Proceedings of the 14th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'14), 15–18 October 2014, Riga, Latvia, p. 2–8. ISBN 978-9984-818-70-2
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

IDENTIFYING MODAL SHIFT BY UTILITY FUNCTIONS TO REACH AN OPTIMAL POINT OF REGIONAL DEVELOPMENT

Tamas ANDREJSZKI[1], Maria CSETE[2], Adam TOROK[3]

[1] Ph.D. student, Budapest University of Technology and Economics, Department of Transport Technology and Economics, Stoczek utca 2 Budapest H-1111, tamas.andrejszki@mail.bme.hu
[2] Associate professor, Budapest University of Technology and Economics, Department of Environmental Economics, Magyar tudósok körútja 2 Budapest H-1117, csete@eik.bme.hu
[3] Assistant researcher, Budapest University of Technology and Economics, Department of Transport Technology and Economics, Stoczek utca 2 Budapest H-1111, atorok@kgazd.bme.hu

The aim of this study was to analyze the modal shift of passengers by analyzing their preferences. If the preferences of passengers are known it is possible to simulate how the modal shift of the investigated area would change if there were some changes in the investigated parameters of transport services (e.g.: implementing a new transport service or introducing new development strategy).

To capture the preferences of passengers stated preference method was used in online questionnaire. Five key factors were identified (from the point of passengers): travel cost, travel time, comfort, safety and environmental efficiency. In these factors three levels was predefined as simplification which made the base of the choice model. At every replier got three alternatives and they were told to choose the best for themselves. From the results of the questionnaire the formulas and the parameters of the mode choice utility function was derived and identified. In the reviewed statistical sample an exponential utility function showed the best matching. For the validation process a probability model was set up to be compared to the proportions of the utilities.

With this utility function it is possible to handle possible future transport services by evaluating the services through the defined five factors. Based on the introduced statistical approach the described method can be used to identify the effect of transport modes on regional development.
Keywords: Utility function; Modal shift; Stated preference

1. Introduction- the explanation of stated preference method

In the literatures stated preference method refers to two different concepts so it is important to define clearly the frames of the examination. The demand can be identified by the curves of indifference so the forecast of demands requires knowing the preferences of the consumers which is equal to the utility function of them. The direct way to interview consumers is a possible method (which is generally used) but it cannot be said a completely suitable tool to get the preferences. The individuals do not have real interest to reveal their preferences because there are no consequences of the answers so their wallet will not feel the gravity of the decision. Their decision is just a reaction to a hypothetic situation. An objective evaluation can be given only if the actual decisions are known so the preferences can be revealed only by the observation of the market behaviour. In 1947 P. A. Samuelson worked out the method of stated preference which makes it possible to simulate approximately the consumer preferences (plus the curves of indifference and the utility function) from factual data (e.g. prices, income, and demanded quantities). Based on Samuelson's work it can be represented that the curves of indifference can be approximately identified from the information of the purchase if exact prerequisites are true *(Karajz, 2008)*.

Nevertheless making interviews and questionnaires seems to be the best way to reveal transport demands of a future transport service (*Kampf* et. al., 2012). According to Kroes and Sheldon's definition the term "stated preference method" refers to a family of techniques which use individual respondent's statements about their preferences in a set of transport options to estimate utility function

(Kroes and Sheldon, 1988). Different stated preference methods are available under a wide variety of names; the best known methods are:

- Conjoint analysis;
- Functional measurement;
- Trade-off analysis;
- The transfer price method.

These methods were originally developed to marketing researches in the beginning of the 70's but a study from 1978 made them known *(Green and Srinivasan, 1978)*. In this study the authors gave the following description to define the conjoint analysis (which seems to be the most suitable for transport purposes): every method that aims to estimate the structure of the consumer preferences and the consumers evaluate options where the levels of the different quantities have been defined before.

2. Design consideration

The first step of designing a stated preference method examination is to identify the relevant variables (factors) and the values belong to each factor (levels). A related task is here the specification of the mathematical formula of the utility function which refers to the authors' hypothesis about how the integrated preference comes from the individual preferences. The linear, additive, compensational model is the mostly used form which has the following structure:

$$U = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n \qquad (1)$$

Here U refers to the complete utility; x_n is the value of factor n and α_n is the utility weight of factor n. It is practical if the sum of the utility weights is 1.

The factors can be defined as continuous variables or as a group of discrete variables also. The stated preference method can also be used to test alternative hypothesises *(Kroes and Sheldon, 1988)*.

The next step in planning is the optimization of the mathematical option combinations. According to the experiences it is worth decreasing the number of options because the respondents can be spared by not answering questions that are trivial. If the number of questions is less the willingness of respondents to fill out completely the questionnaire might be higher because it will not need that much time from them. If the number of factors and the belonging levels is given the needed number of combinations can be calculated.

The factorial structure (*Kroes and Sheldon, 1988*) refers to a combinatorial expression. Because of that the full factorial structure means all the possible combinations of the options and the partial factorial structure means an exact part of the full factorial one. There are values belong to the factors; and an exact value of an exact factor is called "quality". In an option of a question there are more qualities but none of them comes from the same factor (e.g. in a question the first option to choose is a travel that takes 10 minutes, worth 3 €-s and has a low comfort level). The full factorial structure generates too much options and combinations at higher number of factors and levels so the partial factorial structure might be better to go on with.

The examination of stated preference method can be done by two possible ways. The first opportunity is when the questioner creates cards from preference possibilities and options and at each question more cards are given to the respondent. The task of the respondent here is to make a sequence from them. The second opportunity is to make choice option cards which contain predefined questions with predefined options so the task of the respondent is to tick the best option of the card that he/she would choose in the given situation *(Kroes and Sheldon, 1988*).

3. Identification of transport utility function

At every question the respondent is asked to choose one (the best) from three options. Based on the international literature in this model five key factors were considered as playing important role in decision making (*Simecki* et. al., 2013): the travel time, travel cost, comfort, safety and environmental friendliness. Each factor has three values (one bad, one middle and one good) so there are 15 qualities which are the followings:

- Travel time (T) → 30 minutes, 20 minutes and 10 minutes;
- Travel cost (TC) → 4 €, 2 € and 1 €;
- Comfort (C) → not comfortable, more or less comfortable and comfortable;

- Safety (S) → not safety , more or less safety and safety;
- Environmental friendliness (E) → not environmental friendly, more or less environmental friendly and environmental friendly.

These exact qualities come from a transport situation that is given to the respondent so they have their meanings. In the situation the respondent lives in a small town and want to get to the train station to go to work in a weekday morning. By car this journey can be done in 10 minutes if there are no traffic jams. Comfort refers to the quality level of the transport service that was used (e.g. a crowded dirty bus means a non-comfortable mode but a clean car or train can represents a comfortable level; but comfort is not linked to any transport mode) (*Duleba* et. al., 2013). Safety in this meaning refers to the number of accidents that happens on the used section of road per one year. On a not safety section there are 12 accidents per year, the middle value is 4 accidents per year and the most safety value is 0.5 accidents per year. At environmental friendliness the emission level of an internal combustion engine was the sample for the so called bad level. This means more or less 179 g from CO_2 per kilometre (*Kok, 2013*). The good level has almost zero emission like walking and cycling.

If we used all the qualities of all the five factors to create the three options of one question, that would cause 360 questions to be asked (if every quality appears maximum once in one question). To reduce this number at the beginning of the questionnaire respondents are asked to choose the three most relevant factors from the enumerated five. After this decision the following questions will just deal with the chosen three factors and count the non-chosen factors with a zero parameter in the individual utility function. According to this operation there are ten versions of the questionnaire:

(2)

In one question there are always three options. In all the options there are three qualities from three different factors. According to combinatory this means three repeated variation:

(3)

But these 216 questions contain same questions with different order of the options. To have the real number of possible question 216 should be divided by the number of possible ordering:

(4)

These 36 questions are equal to the full factorial structure. For further reduction the trivial questions should be selected. In this case the expression "trivial" refers to those questions that have an option which contains three good qualities or two good qualities and one middle quality. The model handle these questions like these were answered by the respondent in a logical way so they always choose this outstanding option. After the selection of these trivial questions 20 questions remain that can be asked from the respondents. This amount seems to be user friendly and gives the hope of high filling rate.

4. The implemented questionnaire

In the implemented online questionnaire the transport situation was written first. Then the respondent chose the three more relevant factors. From these factors the respondent got 20 questions to answer. The questions were like *Figure 1*. As it can be seen this kind of questions includes partly the appointment of WTP (Willingness to pay) (*Drevs et. al., 2014*).

1. From the following options what would you choose?

O 1) You would pay 4 € to take your journey in 30 minutes in a more or less comfortable vehicle.
O 2) You would pay 2 € to take your journey in 20 minutes in a comfortable vehicle.
O 3) You would pay 1 € to take your journey in 10 minutes in a not comfortable vehicle.

Figure 1. One question of the questionnaire

The questionnaire was filled out correctly by 462 respondents. The ages of the respondents are shown in *Figure 2*. As it can be seen the questionnaire was not representative (at the ages) but because of the time and cost constraints of the examination this was not an expectation.

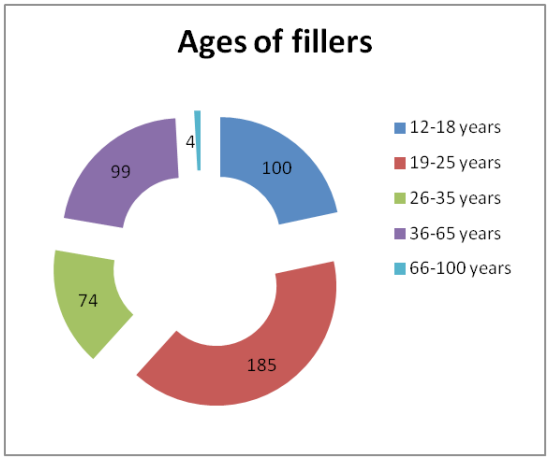

Figure 2. The ages of the respondents

4.1. The algorithm of evaluation

The basis of the evaluation was to give 1-1 point to the qualities of the chosen option and give 0 point to the qualities of the non-chosen options. In this case each quality can gain 20 points as maximum and zero points as minimum. At this moment the 16 questions are added to the real answers then the maximum becomes 36 points. If the given factor is not important for the respondent -so it does not have a high preference – the points of the qualities of this factor will be around the one third of all questions

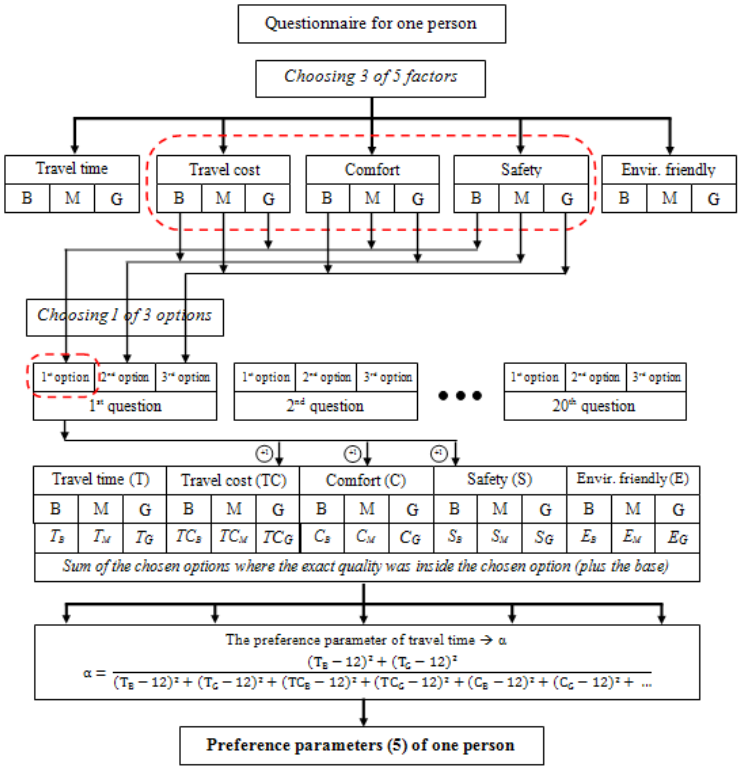

Figure 3. The process of the questionnaire and the evaluation

which is 12. If the factor was relevant for the respondent the good quality might get a higher score or (because avoiding the bad quality of this factor is also a preference) the bad quality might get

a lower score. From these values the individual utility function should be calculated. The parameter of one factor of the individual utility function is rational number in the interval of [0; 1] where:

- 0 is the preference parameter if the factor is absolutely not relevant;
- 1 is the preference parameter if the factor is absolutely relevant.

The parameter of factor 1 should be calculated by the following method. After the points of each quality are summarised the next step is to calculate the square of the distances between the good quality and the average value (which is 12 now) and then to add the square of the distances between the bad quality and the average value. It is enough to examine just the good and the bad quality because (as it was mentioned before) the possible strategies of the respondents do not appear in the middle quality. So getting the middle quality into the algorithm would distort the utility function. After each factor has this basic value this value of factor 1 should be divided by the sum of this basic value of each factor. This process causes that the sum of the parameters gives 1. So if factor 1 is absolutely relevant its' parameter will be 1 and the other factors' parameter will be 0. The whole procedure can be seen in *Figure 3*.

The utility weights of the complete utility function come from the averages of the individual parameters. The structure of utility function was also the object of the examination. Linear, exponential and logarithmical models were considered but the most effective was the following structure:

(5)

In this formula the same notation is used like in equation (1). The values that are connected to the different levels are 1 for bad qualities, 2 for middle qualities and 3 for good qualities. In this case the worst combination of bad qualities causes the zero utility.

4.2. Process of validation

The accuracy of the model can be validated by the examination of the decision situations. The question is that: what is the ratio between the quotient of the utilities of two options and the quotient that shows how many people preferred the first option against the second.

To prepare the probability matrix the first step is to integrate the 10 versions (*Figure 4*). This is not trivial because the versions were filled out by different amount of people and in one question the complete order is not known because the respondent only chose the best option (so the relation of the two not chosen options is not known). So firstly 10 preference matrixes were created in the sizes of 27*27. In the columns and rows there are all the mathematically possible options so one element means that how many times were the option of the row chosen against the option of the column. The second step is the creation of another 10 matrixes called answered matrixes. Here the elements mean that how many times the respondents chose from the two options.

The probability matrix has $3^5 = 243$ rows and 243 columns because in the integration all the 15 qualities of the five factors should be counted with. Every element is the quotient of choosing the option of the row against the option of the column. These options have five dimensions. In this five-dimension option there are 10 three-dimension options that can be found in the 10 preference matrixes and answered matrixes. So to get one element of the probability matrix the appropriate cells of the 10 preference matrixes should be summarised and then this sum should be divided by the sum of the appropriate cells of the answered matrixes.

In the edge of the utility matrix there are the same 243 options as in the probability matrix. One element means the quotient of the utility of the row option and the utility of the column option. Then the next step is to create a matrix in the same size where the elements show the relation between the utility and the probability matrixes. In this validation matrix the value is 1 if the two quotients are similar and 0 if not. Similarity means the followings:

- The row option is better [0; 0.45]
- The row option is similar to the column option]0.45; 0.55]
- The column option is better]0.55; ∞[

So if the quotients are in the same interval the validation matrix element gets 1 if not it gets 0. After having all the values the average of them will give the accuracy of the model. In this case the accuracy level of 73.12% was reached. This level was accepted for further examinations with the created transport utility function.

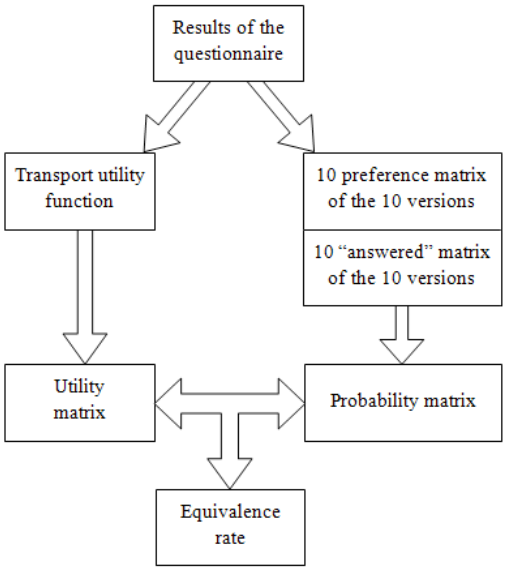

Figure 4. The process of validation

4.3. Evaluation of transport projects by the utility function - Conclusion

With the above demonstrated process the utility of passengers can be determined. The analysis of this hidden and personal utility can help the professional transport planners to identifying the key parameters that play an important role in transport development (*Stenci, Lendel, 2012*) or enhance modal shift (*Cerny* et. al., 2014). This could help to find the optimal path of regional development.

The aim of this study was to show an easy method to statistically analyse passenger preferences. If the preferences of passengers are known it is possible to stimulate modal shift. To capture the passengers' stated preferences an online questionnaire was built. Five passenger focused key factors were identified: travel time, cost, comfort, safety and environmental friendliness. In these factors three levels was predefined as simplification which made the base of the choice model. Although the statistical sample was not representative, this method gives a clear guideline for cities, companies and planners to create their questionnaires and make their sample representative. From the results of the questionnaire the parameters of the mode choice utility function were statistically estimated. An exponential utility function was used as it had the best fit for the examined sample. For the validation process a probability model was set up to be compared to the proportions of the utilities.

With this utility function it is possible to handle possible future transport services by evaluating the services through the defined five factors. It is feasible to compare the possibilities of transport developments, and the opportunity is given to make the comparison by measurable statistical indicators. Based on the introduced statistical approach the described method can be used to identify the effects of transport modes on regional development.

Acknowledgements

The authors are grateful to the support of Bolyai János Research fellowship of HAS (Hungarian Academy of Science).

References

1. Černýa J, Černáa A, Lindab B (2014): Support of decision-making on economic and social sustainability of public transport, *Transport*, 29(1):59-68, doi: 10.3846/16484142.2014.897645
2. DulebaaSz.,ShimazakibY.,MishinabT.(2013):Ananalysisontheconnectionsoffactorsinapublic

transport system by AHP-ISM, *Transport*, 28(4):404-412, doi: 10.3846/16484142.2013.867282
3. Green, P. E., Srinivasan, V. (1978) - Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5, 103-123.
4. Kampf R., Gašparík J., Kudláčková N. (2012): Application of different forms of transport in relation to the process of transport user value creation, *Periodica Polytechnica Transport Engineering*, 40(2):71-75, *doi: 10.3311/pp.tr.2012-2.05*
5. Karajz, S. (2009) - Közgazdasági elméletek (Economical theories), *tutorial notes*
6. Kok, R. (2013) New Car Preferences Move Away from Greater Size, Weight and Power: Impact of Dutch Consumer Choices on Average CO2-emissions. *Transportation Research Part D: Transport and Environment* 21 (June): 53–61. doi:10.1016/j.trd.2013.02.006.
7. Kroes, E.P., Sheldon, R.J. (1988) - Stated preference methods: an introduction, *Journal of Transport Economics and Policy*
8. Kumar M., Sarkar P., Madhu E. (2013): Development of fuzzy logic based mode choice model considering various public transport policy options, *International Journal for Traffic and Transport Engineering*, 3(4):408-425, doi: 10.7708/ijtte.2013.3(4).05
9. Šimecki A., Steiner S., Čokorilo O. (2013): The accessibility assessment of regional transport network in the south east Europe, *International Journal for Traffic and Transport Engineering*, 3(4):351-364, doi: 10.7708/ijtte.2013.3(4).01
10. Štencl M., Lendel V. (2012): Application of selected artificial intelligence methods in terms of transport and intelligent transport systems, *Periodica Polytechnica Transport Engineerin*g, 40(1):11-16, doi: 10.3311/pp.tr.2012-1.02
11. Drevs F., Tscheulin D. K., Lindenmeier J., Renner S. (2014): Crowding-in or Crowding out: An Empirical Analysis on the Effect of Subsidies on Individual Willingness-to-Pay for Public Transportation, *Transportation Research Part A: Policy and Practice* 59 (January): 250–61. doi:10.1016/j.tra.2013.10.023.