Available online at www.sciencedirect.com
ScienceDirect
Procedia - Social and Behavioral Sciences 160 (2014) 499 – 508
Procedia
Social and Behavioral Sciences

XI Congreso de Ingenieria del Transporte (CIT 2014)

Methodology for a study of the perceived quality of public transport in Santander

Valeria Maraglino[a], Luigi dell'Olio[b]*, Dino Borri[a], Angel Ibeas Portilla[b]

[a]*Polytechnic University of Bari, Via Edoardo Orabona 4, 70126 Bari, Italy*
[b]*University of Cantabria, Av. De los Castros s-n, Santander 39005, Spain*

Abstract

This article will primarily deal with improving the quality of a public transport system through the study of the main variables that influence the perceived quality by users, in other words a methodology for modeling the quality of bus services in the Spanish city of Santander, supported by user perception data.
The models calibrated to study the quality of public transport perceived by users are firstly obtained by estimating all the different service attributes, considering mean users' perceptions through an Ordered Probit model, and then studying random variations in users' tastes, applying an Ordered Probit model with random parameters. The models represent the process of quality evaluation based on a limited group of predefined variables. The choice of these variables is important because they are used to explain the selection process to be modeled. The collected data are analyzed and modeled to check the validity of the different variables. Main results suggest consumers would be willing to pay a higher price for improved transport connection networks including the supply of clearer information and the attention to consumer's complaints. Ticket price is also an important attribute mainly for frequent bus users.

Peer-review under responsibility of CIT 2014.

Keywords: Consumer behaviour, Choice models, Ordered Probit Model, Ordered Probit with Random Parameters Model, Public transport system.

* Corresponding author. Tel.: +34-942-202262; fax: +34-942-201703.
E-mail address: delloliol@unican.es

Peer-review under responsibility of CIT 2014.
doi:10.1016/j.sbspro.2014.12.163

1. Introduction

Nowadays, ever greater efforts are made by municipalities to improve public transport systems, in order to achieve sustainable mobility and ensure more healthful environments in our cities.

In order to understand how public transport must be to meet users' needs, it is necessary to define the most important variables determining system quality, from among the many that contribute to the user's perceptions.

It is important when planning market strategies, focused on the development of public transport in urban areas, to understand what users' satisfaction levels depend upon, as well as their expectations for an efficient public system (Bordagaray et al., 2012, 201_, dell'Olio et al. 2011b, Rojo et al. 2012, 2013). An adequate planning strategy should consider the spatial location of bus stops, the re-organization of bus lines, their timing and frequencies across the urban circuit.

Hence the great importance of determining quality in its two different forms in this context: users' *desired* and *perceived quality*. The latter is defined by users' degree of satisfaction with variables that define how public transport works. *Desired quality* (dell'Olio et al., 2011a) differs from *perceived quality* because it does not represent what users feel about daily transport but what they want or would like to find when they travel. In short, it embodies what they envisage as an efficient system (dell'Olio et al., 2010). This article is devoted to an analysis of perceived quality.

In addition, potential users' characteristics have to be considered when planning development policies. The results of assessment of different variables that define public transport quality may depend on these characteristics. The problem is that customers' preferences are uncertain and variable, so the most appropriate model, that estimates users' behavior in regard to perceived quality, needs to be found. It is essential, therefore, to collect information about the characteristics that define the journey and also about how users estimate them.

The paper is structured as follows: Section 2 describes the theoretical approach used in our study; Section 3 discusses into details the case study applied to the public transport (buses) of Santander (Spain) and illustrates the obtained results; and finally, Section 4 summarizes and concludes.

2. Methodology

This section describes the methodology employed in our case study. We illustrate straightforward the features of the sample and finally present the theoretical analysis explaining the choice models used to run our inferential statistic analysis.

Sample. The definition of the sample size is a necessary step before implementing an econometric model. Because of time and cost constraints, it is clearly impossible to cover the entire population, and reach them with a direct survey (Fortini, 2000). For this reason, in statistical methods a random sample of the population is selected, and taken to reflect the entire population. However, in such cases both sampling error, due to working with a sample rather than the entire population, and selection error, due to possible bias even in the case of a random sample, have to be taken into account. Both can generate changes in the results of an investigation. The most commonly used method is the probabilistic one, extracting units from the population in such a way that each element has a known probability of becoming a part of the sample (Marcucci, 2011). Data collection for the survey is carried out by studying users' actual behavior (revealed preference methods), rather than conducting sample surveys, experiments or simulating markets where preferences are collected in a hypothetical choice context (stated preference methods).

Discrete choice models. After concluding the previous steps, the discrete choice models need to be applied, to determine the demand for assets and services, as well as make predictions about users' behavior. They play a very important role in the process of transport modelling and, if properly calibrated, they allow you to make predictions about future states of the studied system, and to intervene in the process of possible development in order to

optimize their function. In our case, perceived quality is related to the concept of utility to the customer, who takes on an absolutely central role (Greene, Hensher, 2010).
The hypothesis is that the greater the utility associated with a service, the higher the probability that a person will choose that one (Marcucci, Gatta, 2007).

Functional form and type of model. To estimate our model we use an Ordered Probit with Random Parameters model. They have been used in a range of applications associated with data arranged in rankings, qualifications or levels (McKelvey, Zavoina, 1975). Generally, an ordered model has a regression format in which the dependent variable y* is a linear function of a group of independent variables x_i and a random term ε.

$$y_i^* = \beta' x_i + \varepsilon_i \tag{1}$$

with $\varepsilon_i \sim F(\varepsilon_i|\Theta)$, and $E[\varepsilon_i] = 0, Var[\varepsilon_i] = 1$. The variable y* can be discretized using:

$$\begin{gathered} y = 0,\ if\ \mu_{-1} < y_i^* \le \mu_0 \\ y = 1,\ if\ \mu_0 < y_i^* \le \mu_1 \\ y = 2,\ if\ \mu_1 < y_i^* \le \mu_2 \\ \ldots \\ y = J,\ if\ \mu_{J-1} < y_i^* \le \mu_J \end{gathered} \tag{2}$$

The model estimates β' and μ. β' are the weights associated with each explanatory variable and represent the importance of each one to the dependent variable. The parameters μ are the limits which define the variable y.
The model assumes a random term ε, which represents the error, with zero mean and a unitary variance.
To obtain the calibration of the model the value of the first answer y corresponds to zero, the lowest threshold parameter corresponds to -∞ and the highest to +∞. Finally, μ_0 is equal to zero.
In the survey, the value scale attributed features the following options: “very bad”, “bad”, “neither good nor bad”, “good”, “very good”. The model works with an ordinal scale and for its estimation, there need to be replies to all the questions. The negative scores “very bad” and “bad” had to be grouped together in the same ordinal numerical category. Considering equation (1), the model applied has the following structure:

$$\begin{gathered} y = 0, (Very\,bad\,or\,bad), if\ -\infty < y_i^* \le 0 \\ y = 1, (Neither\,good\,nor\,bad), if\ 0 < y_i^* \le \mu_1 \\ y = 2, (Good), if\ \mu_1 < y_i^* \le \mu_2 \\ y = 3, (Very\,good), if\ \mu_2 < y_i^* \le -\infty \end{gathered} \tag{3}$$

The results estimate the probability of observing each result of y = 0, 1, 2, 3, a characteristic which does not work with probabilities, but directly estimates a mean value of the dependent variable based on the observed values.
The probability associated with the observed results is:

$$P\left[y_i = jx_i\right] = P\left[\varepsilon_i < \mu_1 - \beta' x_i\right] - P\left[\mu_{j-1} - \beta' x_i\right],\ j = 0,1,2,3 \tag{4}$$

In the Ordered Probit models the random component ε_i assumes a Normal distribution.
The probability functions associated with each y_i is:

$$P\left[y_i = jx_i\right] = F\left[\mu_1 - \beta' x_i\right] - F\left[\mu_{j-1} - \beta' x_i\right] > 0, \quad j = 0,1,2,3 \tag{5}$$

The parameters are obtained thanks to the process of maximum likelihood. The optimization is obtained with the log likelihood function, which is the logarithm of the probability expression above:

$$\log L = \sum_{i=1}^{n} \sum_{j=0}^{j} m_{ij} \log\left[F\left(\mu_j - \beta' x_i\right) - F\left(\mu_{j-1} - \beta' x_i\right)\right] \tag{6}$$

where mij=1 if yi=j, and 0 in other cases.

The survey was made of Revealed Preferences, consisting of the replies to a series of questions about users' experiences with the existing system. The users scored aspects of the general service they received. The use of discrete choice models allows interactions to be introduced which can explain the users' different perceptions, originating from socioeconomic factors (Ortúzar, Willumsen, 2011).
The models calibrated for modeling users' perceived quality of a public transport represent the process of evaluating this quality on the basis of a limited group of predefined variables x_i collected from each of the interviewed users (Marcucci, 2011). The choice of these variables is important because they are used to explain the selection process to be modeled. The collected data are then analyzed and modeled to check the validity of the different variables.

3. Results and discussion

This section illustrates the case study proposed in this work. The goal of the project, therefore, is to propose, develop and analyse methods for modelling the perceived quality of public transport, namely Line 1 of the TUS (Transportes Urbanos de Santander) in Santander, capital of the autonomous community of Cantabria situated in the north coast of Spain.

Santander is a medium-sized city, that covers $36km^2$, with a population of approximately 200,000 inhabitants. Mobility to the north and south of the city is bounded by steep slopes (greater than 15°), because of the hills and valleys that run parallel from North-East to South-West.

Santander's public transport system covers 100% of the municipal area, and for 97% of the route there are stops located at less than 300m distance apart.

For our study, we made four different surveys over a period of three years (2009, 2010, January and December 2011) to examine changes in users' quality perceptions. When choosing the model, it is crucial to take into account the uncertainty of customers' preferences. Failure to consider this randomness can produce inaccurate results and so an untruthful model.

Although different types of models exist, in this project we decided to use the Ordered Probit to consider mean users' perceptions and then to study random variations in users' tastes, with an Ordered Probit with random parameters.

The variables employed to estimate the models with their corresponding meanings are listed below:

Table 1.Variables used in our case study

Main Variables	Description of main variables	Other variables	Description of other variables
LINEA	Indicates the number of the bus line	E3_11	Service information at bus stops
PERIODO	Indicates when the survey was performed	E3_12	Information about available services at TUS offices
SEXO	Indicates the user's gender	E3_13	Information about available services on the web
ESTUDIAN	A student	E4_14	Waiting time at bus stops
TRABAJA	A worker	E4_15	Time to walk to the nearest bus stop
NOTRABAJA	Unemployed	E4_16	Travel time
JUBILADO	Retired	E4_17	Security of TUS
E24	Under 24 years old	E5_18	Courtesy of drivers
E2534	Aged between 25 and 34	E5_19	Attention to the customer asking for information /making claims by phone
E3544	Aged between 35 and 44	E5_20	Attention to the customer asking for information /making claims in person
E4554	Aged between 45 and 54	E5_21	TUS response to complaints
E5564	Aged between 55 and 64	E6_22	Cleanliness and hygiene conditions on the bus
E65	Over 65	E6_23	Heating and air conditioning
FR5	Travel frequency < 5 times per week	E6_24	Comfort
FR5_15	Travel frequency between 5 and 15 times	E6_25	Degree of crowding
FR15_30	Travel frequency between 15 and 30 times	E7_26	Professional skills of bus driver
FR30	Travel frequency is more than 30 times per week	E7_27	Safety during the travel
E1_1	Service provided (frequency)	E8_28	Installation of biodiesel
E1_2	Coverage of different areas	E8_29	Noise pollution
E1_3	Night service / weekend service	E1	Service offered
E1_4	Special service for sporting events, concerts, etc	E2	Accessibility
E1_5	Ticket price	E3	Information
E1_6	System of payment by card	E4	Time
E2_7	Accessibility for people with reduced mobility	E5	Attention paid to the customer
E2_8	Connection with other types of transport	E6	Comfort
E2_9	Facilities for carrying cumbersome baggage, surfing tables, etc…	E7	Safety
E3_10	Information on bus monitor	E8	Environmental impact

After importing all these variables, concerning characteristics of users and buses, we used the software Nlogit version XX to estimate our models.

Ordered Probit model. To obtain a correct estimation of the model it is necessary to omit the variables with negative parameter values and those which are not statistically significant at 95% C.I. level. In particular, the following variables: i) Night service/weekend service; ii) System of payment by card; iii) Information about available services at TUS offices; iv) Level of satisfaction of customer service for claims and information by phone; were removed. Also, interaction terms between consumer choices and other socio-economic variables (e.g. sex, age, frequency) are added to capture market segmentation (due to differences in consumer preferences) in determining the perceived quality of the bus service.

As an example, the variable E8_29F30 represents how consumers who travel more than 30 times per week – F30 – perceive noise pollution – E8_29– as an important attribute in Table 2. The sign of the estimated coefficient of this variable is negative. This tells us that for this type of travelers the perception of noise pollution is a minor problem. The more one travels, the more he/she becomes accustomed with the level of noise pollution. In a similar way we can discuss the variables E2_8F55 (how consumers who travel between 5 and 30 times per week perceives the availability of connections with other transport modes as an important attribute) and E2_8F13 (how workers and unemployed perceive the quality of a face-to-face customer service as an important attribute). According to these types of travelers the presence of the above mentioned attributes do not appear, on average, important in determining the quality of bus service.

Table 2.Estimated results of the Order Probability Model

Attribute	Coefficient	t-ratio
Constant	-6.66	-7.99
E1_1	0.19	2.50
E1_2	0.34	4.83
E1_4	0.22	2.76
E1_5	0.21	4.14
E2_7	0.22	3.34
E2_8	0.43	5.48
E2_9	0.24	3.93
E3_10	0.16	2.51
E3_11	0.12	1.91
E3_13	0.21	2.61
E4_14	0.27	4.13
E4_15	0.26	4.08
E4_16	0.15	2.44
E4_17	0.34	4.58
E5_18	0.45	6.07
E5_20	0.54	4.44

E5_21	0.21	1.66
E6_22	0.18	2.21
E6_23	0.16	2.43
E6_24	0.30	3.35
E6_25	0.36	5.81
E7_26	0.20	2.36
E7_27	0.47	6.49
E8_28	0.43	6.06
E8_29	0.32	4.94
E2_8F55	-0.13	-2.83
E2_8F13	-0.08	-1.44
E8_29F30	-0.20	-1.74
E2_8TRA	-0.08	-1.90
E5_20NTR	-0.15	-1.90
Threshold parameters		
μ_1	2.81	4.13
μ_2	11.21	15.71
μ_3	15.76	19.01
Log likelihood function	-313.16478	
Number of observations	52	

Ordered Probit model with Random Parameters. To better infer on consumer behavior and capture his/her uncertainty levels and the changes in consumer preferences over time, it is necessary to consider the use of an Ordered Probit with Random Parameters. The use of an Ordered Probit with Random Parameters gives us a better log-likelihood value (-306.805) compared to that obtained in the Ordered Probit model (-313.164). Furthermore, consumer preferences are estimated according to a specific probability distribution function (e.g. Gaussian). This means that we have consumers for which their estimated preference value is near to the mean value of the probability distribution function, and other consumers for which their estimated preference value is far from the mean value of the Gaussian function (i.e. some values can reach the extremes of the function).

To improve the models' estimate we use a stepwise method (this method allows the software to select the model with statistically significant variables at 95% C.I) during the iteration process.

Table 3 shows the estimated results of the Ordered Probit with Random Parameters Model.

Table 3.Estimated results of the Order Probit with Random Parameter Model

Attribute	Coefficient	t-ratio
Constant	-13.95	-10.27
E1_1	0.85	6.35
E1_4	0.51	4.10
E2_7	0.38	3.52
E2_8	0.94	6.93
E2_9	0.49	5.25
E3_11	0.31	3.02
E3_13	0.36	3.02
E4_14	0.54	5.20
E4_15	0.57	5.43
E4_16	0.26	2.89
E4_17	0.69	5.75
E5_21	0.51	2.61
E6_22	0.35	2.79
E6_23	0.31	3.10
E6_24	0.65	4.75
E6_25	0.75	7.02
E7_26	0.32	2.70
E7_27	0.95	7.70
E8_28	0.83	7.07
E8_29	0.63	6.36
E2_8TRA	-0.44	-2.49
E5_20NTR	-0.27	-2.26
Means for random parameters		
E2_8F55	-0.24	-3.22
E2_8F13	-0.18	-2.09
E2_8TRA	-0.13	-2.01
E1_1	0.37	3.51
E1_5	0.43	5.17
E3_10	0.36	3.69
E5_18	1.06	8.07
E5_20	1.04	5.37
Scale parameters for random parameters		
E2_8F55	0.34	6.49
E2_8F13	0.48	6.85
E2_8TRA	0.38	7.91
E1_1	0.07	2.39
E1_5	0.25	6.90
E3_10	0.33	9.02

E5_18	0.10	3.56
E5_20	0.10	2.81
Threshold parameters		
μ_1	5.8	5.50
μ_2	22.41	11.56
μ_3	31.87	12.29
Log likelihood function	-306.80531	
Number of observations	64	

From Table 3 above, we can notice that we also used interaction terms (consider for example, E2_8F55, E2_8TRA) to estimate the Ordered Probit with Random Parameters.

During the interaction process with stepwise method, the software considers the observations in which that particular attribute is present and 0 otherwise. By doing so, the estimated variables for the interaction terms would be statistically significant for the group of travelers having that particular selected attribute. The positive sign of the interaction terms tells us that for a given group of travelers, the variable under exam expresses an important attribute according to the preferences of each member of that group, in determining the quality of the bus service.

For non-interaction terms (e.g., E1_1, E1_5), the estimated coefficients represent how the attribute under exam would be relevant for the entire sample.

4. Conclusions

To develop policies, strategies and plans focused on improving urban public transport it is crucial to know which variables influence the quality of the service.

This paper provides a further contribution to the empirical literature on estimating the value of perceived quality by analyzing the case of TUS (TransportesUrbanos de Santander).

Socio-economic analysis has been conducted with some choice models (i.e. Ordered Probit and Ordered Probit with Random Parameter Models) and the presence of a hierarchical structure among the attributes considered has been verified. The performance of the models improves when random parameters are introduced, as the log-likelihood function shows.

The analysis also shows that some users are not concerned with the ease of connection with other transport systems, particularly categories of users such as workers and those who use buses from 5 to 30 times a week. Instead, they attribute a considerable importance to the services offered (e.g. timetables, frequencies, ...), the tickets price, courtesy of drivers, the attention paid to the customerand the information displayed on monitors (on bus only for the ordered probit with random parameter model; and on bus and at bus stops for the ordered probit model).

Ultimately, the results produced by this application, although they refer only to a sample of the whole population, could be usefully employed in future planning of public transport system, highlighting those variables characterizing the service that are perceived as important by users to guarantee and secure an efficient and sustainable mobility.

Acknowledgements

This work was made possible thanks to a grant from the Spanish Ministerio de Economía y Competitividad with additional financing from FEDER funds to the project TRA2012-37659.

References

Bordagaray, M., Ibeas, Á., & dell'Olio, L. (2012). Modeling User Perception of Public Bicycle Services. *Procedia - Social and Behavioral Sciences*, 54, 1308–1316.

Bordagaray, M., dell'Olio, L., Ibeas, A., & Cecín, P. (201_). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica a: Transport Science*, *0*(0), 1–17. doi:10.1080/23249935.2013.823579

dell'Olio, L., Ibeas, A., & Cecin, P. (2010). Modelling user perception of bus transit quality. *Transport Policy*, 17(6), 388–397.

dell'Olio, L., Ibeas, A., & Cecin, P. (2011a). The quality of service desired by public transport users. *Transport Policy*, 18(1), 217–227.

dell'Olio, L., Ibeas, A., Cecin, P., & dell'Olio, F. (2011b). Willingness to pay for improving service quality in a multimodal area. *Transportation Research Part C-Emerging Technologies*, *19*(6), 1060–1070.

Fortini, M., (2000), Linee guida metodologiche per rilevazioni statistiche. ISTAT, Roma

Greene, W. H., and Hensher, D. A. (2010). Modeling ordered choices: a primer. *New York: Cambridge University Press.*

Marcucci E., Gatta V., (2007). Modelli a scelta discreta per il benchmarking della qualità nel trasporto pubblico locale, *I trasporti e il mercato globale, Franco Angeli*.

Marcucci E., (2011). Scelte di trasporto e modelli a scelta discreta, *CollanaEconomia e Politica Industriale*, Franco Angeli, Milano.

McElvey, R. and Zavoina, W. (1975) A Statistical Model for the Analysis of Ordered Level Dependent Variables. *Journal of Mathematical Sociology*, 4, pp. 103-120.

Ortúzar, J. D., Willumsen, L. G., (2011). Modeling Transport, 4th edition. *John Wiley and Sons*

Rojo, M., Gonzalo-Orden, H., dell'Olio, L., & Ibeas, Á. (2012). Relationship between service quality and demand for inter-urban buses. *Transportation Research Part a-Policy and Practice*, *46*(10), 1716–1729.

Rojo, M., dell'Olio, L., Gonzalo-Orden, H., & Ibeas, A. (2013). Interurban bus service quality from the users' viewpoint. *Transportation Planning and Technology*, *36*(7), 599–616.