usage of different types of utility functions for each generalized cost parameter would be needed, which would mean a much too high number of combinations. For final EVA mode choice model EVA 2 utility functions were chosen, even though final log-likelihood for some utility functions for trip purpose work is higher. The reason is that no outstanding results for different generalized cost parameters and different purposes (which give us different values of final log-likelihood) were found.

REFERENCES

http://tu-dresden.de/die_tu_dresden/fakultaeten/vkw/ivs/tvp/Downloads/Heft-5.pdf