ASSESSING THE POTENTIAL FOR A NEW SUBURBAN RAIL: DO USER CHARACTERISTICS MATTER?

S. BASBAS[1], A. NIKOLAIDOU[2], P. PAPAIOANNOU[2], I. POLITIS[2]
[1]Faculty of Rural & Surveying Engineering, Aristotle University of Thessaloniki, Greece
[2]Faculty of Civil Engineering, Aristotle University of Thessaloniki, Greece

ABSTRACT

This article examines the role user-specific characteristics can play on assessing a new public transit service. The case under consideration is the new Western Suburban Rail service in Thessaloniki, Greece, which will cover a suburban area. Binary logit models were developed in order to investigate the potential demand for this project. A revealed and stated preference questionnaire survey was conducted. The sample of the survey was stratified according to the three current travel options, namely car, regular-fare bus and low-fare bus. The revealed preference survey included questions about trip and socio-economic characteristics of each user. The stated preference part examined the potential for the new mode in comparison to the current travel modes through changes in travel cost, travel time and number of transfers. Six binary logit choice models were developed in total, two for each user group. For each group, a simplified model was initially calibrated only with the mode-specific variables; at a next step user-specific variables were added and a second extended model was created. The results show significant relationship of mode- and user-specific variables to the final choice for all the three user groups. Likelihood ratio tests were used to examine whether user-specific attributes improve the overall fit of the models or not, compared to traditional choice models of travel time, travel cost and number of transfers. The outcomes show that the inclusion of user-specific variables improved the overall fit and the explanatory power of the models.
Keywords: choice models, discrete choice models, revealed and stated preference, suburban rail, time and cost elasticities.

1 INTRODUCTION

The success of a new public transit (PT) mode or service depends on a large number of inter-related factors pertaining to economic, functional and demand characteristics. High functional standards normally translate to higher investment and operation and maintenance costs but at the same time improve attractiveness and demand. Determining the right trade-off amongst them is the subject of well-prepared and formulated demand and revenue forecasting studies as well as feasibility studies which assess the economic value of a specific project or intervention.

A question that is of relevance is how much the trip and the user characteristics of those who will potentially use the new service play an important role in the outcome of such studies. Do sociodemographic data of a geographical area affect for example mode choice and ultimately transport demand compared to another geographical area with different characteristics?

This article attempts to provide responses to this question by examining how specific trip and user characteristics, including trip purpose, gender, age group, occupation, etc., may affect mode choices and future demand of new PT services. More specifically the article will present the research made in a real case in Thessaloniki, Greece, which is about the creation of a new suburban railway service covering a geographical extended area not currently served at all by rail but only by buses and private cars.

The research study has used data provided by the local partners of a European project consortium, entitled RAIL4SEE. The specific project has been co-funded by the European Union SEE Programme. The data mainly refer to revealed and stated preferences collected through a questionnaire survey of persons travelling between origin and destination points that will be served in the future by the specific new suburban rail service.

It must be noticed at this point that the authors of the article did not participate in the questionnaire-based survey but they used the raw data, which they processed afterwards according to the needs of the current research.

The article is structured as follows: after this introductory section, the following is devoted to the description of the undertaken research. The development of choice models is described in Section 3 whereas the models results are given in Section 4. Finally, the conclusions drawn are given in the last section of the article.

2 DESCRIPTION OF THE UNDERTAKEN RESEARCH AND THE DESCRIPTIVE STATISTICS RESULTS

2.1 Introduction

The aim of this research as previously mentioned has been to investigate how much the various trip and user characteristics of the potential users will affect mode choice and consequently demand for the new rail service. The research was mainly based on data gathered through a questionnaire survey conducted within the context of a study about the specific new rail service. The design and the execution of the survey were accomplished by trained staff of HIT, a partner of the RAIL4SEE project consortium. The survey was conducted during the period 12–18 December 2013 at various locations along the planned rail route and at the main origin/destination location, the Thessaloniki railway station. A short description of the project under consideration is provided below for completeness.

2.2 The project

The RAIL4SEE project aims to develop in South East Europe a network of 11 railway hubs in 11 cities (Bologna, Venice, Trieste, Ljubljana, Vienna, Bratislava, Budapest, Thessaloniki, Sofia, Zagreb and Bucharest) in order to:

- effectively connect these cities by train (new routes, better management of the existing services, passenger information technologies, 'integrated' and electronic ticket);
- promote the development of public transport based on a fixed track system at metropolitan and regional level;
- implement new forms of cooperation between bodies of government responsible for the development of new services for fixed track systems and combination with other means of public transport; and
- use new technologies for passenger information and ticket payment with the aim of the international movement of passengers.

For Thessaloniki, one of the purposes of the RAIL4SEE project is launching a new suburban rail link between the port of Thessaloniki and the western suburbs of the metropolitan area of Thessaloniki and other major traffic generators located nearby (university, industrial area, etc.).

Namely the aim of the project is the functional interface with an alternative fixed track mode of transport of the centre of Thessaloniki with the western urban and suburban zone, as well as the strengthening of public transport in Thessaloniki which are traditionally underserved, with minimum resource consumption.

Also, with this project will be achieved the interconnection of the urban public transport of Thessaloniki with the national railway network (e.g. with satellite cities around the city of Thessaloniki) and the decongestion of the central passenger railway station.

The implementation of the above project will include the reconstruction and reuse of existing inactive parts of the railway network and/or the construction of new sections in order to operationalize the suburban rail link of the areas in west Thessaloniki with the city port.

2.3 The questionnaire survey

In order to estimate the potential demand of the proposed suburban rail link, a combined revealed and stated preference questionnaire survey was conducted. The sample of the survey was stratified according to the three current travel options; by car, by regular-fare bus and by reduced fare bus (available only for special categories such as students, elderly, etc.). This separation of the sample was performed based on the fact that for each user category different games have been formed, in which the values of time and travel cost were different depending on the mode of transport already used by the respondent.

The questionnaire consisted of three sections. The first section is about the trip characteristics of the respondent, for example the mode used to travel to his/her destination, the travel frequency, the estimated trip cost and time, the trip purpose, the origin point, the final destination, etc. The second section of the questionnaire pertains to the stated preference experiment of hypothetical scenarios regarding the choice of the respondent amongst the current travel options and the new proposed suburban rail link. The third section of the questionnaire consisted of questions regarding interviewees' socio-economic characteristics.

A prerequisite for somebody to take part in the completion of the questionnaire was the making of a trip by bus or car in the recent past between a preselected location along the new rail route and the Thessaloniki's port area. The latter could be either one of the trip ends or a connection point.

Another important element regarding the design of this questionnaire was the developing of the scenarios relating to the stated preference section of the questionnaire and the intentions of the respondent. The stated preference survey is a highly specialized research with particular relevance to transport authorities and operators, who want to create a new service, because they should have a correct understanding of the impact of the specific service to users before they make any investment (Papaioannou et al. [1]).

With the particular stated preference survey users were asked about what they would choose for a given situation that they know or it is described to them. The utility of stated preference surveys lies in that it is possible to quantify the preferences of respondents for some characteristics of services in relation to others (Basbas [2]). The most widely used technique of these surveys is the so-called conjoint analysis which was used in this survey (Hensher & Button [3]).

In this particular survey it was decided only to investigate the parameter of fare level of the proposed suburban rail link (one independent variable) in comparison with given values of fare and of today's trips by bus or by car respectively. The parameter of the existence of transfers or not has entered the games, where it exists, not as a criterion of the selection of respondents but as a description of actual and future mobility with the various means of transport. It was considered appropriate to be incorporated in the description of the current movement in the sense that users should be aware of the necessity for transfers, as for many of them it is a kind of burden (penalty) in the choice of travel mode.

Regarding the variation levels of the different values specified in the games, these had to be plausible and realistic. Therefore, for every stop of the suburban rail (corresponding approximately to a survey position) estimates for time and travel costs from each survey location to the port of Thessaloniki with the different means of transport were calculated.

The suburban rail fare levels chosen were €0.80, €1.00 and €1.20. The variation was under the logic that the lower fare is approximately equal to the current bus fare, the median fare level is approximately equal to the existing railway fare for the shortest rail distance between two consecutive stations and the higher level corresponds to a logical fare increase by 20% (TREDIT [4]).

In total, 534 people were interviewed. The development of the choice models is the main work presented in this article (Nikolaidou [5]). After data clearance and missing data manipulation, 479 responses were finally considered as valid. Table 1 presents the descriptive statistics for the three different users categories for the following variables: gender, age, income, occupation, purpose of the trip, frequency of the trip, ownership of a private car and the mode choice.

Table 1: Survey sample descriptive statistics.

Variable/interval		Car users N = 233		Bus users (normal fare) N = 142		Bus users (reduced fare) N = 104	
		Freq.	%	Freq.	%	Freq.	%
Gender	Female	101	43.3	91	25.4	59	54.8
	Male	132	56.7	50	74.6	45	45.2
Age	<18	0	0.0	1	0.7	1	0.7
	19–24	60	25.8	30	21.3	30	21.3
	25–54	156	67.0	100	70.9	100	70.9
	55–64	16	6.9	9	6.4	9	6.4
	>65	1	0.4	1	0.7	1	0.7
Income	Unemployed	20	9.3	23	18.0	23	18.0
	0–800€	66	30.8	52	40.6	52	40.6
	€801–€1600	92	43.0	48	37.5	48	37.5
	€1601–€2400	28	13.1	5	3.9	5	3.9
	>€2401	8	3.7	0	0.0	0	0.0
Occupation	Student	46	19.9	15	10.6	81	77.9
	Free lancer	64	27.7	34	24.1	3	2.9
	Employee	101	43.7	60	42.6	8	7.7
	Retired	11	4.8	17	12.1	8	7.7
	Other	9	3.9	15	10.6	4	3.8
Trip_purpose	Work	41	17.6	33	23.2	3	2.9
	Within work	35	15.0	25	17.6	7	6.7
	Tourism-leisure	68	29.2	30	21.1	28	26.9
	Shopping	36	15.5	22	15.5	13	12.5
	Education	13	5.6	10	7.0	29	27.9
	Private activity	25	10.7	14	9.9	7	6.7
	Change mode	6	2.6	7	4.9	12	11.5
	Other	9	3.9	1	0.7	5	4.8

	>1 daily	2	0.9	0	0.0	0	0.0
	Everyday	41	17.7	29	20.6	29	20.6
	3–4 times/week	46	19.8	33	23.4	33	23.4
Trip_frequency	1–2 times/week	67	28.9	39	27.7	39	27.7
	2–3 times/ month	31	13.4	19	13.5	19	13.5
	1 time/month	39	16.8	20	14.2	20	14.2
	other	6	2.6	1	0.7	1	0.7
Car_own	No	16	6.9	85	59.9	87	83.7
	Yes	215	93.1	57	40.1	17	16.3
Choice	Current mode	61	26.2	25	24.8	30	28.8
	Suburban	172	63.8	107	75.2	74	71.2

3 DEVELOPMENT OF THE CHOICE MODELS

3.1 Utility maximization and theory of consumer behaviour

The study employs binary choice models in order to examine a number of parameters that may have an effect on the choice or not of the alternative fixed track system. The three different sets of games that were formed for each user category are compared each time with the new proposed suburban rail link.

The principle on which binary models are based is that 'The possibility of a commuter to choose a particular "solution" (trip, transport mode, etc.) is a function of the socio-economic status and the relative "attractiveness" of the particular solution' (Dyer [6]).

The attractiveness of a solution for the trip of a passenger depends on the utility of this solution (Basbas et al. [7]). The concept of utility is a theoretical concept defined as 'one factor that the commuter wishes to maximize as he travels'. Choice models assume that the probability of an individual to choose a particular transport mode depends upon the utility gained from travelling by that mode as compared to the utility corresponding to another alternative. In conclusion the mode with the higher utility will be chosen by each individual. The theoretical utility of a mode of transport i to a user t is expressed by the following mathematical relationship (Louviere et al. [12]):

$$U_{it} = V_{it} + \varepsilon_{it} = \beta_t x_{it} + \varepsilon_{it} \qquad (1)$$

where U_{it} is the utility gained by individual (t) for using mode (i) , V_{it} is the systematic or observed term of the utility function, β_t is the vector of unknown (estimated) parameters related to variables x_{it} (mode and user specific) and ε_{it} random components (called disturbances of the utility.

The observed term V_{it} is a function f(x_{it}, S_{it}) of the operational characteristics of the transport mode i (x_i) such as time in vehicle, travel cost, travel distance, etc., and socio-economic characteristics of the user t (S_t), such as age, income, possible psychological step stage for changing travel mode, etc.

Since the individual (t) will choose the mode with the higher utility, in the case of a binary choice model, the following relationship between the utilities of the two alternative modes i and j exists (Louviere et al. [12]):

$$U_{it} > U_{jt} \qquad (2)$$

And the probability of the mode *i* to be chosen by individual (*t*) is written as (Louviere et al. [12]):

$$P_{it} = \text{Prob}\left(U_{it} > U_{jt}\right) \Rightarrow P_{it} = \text{Prob}\left[\left(V_{it} + \varepsilon_{it}\right) > \left(V_{jt} + \varepsilon_{jt}\right)\right]$$

$$\Rightarrow P_{it} = \text{Prob}\left[\left(V_{it} - V_{jt} + \varepsilon_{it}\right) > \left(V_{jt} + \varepsilon_{it}\right)\right] \quad \forall \ i \neq j \tag{3}$$

In the general form of a binary choice model, the possibility of choosing the mode *i* is expressed as (Louviere et al. [12]):

$$P_i = \frac{e^{(\beta_0 + \Sigma j\beta_j x_j)}}{1 + e^{(\beta_0 + \Sigma j\beta_j x_j)}} = \frac{1}{1 + e^{(\beta_0 + \Sigma j\beta_j x_j)}} \tag{4}$$

3.2 Model specification

In order to examine how and at what extent the personal characteristics of the individual can improve the explanatory power of the choice models for the three alternative fixed track systems, two binary choice models were set up for each type of user: a simplified model including only mode-specific variables like travel time, travel cost and number of transfers (model 1) and an extended model including also socio-economic variables like the age, the gender, the income, etc. The process was as follows: first the simplified model was calibrated and afterwards the additional socio-economic variables were successively added. Alternative specific constants were used for the existing travel modes. As an initial assumption, alternative specific beta coefficients were considered for the variables of travel time and cost; however they were replaced with generic ones in case the results indicated similarities between the modes; this represents an assumption that a minute and/or a cent has the same marginal disutility no matter if it occurs on the current or the examined transport mode (Ben-Akiva & Lerman [8]). For the binary choice experiments, the ordinal and nominal variables of frequency, trip purpose, gender and age were coded as dummies.

In total 12 binary choice models were developed, 4 (2 simplified and 2 extended) for each user category. The models were calibrated through SPSS, a software package used for statistical analysis and the parameters were estimated through the maximum likelihood estimation (MLE) method (Shumacher & Lomax [9]).

MLE approach assumes that a given sample could be generated by different population and is more likely to come from one population rather than another. The likelihood function *L** for *N* observation and *K* parameters can be defined as

$$L^*\left(\beta_1, \beta_2, \ldots, \beta_\kappa\right) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}} \tag{5}$$

and the logarithm of *L**, defined as *L*, will be written as

$$L\left(\beta_1, \beta_2, \ldots, \beta_\kappa\right) = \sum_{n=1}^{N} [y_{in} \log P_n(i) + y_{jn} \log P_n(j)] \tag{6}$$

For the comparison of the two models, the simplified and the extended, a likelihood ratio test was performed each time. The likelihood ratio test usually is used to measure the performance of one model relative to other. The mathematical formula for the two models is written as follows (Shumacher & Lomax [9]):

$$-2\log L \; = \; - \left(2\log_{(\text{model 1})} - \; 2\log_{(\text{model 2})} \right) \tag{7}$$

The result of eqn (7) is compared against the critical tabulated chi-squared value at a specific level of confidence for given degrees of freedom. If the estimated L^* value exceeds the critical chi-squared value, the null hypothesis is rejected, meaning that the extended model 2 has a better model fit (e.g. the inclusion of the individual-specific variables does significantly improve the model fit).

4 MODEL RESULTS

4.1 Binary choice models

This section presents the results of the binary choice models that have been developed within the framework of the study for each of the three population segments: the car users, the bus users with regular fare and the ones with reduced fare. Tables 2–4 illustrate the parameter estimates of the binary choice models for each type of user respectively and present the results of the likelihood ratio tests for the comparison of the simplified (Model 1) and the extended models (Model 2). Blank cells indicate non-significant parameters (sig >0.05).

Table 2: Parameter estimates of the binary choice models for car users.

Variable	Model 1			Model 2		
	B	β	sig.	B	β	sig.
Constant						
Specific						
Mode						
User						
Freelancer				0.520	0.054	0.011
Retired/housekeeping				2.122	0.105	0.004
Low income				0.360	0.036	0.044
High income				–0.580	–0.045	0.020
Very high income				1.291	0.057	0.042
Generic						
ΔT	–0.154	–0.255	0.000	–0.119	–0.192	0.000
ΔC	1.308	0.232	0.000	0.995	0.172	0.000
Nagelkerke R^2		0.132			0.213	
Hosmer and Lemeshow test		0.038			0.904	
–2 Log likelihood		895.908			774.666	
–2 Log L			121.242			
LR test		Significant difference at 95% confidential level				

Table 3: Parameter estimates of the binary choice models for bus users with regular fare.

Variable	Model 1			Model 2		
	B	β	sig.	B	β	sig.
Constant						
Specific						
Mode						
Length	–0.076	–0.078	0.001	–0.102	–0.096	0.000
Transfers	0.688	0.114	0.000	0.467	0.071	0.001
User						
Education				1.759	0.085	0.019
Many times_in the week				1.048	0.081	0.001
Freelancer				0.987	0.080	0.001
Retired/housekeeping				1.378	0.085	0.002
Generic						
ΔT	0.166	0.150	0.000	0.163	0.136	0.000
Nagelkerke R^2		0.349			0.437	
Hosmer and Lemeshow test		0.001			0.063	
–2 Log likelihood		454.818			409.464	
–2 Log L			45.354			
LR test			Significant difference at 95% confidential level			

Table 4: Parameter estimates of the binary choice models for bus users with reduced fare.

Variable	Model 1			Model 2		
	B	β	sig.	B	β	sig.
Constant	6.046		0.000	6.735		0.000
Specific						
Mode						
Cost_train	–6.673	–0.167	0.000	–7.205	–0.180	0.000
User						
Unemployed				–0.640	–0.065	0.022
Generic						
Nagelkerke R^2		0.161			0.199	
Hosmer and Lemeshow test		0.588			0.841	
–2 log likelihood		337.636			312.947	
–2 Log L			24.689			
LR test			Significant difference at 95% confidential level			

as are presenting the results of the likelihood ratio tests for the comparison of the simplified (model 1) and the extended models (model 2).

Apart from the unstandardized betas estimates (b), the tables are also presenting the standardized betas (β) for comparative analysis of the strength of the prediction across the variables. The calculation of the standardized betas was done through the following formula (King [10]):

$$\beta = \frac{1}{1+\exp\left(-\left(\ln\dfrac{P_{\text{Ref}}}{1-P_{\text{Ref}}}\right)+\dfrac{1}{2}\hat{b}_s\right)} - \frac{1}{1+\exp\left(-\left(\ln\dfrac{P_{\text{Ref}}}{1-P_{\text{Ref}}}\right)-\dfrac{1}{2}\hat{b}_s\right)} \qquad (8)$$

where P_{Ref} = a probability value used as a reference point, $\hat{b}$ = the unstandardized logistic regression coefficient, and s = the sample standard deviation.

In order to use the formula, the following steps were taken: (a) calculation of the sample standard deviation for each variable, (b) calculation of the logistic regression predicted probabilities, (c) calculation of the mean of those values to obtain P_{Ref}, and (d) substitution of s, $\hat{b}$ and P_{Ref} into the function.

4.2 Binary choice models for car users

In the two models that were constructed for this user category it was found that the constant terms were not statistically significant and thus car users have no prior preference for any of the examined means of transport.

In the particular models, instead of the mode-specific variables generic variables were used. More specifically, the increase in travel time resulted in a reduced probability of the selection of the suburban approximately by 14%, while the increase of travel cost will increase the same probability by 78% (the presented percentages were calculated with the use of eqn (4)). It is worth noting that variables ΔT and ΔC were defined as the travel time and cost difference of the suburban from the car. Car users mainly choose at a greater degree the suburban rail when the travel cost of the car is substantially increased, whereas when the travel time of car is increased they don't show willingness to switch to the new transport mode.

The variables related to the socio-economic characteristics of users are two dummy variables for the occupation and three dummies regarding the income of the respondents. With regard to the occupation, the freelancers and the retired choose to a greater extent the suburban than the rest of the sample, while in the case of income users who declare income of €0–€800 tend more to the suburban unlike those who belong to the category €1601–€2400.

The standardized coefficients of the models show that the influence of the generic variables is clearly greater than that of the variables related to socio-economic characteristics of the users.

According to the results of the statistical tests the simplified model does not fit the sample data, while the addition of variables of socio-economic interest significantly improves the adaptability of the model.

4.3 Binary choice models for regular-fare bus users

In the case of models for regular-fare bus users, it was found that the constant term of the logistic regression is not statistically significant and therefore users do not express a prior preference for any of the means of transport in question.

Regarding the mode-specific variables, statistically significant were found to be travel length and the number of transfers. The increase in travel length affects negatively the possible selection of the suburban rail opposed to the increase of the number of interchanges.

In the particular models the generic variable ΔT was used relating to the travel time difference between the bus and the suburban rail. The calculation of the odds ratio shows that increasing the difference in travel time by one unit increases the possibility of selecting the suburban rail instead of the bus by 54%. Bus users namely choose the transport mode with the least travel time for a specific movement.

Regarding the socio-economic characteristics of the respondents those who make the specific trip for education purposes are more likely to choose the suburban rail than the other modes, whereas those making the same trip several times during the week also choose to a greater extent than the others the suburban rail. Freelancers and retired persons are more likely to switch from the bus to the new suburban railway.

The calculation of the standardized coefficients of the models indicates that greater level of influence to the mode choice seems to have the travel time difference from the other variables of the models.

The statistical tests undertaken show a good fit of the extended model unlike the simplified one. Moreover, the introduction in the model variables related to the socio-economic characteristics of the sample not only improved the factor Nagelkerke R Square but also improved the adaptability of the model to the data.

4.4 Binary choice models for reduced fare bus users

In both models, simplified and extended, the constant term is statistically significant, meaning that the users of this category have a prior preference for the suburban rail against the bus, when time, cost and number of transfers are equal for both modes.

The only one of the independent variables that seems to have a statistically significant influence on the mode choice, and in particular negative, is the travel cost of the suburban as specified by the games. According to the results of both models, an increase in travel cost by 1% leads to a reduction of users probability to choose the suburban railway almost by 100%. The users of this category, because they travel with a very low fare, are particularly vulnerable to a possible change of travel cost.

The fact that someone is unemployed reduces the chances of choosing the new suburban service by 47%. This decrease can probably be explained by the fact that these are people who may be unemployed and therefore prefer the bus, because discounted fare is lower than that of the suburban rail in all games tested.

Based on the calculation of the standardized coefficients of the models the travel cost is three times more important in terms of mode choice than the employment status of the respondents.

The results of statistical tests performed indicate that the models fit well in the sample data. Also, the introduction of variables related to the socio-economic characteristics of the respondents improves the adaptation of the models to the data.

4.5 Travel cost and travel time elasticities

Having developed the above described set of models travel cost and travel time elasticities can be calculated. Elasticity is defined as the percentage change of a variable value associated with the change in the value of another variable by one percentage point.

The basic equation for calculating the elasticity at each binary choice model is the following:

$$E_{X_{iq}}^{P_{iq}} = \beta_i X_{iq} (1 - P_{iq})$$

(9)

From the above equation, it is observed that the elasticity gets a zero value when the probability P_{iq} becomes one, whereas when the corresponding probability becomes zero, the elasticity is equal to the product $\beta_i{}^* X_{iq}$ (Dunne [11]).

The above equation is useful for calculating direct elasticities. However, to calculate the aggregate elasticity a method known as 'sample enumeration method' was used, since the use of means of the values X_{ikq} of the sample or the medium probability P_{iq} is not appropriate for non-linear logistic regression models. In the sample enumeration approach, individual probability is first calculated and then aggregate by weighting each individual elasticity by the individual's estimated probability of choice. According to the method used, the aggregate elasticity is calculated with the following formula (Louviere et al. [12]):

$$E_{X_{jq}}^{\bar{P}_i} = \left(\sum_{q=1}^{Q} \hat{P}_{iq} E_{X_{jq}}^{P_{iq}} \right) \bigg/ \sum_{q=1}^{Q} \hat{P}_{iq}$$

(10)

where $\hat{P}_{iq}$ is an estimated choice probability and P_{iq} refers to the aggregate probability of choice probability i (Ben-Akiva & Lerman [8]).

The elasticity values estimated, using the model of logistic regression, enable the estimation of demand sensitivity to changes of various factors that affect it, in this case the time and travel cost, and thus allow for the evaluation of various policies that can be applied in the field of transport.

Table 5 presents the travel and time elasticities that were calculated for both simplified and extended models for all three users categories.

Table 5: Travel time and cost elasticity for all models.

Car users	Model 1	Model 2
Travel time elasticity (tt_car)	–0.354	–0.319
Travel cost elasticity (cost_train)	–0.583	–0.584
Regular-fare bus users	**Model 1**	**Model 2**
Travel time elasticity (tt_bus)	0.655	0.801
Travel time elasticity (tt_train)	–0.525	–0.606
Travel cost elasticity (cost_train)	–1.002	–0.863
Reduced fare bus users	**Model 1**	**Model 2**
Travel time elasticity		
Travel cost elasticity (cost_train)	–1.346	–1.413

From elasticities calculated for car users it comes out that both travel time by car and travel cost of the suburban rail are inelastic for the users of this category. Furthermore, it was found that car users are more sensitive to a possible change in the travel cost of the suburban rail with respect to a change in travel time by car. In the case of travel time it was found once again that in a possible increase in travel time users still choose their car instead of the suburban rail.

In the case of bus users paying a regular fare, users are affected most by the travel cost than by time. Regarding travel times of the two modes of transport, when the travel time by bus increases, the probability of selecting the suburban rail also increases unlike suburban rail in which case when the travel time increases the corresponding probability decreases. For bus users with reduced fare, travel cost of the suburban is elastic.

Comparatively, with regard to the travel cost of the suburban railway, demand is inelastic for car users and elastic for bus users. This means that the bus users are more affected by a change in cost, something expected for this user category where the travel cost is comparatively lower for car users. These results are consistent with the international literature (Litman [13]).

Finally, the inclusion of the user-specific attributes in the models, although it seems that it does not have a great effect on the magnitude of the elasticity, causes a relative increase in the travel time elasticities (e.g. the travel time elasticities are higher in most cases for model 2) and a relative decrease in the cost elasticities (e.g. the travel cost elasticities are lower for model 2 with the exception of bus users with reduced fare ticket).

5 CONCLUSIONS AND FURTHER RESEARCH

The research undertaken and presented in this article was focused on the likely effects that socio-economic characteristics of potential users of a new suburban rail link may have on mode choice and hence on expected demand of that new service. The investigation is based on data collected by a revealed and stated preference conducted in the city of Thessaloniki in Greece and more specifically at locations along the new rail link. The survey sample was stratified according to the current mode used in three categories, namely car users, bus users paying regular fare and bus users paying reduced fare (concession fares). Twelve binary choice models were developed for the purposes of this research using the MLE method, from which two were simplified and two were extended per user category.

The trip and user characteristics examined include travel time and cost, number of transfers, gender, occupation, trip frequency, etc.

The binary choice models developed for car users indicate that these trip makers insist on using their car even if travel times increase in relation to the travel time by the future rail service. Comfort seems to play a very strong role to the car users. Furthermore, the fact that they may have to suffer at least one transfer works in a negative way as expected.

In the case of bus users paying the regular fare, the results indicate that a likely increase of travel length and time has a deterring effect in choosing the suburban rail. This result may be explained on the grounds that the proposed suburban line covers a certain part of the total journey of users while the remaining part is served by the existing bus lines. Therefore, when the travel length increases users are not willing to switch to the suburban railway to cover a part of their entire journey.

Another finding is that the model standardized coefficients for all three categories show that the mode-specific and generic variables have more influence on mode selection than the variables related to the socio-economic characteristics of the respondents. Equally interesting findings emerge from the calculation of the travel time and cost elasticity values for the three models as well as from the expected travel behaviour of certain user subcategories such as the freelancers and the unemployed.

In all models developed, the inclusion of the users' socio-economic characteristics had an additive effect on their explanatory power. In statistical terms all user characteristics were proven significant. In this respect, using the extended models instead of the simplified ones seems to be the right choice. Apparently the specific findings are strongly related to the functional characteristics of the specific suburban rail case. However, we dare to say that even in other cases, the user characteristics will be proven valuable in estimating mode choice and future demand.

The model application for new service planning purposes requires the availability of such data for the population in the study area. This is a difficult task especially if no periodical transport surveys or census on household and trip characteristics take place. The present research in fact is a way to stress the need to obtain such data and information on a systematic way.

The rather small sample used for the purposes of this research poses definitely certain limitations. Although 500 questionnaires were initially collected, the valid ones were substantially reduced in the analysis phase since a lot of them corresponded to users who were already using the existing suburban rail services.

In the particular stated preference survey, a series of three games was presented to the respondents each time. Respondents were asked to answer each time whether they would switch to the new suburban rail. This technique, although it helps to explore the willingness of users to choose between different means of transport, does not permit the determination of the alternative that is expected to generate the largest utility for the user. Namely, under what conditions of travel time and cost the user is more willing to choose the new transport mode.

Other issues to be considered in a similar investigation could be a different segmentation of the affected population by trip purpose, trip length, or even specific income levels. Finally, the use of latent factors would be quite useful to better understand attitudes and eventually travel behaviour of the potential users.

REFERENCES

[1] Papaioannou, P., Basbas, S., & Vougioukas, M., The use of stated preference technique in evaluating a passenger information system: The EuroBus/Popins/Thepis experience. Public Transport Planning and Operations. *Proceedings of Seminar, Held at the PTRC European Transport Forum*, Brunel University: England, 1996.

[2] Basbas S., Stated preference surveys and the valuation of urban transport systems. *Proceedings of the 14th International Conference: Urban Transport and the Environment in the 21st Century – Urban Transport XIV*, ed. C.A. Brebbia, Vol. 101, Wessex Institute of Technology, *Transactions on the Built Environment*, Malta, 1–3 September 2008, pp. 3–14.

[3] Hensher, D.A. & Button, J., *Handbook of Transport Modeling*, 3rd ed., Pergamon: Sydney, 2000.

[4] TREDIT S.A., *Users' Survey Design of the Services of TRAINOSE, Education of Researchers and Data Analysis for the Project RAIL4SEE – Cities of Southeast Europe as Rail Links*, Thessaloniki, 2014.

[5] Nikolaidou, N., *Development of Binary Choice models for the case of a Suburban Rail In the City of Thessaloniki, Greece*, Master Thesis, Interdepartmental Course 'Design, Planning and Organization of Transport Systems', Supervisor: Prof. Panos Papaioannou, Aristotle University of Thessaloniki, Greece, 2014.

[6] Dyer, C., *Research in Psychology: A Practical Guide to Methods and Statistics*, Blackwell Publishing: Hodoken, NJ, 2006.

[7] Basbas, S., Tetou, V. & Politis, I., Ordinal and binary logistic logit models for examination of behavioral, infrastructure and perception factors influencing biking. *Proceedings of the 19th International Conference: Urban Transport and the Environment in the 21st Century – Urban Transport XIX*, ed. C.A. Brebbia, Vol. 130, Wessex Institute of Technology, *WIT Transactions on the Built Environment*, Kos, Greece, 29–31 May 2013, pp. 573–584.

[8] Ben-Akiva, M.E. & Lerman, S.R., *Discrete Choice Analysis: Theory and Application to Travel Demand,* Vol. 9. MIT press, 1985.

[9] Shumacher, R.E. & Lomax, R.G., *A Beginner's Guide to Structural Equation Modeling*, 2nd ed. London: Lawrence Erlbaum Associates, 2004.

[10] King, J.E., Standardized coefficients in logistic regression. *Paper Presented at the Annual Meeting of the Southwest Educational Research Association*, San Antonio, TX, 2007.

[11] Dunne, J.P., Elasticity measures and disaggregate choice models. *Journal of Transport Economics and Policy*, **XV111(2)**, pp. 189–197, 1984.

[12] Louviere, J.J., Hensher, D.A. & Swait, J., *Stated Choice Methods,* Cambridge: Cambridge University Press, 2000

[13] Litman, T., *Understanding Transport Demands and Elasticities: How Prices and Other Factors Affect Travel Behavior.* Canada: Victoria Transport Policy Institute, 2013.