sustainability

MDPI

Article

Weighted Variables Using Best-Worst Scaling in Ordered Logit Models for Public Transit Satisfaction

Iván Manuel Mendoza-Arango [1], Eneko Echaniz [2], Luigi dell'Olio [2,*] and Eduardo Gutiérrez-González [1]

[1] Instituto Politécnico Nacional, Sección de Estudios de Posgrado e Investigación, Av. Te 950 Alcaldía Iztacalco Col. Granjas Mexico, CDMX 08400, Mexico; ivanmendozarango@gmail.com (I.M.M.-A.); egutierrezg@ipn.mx (E.G.-G.)

[2] Sustainable Mobility and Railways Engineering (SUM+LAB), Department of Transportation, University of Cantabria, Av. de Los Castros 44, 39005 Santander, Cantabria, Spain; echanize@unican.es

* Correspondence: delloliol@unican.es; Tel.: +34-942-201-566

check for updates

Abstract: Customer overall satisfaction regarding a public transport system is dependent on the satisfaction of the users with the attributes that make up the service, as well as the contribution that each of these attributes makes to explain the overall satisfaction. A common way of analysing the contribution of service attributes to explain overall satisfaction is through the use of ordered logit or probit models. This article presents an ordered logit model that considers the weighting of independent variables through the explicit importance calculated on the basis of a best-worst case 1 choice task. For the calculation of importance, a multinomial logit model has been estimated which considers the heterogeneity of the sample through systematic variations in user tastes. In this way, it is possible to establish a level of importance of each specific attribute for each type of user. The results show that the importance varies considerably depending on different socio-economic and mobility-base variables. On the other hand, the inclusion of the weighted variables in the ordered logit model improves its fit. Therefore, the results make possible to develop policies focused on improving satisfaction on specific user targets.

Keywords: user satisfaction; public transport; best-worst scaling; ordered logit; discrete choice

1. Introduction

The satisfaction with a public transport system is conditioned by the satisfaction about the different attributes that compose it. In the same way, the importance that users place on the attributes also affects their perception. Several studies [1–3] have shown that the demand associated to a public transport system is directly linked to users' satisfaction. Therefore, identifying the factors that most affect satisfaction is essential for developing policies focused on improving the service and increasing patronage.

The usual way to obtain these key attributes or factors is through the use of models. Overall satisfaction is usually represented as the dependent variable, which is estimated from the service attribute ratings. The most common methods used in the last decade have been structural equations and discrete choice models. In the use of structural equations, the works [4–10] stand out. This method has proven to be very useful for understanding the different mechanisms that explain users' satisfaction. Structural equations have been applied in different cases such as road base public transport systems [11–15], as in other transport modes [6,8–10,16]. However, there are other methods in the literature that have also proven to be reliable and useful for analysing quality in public transport systems.

These include the use of decision trees [17–21], multi-criteria analysis [22,23], factor analysis [24–27], neural network methods [28–30] or multiple regression methods [14,31,32].

As for discrete choice models, more specifically ordered data models, they have been widely used in the field of quality in public transport. The studies [28,33–37] are a good example of how ordered models can be used for analysing user satisfaction on public transport. One of the advantages of ordered data models is the possibility of capturing non-linearity between different choices. This corresponds to the reality of the applications, since it has been shown that it is more costly to improve the satisfaction towards an initially well rated service than to improve a service which is considered to be bad or deficient. The nonlinearity of the dependent variable is defined by the threshold parameters estimated in the model. Another way based on a discrete choice model has been to conduct stated preference (SP) surveys, which show the respondents a number of choice tasks and ask them to choose the preferred one. The SP data is analysed using different kinds of discrete choice models. For example, Román et al. [38] used the multinomial logit model (MNL) and a mixed logit model (ML) to examine public transport services in Gran Canaria (Spain). Similarly, dell'Olio et al. [39] used an MNL model to analyse the quality desired by future users.

Various studies [31,33,35,39,40] have shown that the socio-economic variables of the respondents affect the perception of quality in public transport systems. Those variables are mainly gender, age, work situation and trip purpose. Similarly, [39] and [33] show that the specific characteristics of the different lines that make up a public transport system also condition users' perception of quality. Also, in the study developed by Krueger et al. [41] it was shown that individual's perception of the beliefs of others regarding a specific behaviour can be linked to mobility choices. Proving that the attitudes towards a specific transport mode are influenced by both, the normative beliefs and socioeconomic characteristics.

In the work of Echaniz et al. [42] several ordered probit models were developed in which the systematic and random heterogeneity of users was considered and the variables were weighted according to the importance of the attributes. The weighting of the variables was calculated using the results of ranking based questions. Where respondents had to order some variables within different groups from the most important to the least important. The final conclusion of this study showed that the perception of the overall satisfaction of a public transport service was significantly affected by the different characteristics of the users. However, the importance of the attributes did not turn out to be completely relevant when it came to improving the fit of the models. Moreover, it was observed that the ranking method used to obtain attribute importance levels was very time consuming, requiring a high cognitive effort of the respondents. The study developed in this paper intends to take a further step in the research carried out by Echaniz et al. [42] by weighting the variables of an ordered model by using the importance levels obtained through a Best-Worst type questions. Being this type of questions faster and more cost-effective than ranking based questions.

Best-worst (BW) [43] questions are based on showing the respondent a set of choices in which the respondent must choose the best and worst option. The use of this type of question in the field of transport is quite limited [44–46], even more so in the specific case of studying quality on public transport. In this last field, the work of Beck and Rose [47] and that of Echaniz et al. [48] stand out, in both cases it was established that the importance derived from the BW questions was more similar to reality compared to the more traditional method. Beck and Rose [47] do not consider the heterogeneity of the sample when defining the importance of the different attributes of the system, obtaining constant values for all types of users. Echaniz et al. [48] consider the heterogeneity of the sample by applying a mixed logit model with random parameters when analysing BW data. In this case, it was demonstrated that there is variability in the sample, however, when applying random parameters this variability is not related to any objective factor. Therefore, it is not possible to establish the drivers that affect the variation on importance levels. To study this fact in more detail, this paper develops a BW-based model that considers the systematic variations of the sample, defining those variables that most influence the variation of the perception of importance.

To sum up, the contribution of this article to the existing literature focuses on further developing the studies carried out in Echaniz et al. [42] and Echaniz et al. [48]. First, the BW data is modelled by considering the heterogeneity of the sample. Which serves to analyse the importance of different attributes that make up a public transport service depending on the type of user. That way, the random variation found in Echaniz et al. [48] is explained through systematic, observable variables. In turn, these modelling results are combined with the ordered logit model, to improve the model's fit and to check whether the importance of the attributes affects the modelling of overall user satisfaction. Which answers one of the questions left open by that study.

The remainder of this paper is structured in to four sections. The next section describes how the data was collected and develops a short descriptive analysis of the sample. Section 3 describes the data, the survey design and the sample. Section 3 explains the modelling method used, with estimation results presented in Section 4. The paper ends with the conclusions of the main findings, with some directions for future study.

2. Data

2.1. Survey

In order to obtain the necessary data for the estimation of the models, a survey divided into two parts was proposed. In the first part, information was collected regarding the socioeconomic characteristics of the respondent and their mobility preferences, from each of them the following were obtained: gender, age, employment situation, alternative mode of transport to make the same trip, trip purpose (specifying the purpose for origin and destination), bus usage level (specifying the number of uses per week) and finally, income level, defining the monthly salary. Of all these questions, the one related to the salary was kept as optional due to the fact that it is usually sensitive for certain users. The second part of the survey consisted of a series of questions associated with the perceived quality for a set of 24 service attributes: Use of hybrid buses, access time to bus stop, egress time from alighting stop to destination, vehicle cleanliness, ease of transfer, information at stops, information on board, comfort of the buses, service reliability/punctuality, driver's kindness, quality of bus stops, information on mobile app, line coverage, information on the web page, priority seats for people with reduced mobility (PRM), waiting time, crowding level, readability of map design, in-vehicle travel time, service frequency and timetables, driving style, price/fare, heating/air conditioning and noise. In turn, the satisfaction questions were divided into two different exercises. Each respondent had to answer a total of three questions. In each of the questions, a subset of four out of the 24 previously listed attributes was displayed. These sets were randomly generated, with the restriction that an attribute could not be shown more than once to each respondent and that the sample had to be balanced across the entire sample. For each set, respondents had to evaluate the attributes shown on a 5-point Likert scale, from "very bad" to "very good". Once the attributes were evaluated, a choice was made based on BW case 1 type, where the respondent had to choose the attributes that he or she considered to be the most important and the least important of those shown in the subset. This question, composed of the rating and the BW exercise, was repeated three times in each survey, once for each subset. At the end of the survey, all respondents were asked to rate the service as a whole, defined as the overall satisfaction. The overall satisfaction was obtained following the same 5-point rating scale used for the attributes.

2.2. Sample

The surveys were run between October and November 2017 in the city of Santander (Spain). The surveys were conducted face-to-face on 4 urban bus lines (L1, L2, L3 and L13) operated by the city's public entity. A total of 808 surveys were carried out according to the demand of passengers on the lines, obtaining a slightly larger samples for lines 1 and 2 which move a slightly higher number of passengers.

According to the socioeconomic information collected in the survey (Table 1), the sample is composed of a larger number of women 67%, this segment may be slightly overrepresented, although a larger number of female users of the service were observed. With regard to the age range, 25% of the respondents were young people under 25 years of age while the proportion of the rest of the age groups remained close to 15% in all cases. For the older age group, it was decided to make a differentiation between users between 65 and 75 and those over 75, since the average age of the city centre is increasing and requires more specific analysis. Almost half of the sample (47%) was working, while 25% of them were students. Retirees or pensioners represent 17% of the sample. As an alternative mode of transport, the respondents had available mainly the private vehicle, almost 50% considering the option of "driving" and "accompanying". On the other hand, the respondents did not have any other mode of motorized transport to make the same trip, with 47% of the users choosing the option "other", which mainly referred to making the same trip on foot. Therefore, a large part of the users in the sample can be considered captive users of public transport.

Table 1. Socioeconomic attributes of the sample.

Group	Variable	Acronym	Percentages	
Gender	Male	M	33%	
	Female	F	67%	
Age	<25	A1	25%	
	25–34	A2	14%	
	35–44	A3	15%	
	45–54	A4	17%	
	55–64	A5	15%	
	65–75	A6	11%	
	>75	A7	4%	
Work status	Housekeeper	W1	5%	
	Employee	W2	47%	
	Retired	W3	17%	
	Student	W4	24%	
	Unemployed	W5	8%	
Other available transport systems	Car (driving)	AT1	35%	
	Car (accompanying)	AT2	12%	
	Bike	AT3	6%	
	Motorcycle	AT4	3%	
	Other (mainly walking)	AT5	44%	
			Origin:	Destination:
Trip purpose	Home	P1	46%	29%
	Work	P2	22%	25%
	Studies	P3	9%	13%
	Health	P4	4%	5%
	Shopping	P5	5%	7%
	Leisure	P6	10%	13%
	Other	P7	5%	9%
Number of trips made by bus per week	<5	TF1	26%	
	5–15	TF2	54%	
	15–30	TF3	18%	
	>30	TF4	1%	
Monthly income	<900€	I1	7%	
	900€–1500€	I2	20%	
	1500€–2500€	I3	17%	
	>2500€	I4	14%	
	No answer	I5	42%	

Most of the trips were home-related, work being the second most important reason and about 10% of the trips had as origin or destination some place of leisure. Showing that public transport in the city is mainly used for commuting purposes. In relation with the previous results, the frequency of use of the bus by the respondents was mainly less than 15 trips per week, which represents regular users of the bus on weekdays, not using it too much on weekends. Nearly a quarter of the sample (26%) was an occasional user with a marginal use of less than 5 trips per week. The income level of the sample is mainly average, but a large part of the sample (42%) preferred not to answer this question, making clear the personal nature of the question and the reluctance of respondents to provide this information.

2.3. Rating Scale Results

Table 2 shows the results obtained in the rating of the 24 attributes and the overall satisfaction. For ease of understanding, the averages have been calculated by recoding the qualitative responses on a numerical scale between 0 and 10. 0 being the most negative response ("Very Bad") and 10 the most positive response ("Very good"). The attributes are ordered from highest to lowest level of satisfaction. The best rated attribute was the use of hybrid vehicles, obtaining a score of 8.10, clearly the highest of all the attributes. In second and third place are the access time and egress time to and from the stops, this fact makes sense due to the great density of stops available in the city. The separation between stops is lesser than 400 m at any point of the network and the average is close to 300 m. The overall satisfaction of the service was 6.73 with a standard deviation of 2.01.

Table 2. Rating scale results.

Order	Attribute	Acronym	Mean	Std. Deviation
1	Use of hybrid buses	HY	8.10	1.91
2	Access time to bus stop	AT	7.34	2.24
3	Egress time from alighting stop to destination	DT	7.29	2.22
4	Vehicle cleanliness	CL	7.04	1.75
5	Ease of transfer	TR	6.96	2.24
6	Information at stops	IS	6.90	2.43
7	Information on board	IB	6.83	2.24
8	Comfort of the buses	CM	6.77	1.86
9	Service reliability/punctuality	SR	6.75	2.16
10	Driver's kindness	DK	6.57	2.12
11	Quality of bus stops	ST	6.56	2.03
12	Information on mobile app	IM	6.53	3.19
13	Line coverage	LC	6.50	2.07
14	Information on the web page	IW	6.46	2.39
15	Priority seats for people with reduced mobility (PRM)	RM	6.27	2.22
16	Waiting time	WT	6.26	2.27
17	Crowding level	OC	6.25	2.17
18	Readability of map design	MD	6.21	2.46
19	In-vehicle travel time	TT	6.18	2.14
20	Service frequency and timetables	SE	6.10	2.42
21	Driving style	DS	5.98	2.15
22	Price/Fare	PR	5.83	2.35
23	Calefaction/air conditioning	CA	5.77	2.49
24	Noise	NO	5.71	2.06
	Overall satisfaction	OS	6.73	2.01

3. Model Specification

3.1. Multinomial Logit for Best-Worst Scaling

The structure of the survey has been based on BW case 1 exercises. Each BW case type requires a different modelling specification [43]. For this study a specification based on the logit model has been adopted, which assumes that the non-observable part of the utility is distributed according to a type 1 GEV function or Gumbel distribution with the random variables distributed identically and independently. More specifically the multinomial logit model (MNL).

A total of K attributes have been defined in the survey. In each BW choice the respondent had to choose between 4 attributes of a subset Y. The probability of choosing attribute b as best and attribute $w \neq b$ as worst within the subset Y is defined as $P_{BW}(bw \mid Y)$. The survey instrument was programmed in such a way that the respondent cannot advance if the same attribute was selected as both best and worst attributes.

The choice probability according to the Logit specification can be calculated as defined in (1). This specification is also called the Maxdiff model [49].

$$P_{BW}(bw|Y) = \frac{\exp[v(b) - v(w)]}{\underset{\substack{l,k \in Y \\ l \neq k}}{\Sigma} \exp[v(l) - v(k)]} \tag{1}$$

where $v(.)$ is the observable utility components specified as a linear-in-parameter function of attributes such as $v(k) = \delta_k y_k$ where y_k is an indicator vector of 0 and 1 ($y_k = 1$ when the attribute k is shown to the respondent i and 0 otherwise. In this way, the parameter estimate δ_k could be interpreted as the importance or level of attribute k relative to the reference/base attribute which has $\delta_0 = 0$. The previous notation considers the parameter δ_k to be constant for all respondents i. This is a strong assumption that may not be realistic considering the heterogeneity of the analysed sample. To make this assumption more flexible, a variation of the parameter δ has been included, considering the systematic heterogeneity of the sample. Thus, $\delta_i = \delta + \Lambda z_i$ where δ remains the vector of constants associated with attributes k, while Λ is the vector of parameters associated with the socioeconomic variables z_i of each individual i.

The maxdiff model assumes that the respondent simultaneously chooses the best and the worst options; however, it may be possible that the respondent selects the best option first, then eliminate this attribute out of the choice set before selecting the worst option. In this case, the repeated best-worst model specification could be more appropriate [50]. However, empirical studies [51] have shown that these alternative model specifications are likely to produce similar results. For this specific case, since both models have also shown almost identical results, we opted for the maxdiff model due to the slightly better model fit.

3.2. Ordered Logit Model

Ordered models, in their most contemporary form based on regression, were proposed by McKelvey and Zavoina [52,53]. The objective of these models is to analyse ordered, categorical and non-quantitative choices.

A latent regression model is established, where the dependent variable q_i^* is estimated based on a regression composed by some observable systematic variables v_i that multiply the marginal utilities θ' and finishes with an unknown random component ε_i. The conventional assumptions are that ε_i is continuous, random and follows a certain cumulative distribution function (CDF), $F(\varepsilon i | xi) = F(\varepsilon i)$:

$$q_i^* = \theta' \mathbf{v}_i + \varepsilon_i, \ i = 1, \dots, n \tag{2}$$

The latent variable q_i^* is divided into different bands J delimited by $J + 2$ threshold parameters μ. The discrete latent variable is defined as q_i.

$$
\begin{aligned}
q_i \ &= 0 \text{ if } \mu_{-1} < q_i^* \leq \mu_0, \\
&= 1 \text{ if } \mu_0 < q_i^* \leq \mu_1, \\
&= 2 \text{ if } \mu_1 < q_i^* \leq \mu_2, \\
&= \dots \\
&= J \text{ if } \mu_{J-1} < q_i^* \leq \mu_J.
\end{aligned} \tag{3}
$$

In this study, the unobservable continuous variable q_i^* represents the overall service satisfaction, with q_i being the discrete observable variable of satisfaction. The J bands represent the response options of the Likert scale. The variable v_i is the satisfaction rating of the 24 attributes included in the survey made by each respondent i.

Equations (2) and (3) assume that neither parameters θ nor thresholds μ vary across individuals. This assumption of homoscedasticity is arguably strong but can be relaxed. To include heterogeneity in the parameters, the effect that systematic or random variations in users taste have can considered. Previous studies have already demonstrated the influence of socioeconomic attributes on ordered models [42]. That same study established the need to analyse in more detail the influence of the importance that users place to different attributes in the models. In this study, the form of inclusion of importance has been modified based on the results of the BW models explained above, which, in parallel, consider the heterogeneousness of the sample. To this end, it is considered that the parameter θ_i is composed by a constant term θ to which a variation $\Delta\delta_i$ is added, $\theta_i = \theta + \Delta\delta_i$. The parameters Δ are estimated considering the results of the MNL model based on BW, δ_i, for each individual i. In order to include the parameters of the MNL model into the ordered model it has been necessary to perform a normalization of the resulting parameters δ_i. The δ have been normalized for each individual i. The parameter with the highest value has been given the value 1, while the parameter with the lowest value has been given the value 0. The rest of the parameters have been weighted linearly between the two values.

In accordance with the survey structure defined in Section 2.1, respondents evaluated only 12 of the 24 attributes included in the survey. Due to this lack of information, it is necessary to complete the database in order to estimate the ordered models. The method used to complete the sample has been based on Multiple Imputation [54,55]. The study developed by Echaniz et al. [56] showed that it is possible to obtain conclusive results using this imputation method for satisfaction studies in public transport. Multiple imputation is estimated by using a procedure called Fully Conditional Specification (FCS), which uses an iterative Monte Carlo method with Markov chains [57]. The FCS approach is based on variable-by-variable imputation of data, specifying an estimation model for each of the variables with missing data. By using the multiple imputation method a small prediction error is assumed. Echaniz et al. [56] showed that estimating a model by using multiple imputation to complete the database leads to a lower model fit. The differences on fits between full dataset models and partial data models was empirically proven to be 2% less correct predictions when multiple imputation is used to complete the missing information. The difference between fits was considered acceptable after testing the models using a Vuong test for non-nested models [58].

With the database completed, the estimation of the models is done as usual. To estimate the model some normalizations are required [59]. The associated probabilities are defined as:

$$\mathrm{Prob}\big[q_i = j|v_i\big] = \mathrm{Prob}\big[\varepsilon_i \leq \mu_j - \theta' \mathbf{v}_i\big] - \mathrm{Prob}\big[\mu_{j-1} - \theta' \mathbf{v}_i\big],\ j = 0, 1, \ldots, J \tag{4}$$

The model (4) is estimated using maximum likelihood estimator which maximises the log-likelihood function defined as follows:

$$\log L = \sum_{i=0}^{n} \sum_{j=0}^{J} m_{ij} \log\big[F(\mu_j - \theta' \mathbf{v}_i) - F(\mu_{j-1} - \theta' \mathbf{v}_i)\big] \tag{5}$$

where $F(.)$ is the cumulative distribution function; $m_{ij} = 1$ if $q_i = j$ and 0 otherwise.

4. Results and Discussion

The results of the data modelling are shown below. A total of six models are presented. The first two correspond to the BW choice made in the survey. The next four correspond to the ordered models estimated with the rating data obtained in the survey, which are combined with the results of the BW

modelling to compare the improvement in fit achieved by including the importance of the different attributes in the ordered models.

4.1. Modelling Results from Best-Worst Scaling

The following table (Table 3) shows the result of the estimated MNL model with the BW data. In this first model, it is assumed that there is no heterogeneity in the sample, so the estimated parameters are constant and equal for the whole sample. The meaning of the parameters is interpreted based on their value.

Table 3. MNL model for BW data.

Variable	Acronym	Parameter
Access time to bus stop	AT	0.346 (2.84)
Waiting time	WT	0.870 (7.04)
In-vehicle travel time	TT	1.121 (8.82)
Egress time from alighting stop to destination	DT	0.512 (4.17)
Price/Fare	PR	0.742 (6.12)
Ease of transfer	TR	0.043 (0.35)
Service frequency and timetables	SE	1.143 (9.17)
Service reliability/punctuality	SR	1.218 (9.86)
Line coverage	LC	1.117 (8.95)
Information at stops	IS	0.202 (1.64)
Information on the web page	IW	−1.118 (−8.97)
Information on board	IB	−0.869 (−6.99)
Crowding level	OC	0.360 (2.97)
Calefaction/air conditioning	CA	−0.688 (−5.57)
Priority seats for people with reduced mobility (PRM)	RM	0.503 (4.07)
Comfort of the buses	CM	−0.006 (−0.05)
Vehicle cleanliness	CL	−0.228 (−1.88)
Driving style	DS	0.356 (2.88)
Driver's kindness	DK	−0.494 (−3.99)
Use of hybrid buses	HY	−0.035 (−0.28)
Noise	NO	−0.645 (−5.26)
Information on mobile app	IM	−0.160 (−1.28)
Quality of bus stops	ST	−0.533 (−4.29)
Readability of map design	MD	0
Log-likelihood		−5373.256
AIC/*n*		4.452

For the modelling requirements it is necessary to set the value 0 to one of the parameters in order to estimate the rest. In this case, the reference parameter chosen has been the "readability of map design" (MD). The rest of the attributes have an associated parameter that can be greater or less than 0. The value of the parameter represents the importance of the attribute to which it is associated. The greater the value of the parameter, the greater the importance of the corresponding attribute. On the contrary, if a parameter has a negative value, it means that the importance of that attribute is less than the attribute established as a base. The differences in the values of the parameters represent the relative differences in importance levels between the different attributes, assuming a greater difference in values a greater relative difference between one attribute and another. The statistical significance of each parameter is established by the *z* value shown in parentheses.

The values in Table 3 show that the attributes that are most important to users are those directly related to the characteristics of the service, i.e., in order of highest to lowest importance, "service reliability/punctuality", "service frequency and timetables", "in-vehicle travel time" and "line coverage". By contrast, the less important variables are "information on the web page" and "information on board". The vast majority of the parameters have been found to be statistically significant at a 95% level, showing a z-values greater than 1.96.

The second model estimated using BW data is shown in Tables 4 and 5. In this case, it has been considered that the sample is heterogeneous, in such a way that the socio-economic characteristics observed in the survey affect the perception of the importance of the attributes. The way to include heterogeneity has been by including interactions in the utility functions of each attribute, as explained in Section 3.1. To simplify the results in Tables 4 and 5, only those interactions that have been found to be significant at least at a 90% confidence interval are shown, omitting the rest of the interactions. At the same time, it has been decided to omit the value z for reasons of simplicity. The coding used to include the socio-economic variables has been by means of effect coding [60]. This coding is similar to the dummy coding with the peculiarity that it allows to capture the effect associated with the reference variable of each group that is not included in the model. This effect is calculated as the sum of the negatives of the rest of the parameters belonging to the same group. For each group of socioeconomic variables (Table 1) the last variable of each groups have been takes as the reference variable. Additionally, the time of the day when the surveys were carried out have been taken into account.

Table 4. MNL model for BW data considering customer heterogeneity (1).

Var.	Cons.	F	L1	L2	L3	AT1	AT2	AT3	AT4	I1	I2	I3	I4	P2	P3
AT	0.346	-	−0.427	0.633	-	0.228	-	-	-	-	-	-	-	-	-
WT	0.903	-	-	-	0.561	-	-	-	-	-	-	-	-	-	-
TT	1.332	-	-	-	-	-	-	-	-	-	-	-	0.410	−0.340	0.359
DT	0.450	-	-	-	0.264	-	-	-	-	-	-	-	-	-	−0.462
PR	0.691	-	-	-	0.352	-	-	-	-	-	-	-	-	−0.426	-
TR	−0.056	-	-	-	-	-	-	-	-	-	0.242	-	-	-	-
SE	1.436	-	−0.361	-	0.383	-	0.334	-	-	0.354	-	-	-	-	-
SR	1.267	-	−0.276	-	0.342	-	-	-	-	-	-	-	-	-	-
LC	1.041	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS	0.270	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IW	−1.275	-	-	-	0.463	-	-	-	-	-	-	-	-	-	-
IB	−0.990	-	-	-	0.359	-	-	-	-	-	-	-	-	-	-
OC	0.750	-	-	-	0.475	−0.500	-	0.590	-	-	-	-	-	-	-
CA	−0.770	-	-	-	0.403	-	-	-	-	-	-	-	-	-	-
RM	0.797	0.258	-	-	-	-	-	-	-	0.503	-	−0.425	-	-	-
CM	0.159	-	-	-	0.671	-	-	-	-	-	-	-	-	-	-
CL	−0.234	-	-	-	0.437	-	-	-	-	-	-	-	-	−0.374	-
DS	0.343	0.210	−0.550	-	0.383	-	-	-	−0.304	-	-	-	-	-	-
DK	−0.385	-	−0.410	-	0.781	-	-	-	-	0.549	-	-	−0.594	−0.593	-
HY	−0.083	-	−0.287	0.434	0.428	-	-	-	-	-	-	-	-	-	-
NO	−0.592	-	0.443	-	0.527	-	-	-	-	-	-	-	-	-	-
IM	−0.464	-	-	-	0.304	-	-	0.329	-	-	-	-	-	-	-
ST	−0.435	-	−0.458	-	0.553	-	-	-	-	-	-	-	-	−0.410	-
MD	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. MNL model for BW data considering customer heterogeneity (2).

Var.	P5	A1	A2	A3	A4	A6	W1	W2	W3	W4	TF1	TF2	TF3	H1	H2	H3	H4
AT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DT	-	-	0.607	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PR	-	-	-	-	-	-	-	-	-	−0.567	-	0.325	-	-	-	-	-
TR	-	-	-	-	-	-	-	-	-	-	-	-	0.734	−0.474	-	0.520	-
SE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.420	-	-
SR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LC	-	-	-	-	0.460	-	-	-	-	-	-	-	0.523	0.414	-	-	-
IS	-	-	-	-	-	-	-	-	-	-	-	-	-	0.693	−0.535	0.539	-
IW	-	-	-	0.460	-	-	-	-	-	-	-	-	-	−0.489	-	-	-
IB	-	-	-	-	-	0.575	-	-	-	-	-	-	-	-	-	-	-
OC	-	−0.557	-	-	-	-	-	−0.331	-	-	-	-	-	-	-	-	-
CA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RM	-	-	-	-	-	-	1.054	−0.521	−0.435	-	-	-	-	-	-	-	-
CM	-	−0.437	-	-	-	0.441	-	-	-	-	−0.523	-	−0.462	-	-	−0.483	0.336
CL	-	-	-	-	-	-	-	-	-	-	0.538	-	-	-	-	-	-
DS	−0.519	-	-	-	-	-	0.490	-	-	-	−0.589	-	-	-	-	-	-
DK	-	−0.739	-	0.724	-	-	-	-	-	-	-	-	-	-	−0.356	-	0.410
HY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NO	-	-	-	-	-	0.486	-	-	-	-	-	-	-	-	-	-	-
IM	-	0.632	0.534	-	-	−0.649	-	-	-	-	-	0.359	0.438	-	-	-	-
ST	-	-	-	−0.455	-	-	-	-	-	-	-	-	-	-	-	-	-
MD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Log likelihood	−5134.770
AIC/*n*	4.331

Five periods were defined: morning peak hour (H1), midday peak hour (H2), afternoon peak hour (H), morning off peak hour (H4) and afternoon off-peak hour (H5). As for the other variable groups, the last variable have been chosen as the reference variable.

The way to interpret the results of these models is through the signs and parameter values of the interactions. The constant parameters of each attribute show at the first glance the level of importance of each attribute in a similar way to the constant-only model (Table 3), however, in this case it is not conclusive because depending on the person being analysed, the final resultant values can alter the order. Overall, the most important attributes are the service frequency (SE), the travel time (TT) and the service reliability (SR).

On the other hand, the least important attributes are the information on the web page (IW) and the information on board the bus (IB). Analysing the heterogeneity of the sample, when an interaction has a positive value, it represents that that user segment considers the interacting parameter as more important than the rest, while if an interaction is negative, the effect is the opposite, the attribute is less important for that user. The dispersion of interactions is considerable, showing that the importance of different attributes are affected by different user characteristics. The variable that most influences the perception of importance is the analysed line, especially in the case of line 3 (L3), where the waiting time (WT), the egress time (DT), the price (PR), the service frequency (SE), the service reliability (SR), the information on the web page (IW), the information on board the bus (IB), the crowding level (OC), the calefaction system (CA), the comfort of the buses (CM), the cleanliness (CL) the driving style (DS), the drivers' kindness (DK) and the use of hybrid buses (HY) show positive values, meaning that for the line 3, those attributes show a greater importance than for the rest of the lines. Which also can be interpreted as those attributes not affected by L3 variable are considered lesser important for the people that use line 3.

Lines 1 and 2 show a similar importance levels across the board, with some exceptions: access time (AT) service frequencies and reliability (SE and SR), drivers' kindness and driving style (DK and DS), the use of hybrid vehicles (HY), noise (NO) and quality of bus stops (ST). Most of those variations are negative for line 1 (L1), being less important for the users of that line. The gender of the users does not greatly affect the perception of the importance of the attributes. Looking at alternative modes of transport, users with the possibility of taking the car (AT1) give more importance to the access time, while they consider occupation less important, while those who can make the trip by bicycle consider it more important (AT3). The level of income of the users affects in a varied way, those with a high level of income (I4) give more importance to the time spent on board the bus. Regarding trip purposes, household-related trips (P2) show less importance than price (PR) and the on board travel time, while for leisure trips (P3), the travel time is more important. For health-related trips (P5), the importance of the driver's driving style (SD) becomes less important. In terms of age, younger age groups (A1 and A2) give greater importance to mobile phone information (MI) than others, with older users (A6 and A7) showing the opposite effect. With regard to the employment situation, domestic workers consider the space for people with reduced mobility and the driving style to be more important, workers consider crowding level to be less important and, like students, they consider the space for people with reduced mobility of lower importance. Pensioners consider the price less important, which is understandable due to the heavy subsidy for retired people offered by the transport service. The most frequent users of the service (TF3) consider the ease of transfers (TR) and the coverage of the lines (LC) to be of greater importance. The time frames mainly affect transfers and line coverage.

Those variables used as base do not appear in the model, however, as mentioned before, because the codification used in the modelling for the socioeconomic variables was effects codding [60], it is possible to study how the importance vary across the user segments defined by the base variables. The impact of a base variable of a group is calculated as the sum of the negative values of the parameters of the rest of the parameters belonging to that group. For the gender case, as an example, the female variable was included in the model, which means that, male users consider less important the priority seats for people with reduced mobility (RM) and driving style (DS). In those cases where there are more

than two variables within a socioeconomic group, age for example (A1 to A7), the effect of A7, which is not explicitly included in the model, is calculated as the negative sum of the rest. Regarding destination time (DT) the only age group showing a significant parameter is A2 (age 25 to 34), therefore, the effect on A7 would be (−0.607), being the negative sum of all the parameters within that socioeconomic group. Same occurs for all the socioeconomic variable groups.

Comparing the model adjustment indicators of previous models (Tables 3–5), it can be seen that the inclusion of the interactions improves the model adjustment. It is widely known that public transport users are a heterogeneous group of people. Using a homogeneous model to represent the complete sample (Table 3) can be useful to some extend in order to explain those aspects of the service that are important. More or less all the user types consider as most important attributes some key aspects of the service, therefore those aspects can be modelled with an acceptable error through a homogeneous model. However, once you get deeper on the analysis, as the modelling results shown above have proven (Tables 4 and 5), the difference in perception across individuals is significant and can be defined by using observable socioeconomic variables. By considering the heterogeneity of the sample, it is possible to understand the preferences of each user group. Therefore, improving those most important aspects can maximize the efforts made for increasing user satisfaction regarding public transport services.

4.2. Ordered Logit Modelling Results

The ordered logit modelling results are shown below (Tables 6 and 7). Model 1 shows the basic model, where the dependent variable OS is estimated directly based on the ratings made to the attributes. The model shows those parameters that have been found to be statistically significant, the *z*-value is shown in parentheses. The second model shows the ordered model with the variables weighted by the parameters estimated in the MNL model with the BW data. In this second model all the weighted variables have been include. However, a large number of them have turned out to be statistically non-significant. In model 3, model 2 has been refined considering the weighted variables that have turned out to be statistically significant, reducing in a great number the variables included in the model and facilitating their interpretation. Finally, Model 4 shows an only weighted variable model.

Table 6. Estimated ordered logit models (1).

Variables	Model 1	Model 2	Model 3	Model 4
Constant	−1.305 (−2.88)	−1.968 (−3.65)	−1.716 (−3.43)	−0.209 (−0.58)
Access time to bus stop	0.151 (2.80)	0.055 (0.21)	0.045 (1.77)	-
Waiting time	0.142 (3.14)	0.265 (2.34)	0.574 (2.87)	-
In-vehicle travel time	0.080 (3.71)	0.044 (0.19)	0.028 (2.65)	-
Egress time	0.067 (3.07)	0.119 (0.77)	-	-
Price/Fare	-	0.076 (0.91)	-	-
Ease of transfer	0.134 (4.24)	0.091 (0.84)	0.162 (3.05)	-
Service frequency and timetables	0.298 (5.11)	0.616 (2.72)	0.412 (4.20)	-
Service reliability/punctuality	0.270 (4.56)	0.464 (1.09)	0.262 (4.35)	-
Line coverage	0.200 (3.59)	0.209 (0.90)	0.195 (3.45)	-
Information at stops	0.155 (2.86)	0.139 (1.27)	0.164 (2.95)	-
Information on the web page	0.101 (3.35)	0.109 (2.58)	0.113 (2.77)	-
Information on board	0.059 (2.68)	0.018 (0.26)	0.060 (2.42)	-
Crowding level	0.145 (3.28)	0.136 (1.01)	0.104 (2.22)	-
Calefaction/air conditioning	0.062 (2.86)	0.085 (1.39)	0.065 (1.98)	-
Priority seats for PRM	0.055 (2.47)	0.119 (1.26)	0.056 (2.46)	-
Comfort of the buses	0.063 (2.63)	0.106 (1.61)	0.079 (1.64)	-
Vehicle cleanliness	0.108 (2.29)	0.044 (0.37)	0.024 (2.00)	-
Driving style	0.078 (2.41)	0.086 (1.01)	0.058 (1.05)	-
Driver's kindness	0.021 (1.98)	0.052 (0.96)	0.042 (0.76)	-
Use of hybrid buses	0.039 (1.87)	−0.094 (−0.67)	-	-
Noise	0.030 (2.66)	0.034 (1.43)	0.033 (2.80)	-
Information on mobile app	0.045 (2.23)	0.035 (0.40)	0.022 (2.12)	-
Quality of bus stops	0.154 (2.76)	0.093 (1.09)	0.165 (2.91)	-
Readability of map design	0.180 (3.25)	0.166 (1.30)	0.173 (3.06)	-

Table 7. Estimated ordered logit models (2).

Variables	Model 1	Model 2	Model 3	Model 4
		Weighted Variables		
δ*Access time to bus stop	-	0.163 (0.37)	0.197 (2.69)	0.151 (2.67)
δ*Waiting time	-	−0.166 (−1.30)	−0.568 (−2.41)	0.139 (2.30)
δ*In-vehicle travel time	-	0.097 (0.39)	0.108 (2.14)	0.069 (2.16)
δ*Egress time	-	−0.030 (−0.13)	0.144 (1.81)	0.054 (1.81)
δ*Price/Fare	-	−0.043 (−0.45)	-	0.236 (3.07)
δ*Ease of transfer	-	0.163 (0.80)	-	0.300 (4.60)
δ*Service frequency and timetables	-	−0.364 (−1.44)	−0.124 (−1.33)	0.255 (3.93)
δ*Service reliability/punctuality	-	−0.231 (−0.48)	-	0.198 (3.16)
δ*Line coverage	-	−0.006 (−0.03)	-	0.136 (1.66)
δ*Information at stops	-	0.031 (0.20)	-	-
δ*Information on the web page	-	0.005 (0.06)	-	0.063 (1.30)
δ*Information on board	-	0.299 (1.06)	0.170 (1.11)	0.212 (2.47)
δ*Crowding level	-	0.048 (0.25)	0.076 (4.43)	0.080 (1.89)
δ*Calefaction/air conditioning	-	−0.017 (−0.16)	-	0.086 (1.38)
δ*Priority seats for PRM	-	−0.071 (−0.53)	-	0.045 (2.25)
δ*Comfort of the buses	-	−0.037 (−0.90)	-	0.120 (1.60)
δ*Vehicle cleanliness	-	0.164 (0.60)	0.229 (2.25)	0.098 (1.95)
δ*Driving style	-	−0.006 (−0.01)	-	-
δ*Driver's kindness	-	−0.006 (−0.09)	-	0.130 (1.56)
δ*Use of hybrid buses	-	0.388 (1.33)	0.223 (2.04)	0.045 (1.24)
δ*Noise	-	-	-	0.051 (1.30)
δ*Information on mobile app	-	0.089 (0.48)	0.115 (1.44)	0.331 (2.06)
δ*Quality of bus stops	-	0.278 (1.08)	-	0.264 (2.32)
δ*Readability of map design	-	0.019 (0.06)	-	1.389 (11.08)
Mu(01)	1.427 (11.07)	1.437 (11.03)	1.442 (11.08)	3.505 (33.32)
Mu(02)	3.628 (33.61)	3.671 (33.20)	3.672 (33.482)	6.773 (46.76)
Mu(03)	6.997 (46.47)	7.131 (45.55)	7.104 (45.91)	−0.209 (−0.58)
Log-likelihood	−818.697	−805.064	−808.399	−833.370
AIC/*n*	2.072	2.1008	2.066	2.1148
Count R^2	0.590	0.596	0.592	0.578

For consistency, the constant term of the model has to be negative and the rest of the parameters (sum of the independent parameter with its weighting in the case of models 2 and 3) have to be positive [42,48]. It can be see that all 4 models comply with those conditions. The value of the parameter represents the contribution that a specific variable makes to explain the overall satisfaction of a user. Higher the value, greater its influence when defining the overall satisfaction of the service. In model 1 it can be seen that the parameters with the highest value are those associated with the attributes "Service frequency and timetables", "Service reliability/punctuality" and "Line coverage". These are very similar to the most important variables derived from the BW exercise. The lowest parameter is associated to the variable "Driver's kindness".

Model 2 shows an improvement in model fit with a log likelihood value closer to 0 and a slightly higher Count R^2. However, AIC/*n* value is higher than in model 1, which indicates that the improvement in fit compared to the considerable increase in variables included in the model is not optimal. Therefore, it is quite possible that the number of variables in the model can be reduced considerably without losing too much capacity to represent reality. The correlation between weighted variables and regular variables is considerable. Therefore, including both in the model generates a lower significance level of all the parameters. Some parameters show a greater significance level on the weighted variables, while, other show a greater significance in the non-weighted part. To improve the significance level of the parameters it is necessary to modify the model by reducing the number of the parameters by choosing the weighted or the non-weighted parameter in each case. The reduction on the number of parameters is shown in model 3, where the most significant parameters have been kept, while the non-significant ones have been omitted from the model.

Model 3 shows a reduced version of model 2. The adjustment fit of model 3 is slightly lower than model 2, however, its fit is still better than that of model 1. Therefore, the inclusion of the weighting based on importance improves the initial fit of the model. Moreover, by reducing the number of parameters in comparison to model 2, their significance levels have increased considerably, making the model more consistent in terms of the estimated parameters. The weighted parameters show a differential effect in the explicit importance give to the contribution of a variable when explaining the overall service satisfaction. When the value of the parameter of a weighted variable is positive, the contribution of the variable increases the greater its importance for the user is. Ergo, the influence of the attribute to which the parameter is associated has more influence on that particular user than on the rest of the users. On the contrary, when the value of the parameter of a weighted variable is negative, the importance of the attribute makes its effect on the general satisfaction lower. This may be due to the fact that respondents who consider certain variables to be important do not really evaluate the service based on those variables, but rather the greater or lesser general satisfaction is related to those attributes that they consider to be more secondary, which are those that influence their perception. The variables affected by the weighting that have shown an acceptable level of significance level are "access time to bus stop", "waiting time", "in-vehicle travel time", "egress time", "service frequency and timetables", "information on board", "crowding level", "vehicle cleanliness", "use of hybrid buses" and "information on mobile app". Except for waiting time and frequency, all other weighted variables show a positive parameter. Particularly noteworthy are those variables that do not show a very high level of significance in model 1, but which, by including the specific importance that they have for each user, do show themselves to be statistically significant, this is the case of the "information on board", "vehicle cleanliness", "use of hybrid buses" and "information on mobile app". These are primarily secondary attributes of low general importance but may show greater importance for a certain segment of users.

Model 4 consider only weighted parameters. The results show that this model is the one with the worst fit in comparison with the other ones. Comparing this weighted only model with model 1, with no weighted variables, it can be seen that weighting all the variables of the model does not benefit the model's fit. As it can be seen in model 3, only some variables are significantly affected by the inclusion of their importance level through a weighting method.

All four models show very similar adjustment fit in terms of the fit indicators. The model showing the best fit regarding log-likelihood and Count R^2 values is model 2. However, in the case of AIC/n indicator, the best positioned model is model 3. In contrast to log-likelihood and Count R^2 values, the AIC/n considers the number on the parameters within a model. In other words, the AIC/n indicator shows the trade-off between the model's fit and the number of parameters required for that fit. Understanding that if the improvement on a model's fit requires too many additional parameters, then the model is considered not optimised. Therefore the AIC/n indicator value ends up being higher and the overall fit of the model is higher. That is the case of model 2, where the adjustment fit is the highest, but because of the considerable number of parameters used in this model, the AIC/n indicator is worse than models 1 and 3.

To sum up, the module comparison has shown that including attribute explicit importance in ordered logit models improves the prediction capability of the models. However, the weighting is not applicable to all the variables, as some of them are not affected by the importance level while explaining the overall satisfaction of a public transport system. Also, a huge correlation between the weighted and non-weighted variables has been found, therefore, including both variables on the model simultaneously decreases the significance level of the parameters. Hence, the inclusion of weighted variables on a model should be selective, by including as weighted only those variables that are significantly affected by the importance level perceived by the user. Finally, the only weighted variable model (model 4) have shown a poor adjustment fit. Consequently, it is not advisable to use that model (weighted only) to analyse satisfaction on a public transport service, being even preferable

to use the non-weighted variables model (model 1), which is more cost effective to estimate and also have shown to work properly in previous studies.

5. Conclusions

This article has developed a method to include weighted variables in ordered logit models that considers the importance that users give to the different attributes that make up a public transport service. To establish this importance level, a multinomial logit model has been estimated based on data obtained through BW type questions.

The modelling of the BW responses has been carried out considering the heterogeneity of the users through the socio-economic information obtained in the survey. The way in which this information has been included in the model has been by carrying out interactions between the socio-economic variables and the attributes evaluated, adding different components to the utility functions of each variable. The results have shown that the importance of the attributes is conditioned by the socioeconomic and travel characteristics of the users, improving in turn the predictive capacity of the models. This result is in line with that established in [48] where it was shown that there was variability in the perception of importance. In this previous study this variability was considered to be random, however, in this study it has been shown that a large part of this variability can be explained by systematic variations. The variable that most affects the perception of importance is the transport service line analysed, which suggests the importance of the segregated analysis between the different lines of public transport that make up a service. In [35] it was shown that the analysis of quality using ordered models also varies from one line to another, so the results of this study reaffirm the need of studying the public transport systems segregating by lines. Additionally, the results presented in [48] show that the importance of nine of the 24 attributes (egress time, service reliability/punctuality, line coverage, vehicle cleanliness, driver's kindness, noise, information on mobile app, readability of map design and information on the web page) did not vary across individuals when considering random variations. However, when studying the systematic variations it has been proven that even those attributes are affected to some extent by the users' characteristics. For example, the importance of egress time (DT) varies depending on the line analysed and for the age group between 25 and 34 like for the rest of the attributes mentioned above, for the exception of the readability of the map design, for which no interaction was found to be statistically significant. Developing an in-depth analysis on changes in user's perception considering systematic heterogeneity of the sample enables more user focused decision makings. Knowing how importance levels vary from one user to another makes it easier for policy makers to encourage specific users to use a public transport by improving those aspects that are important for them. Also, considering systematic variation on users' tastes have shown to be more beneficial than simply considering random heterogeneity, mainly because with systematic variation it is possible to know which is the reason behind differences in perception, but also, because it has been proven that random heterogeneity does not capture some of the variation existing in the sample.

With regard to the weighting of variables in the ordered models, this study wanted to check the questions raised by the previous study developed by Echaniz et al. [42], where it was questioned whether another method of weighting the variables by importance could have a greater influence on improving the predictions made by the ordered models. In this study it has been seen that when the weighting is carried out with the BW-based models the improvement was not completely significant. Therefore, including weighted variables based on the level of importance in ordered do not greatly improve models' fit. Moreover, the number of variables present in the model increases to a great extent without substantially improving the results. Henceforth, coinciding with the results obtained in [42]. However, some initially non-significant variables turn out to be significant when weighted according to their importance level. Those attributes were: "information on board", "vehicle cleanliness", "use of hybrid buses" and "information on mobile app".

For future lines of research, it is worth mentioning that having demonstrated that the perception of the importance of attributes changes from one user to another, this opens the possibility of carrying

out a specific importance-performance study for each target user segment. This study could define those variables that require greater investment in order to improve the satisfaction of different users. In turn, having observed a variation depending on the transport lines, this same study could focus on defining the specific variables to maximize satisfaction in each of the lines that make up the service separately. At the same time increasing the effectiveness of actions at the particular and global level of the service.

References

1. Hensher, D.A.; Prioni, P. A Service Quality Index for Area-wide Contract Performance Assessment. *J. Transp. Econ. Policy* **2002**, *36*, 93–113.
2. Hensher, D.A.; Stopher, P.; Bullock, P. Service quality—Developing a service quality index in the provision of commercial bus contracts. *Transp. Res. Part A Policy Pract.* **2003**, *37*, 499–517. [CrossRef]
3. Mokonyama, M.; Venter, C. Incorporation of customer satisfaction in public transport contracts—A preliminary analysis. *Res. Transp. Econ.* **2013**, *39*, 58–66. [CrossRef]
4. Allen, J.; Eboli, L.; Forciniti, C.; Mazzulla, G.; Ortúzar, J.D.D. The role of critical incidents and involvement in transit satisfaction and loyalty. *Transp. Policy* **2019**, *75*, 57–69. [CrossRef]
5. Allen, J.; Bellizzi, M.G.; Eboli, L.; Forciniti, C.; Mazzulla, G. Latent factors on the assessment of service quality in an Italian peripheral airport. *Transp. Res. Procedia* **2020**, *47*, 91–98. [CrossRef]
6. Eboli, L.; Mazzulla, G. Relationships between rail passengers' satisfaction and service quality: A framework for identifying key service factors. *Public Transp.* **2015**, *7*, 185–201. [CrossRef]
7. Hapsari, R.; Clemes, M.; Dean, D. The Mediating Role of Perceived Value on the Relationship between Service Quality and Customer Satisfaction: Evidence from Indonesian Airline Passengers. *Procedia Econ. Financ.* **2016**, *35*, 388–395. [CrossRef]
8. Machado, J.L.; de Oña, R.; Diez-Mesa, F.; de Oña, J. Finding service quality improvement opportunities across different typologies of public transit customers. *Transp. A Transp. Sci.* **2018**, *14*, 761–783. [CrossRef]
9. Rose, J.M.; Hensher, D.A. User satisfaction with taxi and limousine services in the Melbourne metropolitan area. *J. Transp. Geogr.* **2018**, *70*, 234–245. [CrossRef]
10. Shen, W.; Xiao, W.; Wang, X. Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares. *Transp. Policy* **2016**, *46*, 20–31. [CrossRef]
11. Allen, J.; Muñoz, J.C.; Ortúzar, J.D.D. Modelling service-specific and global transit satisfaction under travel and user heterogeneity. *Transp. Res. Part A Policy Pract.* **2018**, *113*, 509–528. [CrossRef]
12. Eboli, L.; Mazzulla, G. Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *J. Public Transp.* **2007**, *10*, 21–34. [CrossRef]
13. de Oña, J.; de Oña, R.; Eboli, L.; Mazzulla, G. Perceived service quality in bus transit service: A structural equation approach. *Transp. Policy* **2013**, *29*, 219–226. [CrossRef]
14. Das, T.; Apu, N.; Hoque, M.S.S.; Hadiuzzaman, M.; Xu, W. Parameters Affecting the Overall Performance of Bus Network System at Different Operating Conditions: A Structural Equation Approach. *Transp. Res. Procedia* **2017**, *25*, 5063–5075. [CrossRef]
15. de Oña, J.; de Oña, R.; Eboli, L.; Forciniti, C.; Mazzulla, G. Transit passengers' behavioural intentions: The influence of service quality and customer satisfaction. *Transp. A Transp. Sci.* **2016**, *12*, 385–412. [CrossRef]

16. Tam, M.L.; Lam, W.H.K.; Lo, H.P. Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica* **2010**, *6*, 3–17. [CrossRef]

17. Tsami, M.; Nathanail, E. Guidance Provision for Increasing Quality of Service of Public Transport. *Procedia Eng.* **2017**, *178*, 551–557. [CrossRef]

18. Hernandez, S.; Monzon, A.; de Oña, R. Urban transport interchanges: A methodology for evaluating perceived quality. *Transp. Res. Part A Policy Pract.* **2016**, *84*, 31–43. [CrossRef]

19. de Oña, R.; Eboli, L.; Mazzulla, G. Key factors affecting rail service quality in the Northern Italy: A decision tree approach. *Transport* **2014**, *29*, 75–83. [CrossRef]

20. de Oña, J.; de Oña, R.; López, G. Transit service quality analysis using cluster analysis and decision trees: A step forward to personalized marketing in public transportation. *Transportation (Amst.)* **2016**, *43*, 725–747. [CrossRef]

21. Wu, X.; (Jason) Cao, X.; Ding, C. Exploring rider satisfaction with arterial BRT: An application of impact asymmetry analysis. *Travel Behav. Soc.* **2020**, *19*, 82–89. [CrossRef]

22. Nathanail, E. Measuring the quality of service for passengers on the hellenic railways. *Transp. Res. Part A Policy Pract.* **2008**, *42*, 48–66. [CrossRef]

23. Hrncir, J.; Zilecky, P.; Song, Q.; Jakob, M. Practical Multicriteria Urban Bicycle Routing. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 493–504. [CrossRef]

24. Burlando, C.; Ivaldi, E.; Musso, E. An indicator for measuring the perceived quality of local public transport: Relationship with use and satisfaction with the ticket price. *Int. J. Transp. Econ.* **2016**, *43*. [CrossRef]

25. Cats, O.; Abenoza, R.F.; Liu, C.; Susilo, Y.O. Evolution of satisfaction with public transport and its determinants in Sweden identifying priority areas. *Transp. Res. Rec.* **2015**, *2538*, 86–96. [CrossRef]

26. Wen, C.-H.; Lan, L.W.; Chen, C.-H. Passengers Perception on Service Quality and Their Choice for Intercity Bus Services. In Proceedings of the Transportation Research Board, 84th Annual Meeting, Washington, DC, USA, 9–13 January 2005.

27. Javid, M.A.; Hussain, S.; Anwaar, M.F. Passenger's Perceptions on Prospects of Qingqi Paratransit Public Transport Service in Lahore. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2020**, *44*, 185–195. [CrossRef]

28. dell´Olio, L.; Ibeas, A.; de Oña, J.; de Oña, R. *Public Transportation Quality of Service: Factors, Models, and Applications*; Elsevier: Cambridge, MA, USA, 2017; ISBN 978-0-08-102080-7.

29. Garrido, C.; de Oña, R.; de Oña, J. Neural networks for analyzing service quality in public transportation. *Expert Syst. Appl.* **2014**, *41*, 6830–6838. [CrossRef]

30. de Oña, J.; de Oña, R.; Garrido, C. Extraction of attribute importance from satisfaction surveys with data mining techniques: A comparison between neural networks and decision trees. *Transp. Lett.* **2017**, *9*, 39–48. [CrossRef]

31. Mouwen, A. Drivers of customer satisfaction with public transport services. *Transp. Res. Part A Policy Pract.* **2015**, *78*, 1–20. [CrossRef]

32. Guirao, B.; García-Pastor, A.; López-Lambas, M.E. The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. *Transp. Policy* **2016**, *49*, 68–77. [CrossRef]

33. Bordagaray, M.; Dell'Olio, L.; Ibeas, A.; Cecín, P. Modelling user perception of bus transit quality considering user and service heterogeneity. *Transp. A Transp. Sci.* **2014**, *10*, 705–721. [CrossRef]

34. Rojo, M.; dell'Olio, L.; Gonzalo-Orden, H.; Ibeas, A. Interurban bus service quality from the users' viewpoint. *Transp. Plan. Technol.* **2013**, *36*, 599–616. [CrossRef]

35. dell'Olio, L.; Ibeas, A.; Cecín, P. Modelling user perception of bus transit quality. *Transp. Policy* **2010**, *17*, 388–397. [CrossRef]

36. Tyrinopoulos, Y.; Antoniou, C. Analysis of passengers' perception of public transport quality and performance. In *Transportation Systems and Engineering: Concepts, Methodologies, Tools, and Applications*; Engineering Science Reference (an imprint of IGI Global): Hershey, PA, USA, 2015; Volume 2–3, ISBN 9781466684744.

37. Hensher, D.A.; Mulley, C.; Yahya, N. Passenger experience with quality-enhanced bus service: The tyne and wear "superoute" services. *Transportation (Amst.)* **2010**, *37*, 239–256. [CrossRef]

38. Román, C.; Martín, J.C.; Espino, R. Using Stated Preferences to Analyze the Service Quality of Public Transport. *Int. J. Sustain. Transp.* **2014**, *8*, 28–46. [CrossRef]

39. dell'Olio, L.; Ibeas, A.; Cecin, P. The quality of service desired by public transport users. *Transp. Policy* **2011**, *18*, 217–227. [CrossRef]

40. Soltanpour, A.; Mesbah, M.; Habibian, M. Customer satisfaction in urban rail: A study on transferability of structural equation models. *Public Transp.* **2020**, *12*, 123–146. [CrossRef]

41. Krueger, R.; Vij, A.; Rashidi, T.H. Normative beliefs and modality styles: A latent class and latent variable model of travel behaviour. *Transportation (Amst.)* **2018**, *45*, 789–825. [CrossRef]

42. Echaniz, E.; dell'Olio, L.; Ibeas, A. Modelling perceived quality for urban public transport systems using weighted variables and random parameters. *Transp. Policy* **2018**, *67*, 31–39. [CrossRef]

43. Louviere, J.J.; Flynn, T.N.; Marley, A.A.J. *Best-worst Scaling: Theory, Methods and Applications*; Cambridge University Press: Cambridge, UK, 2015; ISBN 9781107337855|9781107043152.

44. Ho, C.Q.; Hensher, D.A. Application of irrelevance of state-wise dominated alternatives (ISDA) for identifying candidate processing strategies and behavioural choice rules adopted in best–worst stated preference studies. *J. Choice Model.* **2017**, *25*, 40–49. [CrossRef]

45. Mulley, C.; Hensher, D.A.; Rose, J. Do preferences for BRT and LRT vary across geographical jurisdictions? A comparative assessment of six Australian capital cities. *Case Stud. Transp. Policy* **2014**, *2*, 1–9. [CrossRef]

46. Hensher, D.A.; Mulley, C.N.; Rose, J.M. Understanding the relationshipbetween voting preferences for public transport and perceptions and preferences for busrapid transit versus light rail. *J. Transp. Econ. Policy* **2014**, *49*, 236–260.

47. Beck, M.J.; Rose, J.M. The best of times and the worst of times: A new best-worst measure of attitudes toward public transport experiences. *Transp. Res. Part A Policy Pract.* **2016**, *86*, 108–123. [CrossRef]

48. Echaniz, E.; Ho, C.Q.; Rodriguez, A.; dell'Olio, L. Comparing best-worst and ordered logit approaches for user satisfaction in transit services. *Transp. Res. Part A Policy Pract.* **2019**, *130*, 752–769. [CrossRef]

49. Marley, A.A.J.; Louviere, J.J. Some probabilistic models of best, worst, and best-worst choices. *J. Math. Psychol.* **2005**, *49*, 464–480. [CrossRef]

50. Dyachenko, T.; Reczek, R.W.; Allenby, G.M. Models of Sequential Evaluation in Best-Worst Choice Tasks. *Mark. Sci.* **2012**, *33*, 828–848. [CrossRef]

51. Greene, W.H. Nlogit 6 Guide. 2016.

52. McKelvey, R.D.; Zavoina, W. Ibm Fortran-Iv Program To Perform N-Chotomous Multivariate Probit Analysis. *Behav. Sci.* **1971**, *16*, 186.

53. McKelvey, R.D.; Zavoina, W. A statistical model for the analysis of ordinal level dependent variables. *J. Math. Sociol.* **1975**, *4*, 103–120. [CrossRef]

54. Rubin, D.B. Multiple imputations in sample surveys—A phenomeno-logical Bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*; American Statistical Association: Alexandria, VA, USA, 1978; pp. 20–28.

55. Rubin, D.B. Formalizing Subjective Notions about the Effect of Nonrespondents in Sample Surveys. *J. Am. Stat. Assoc.* **1977**, *72*, 538–543. [CrossRef]

56. Echaniz, E.; Ho, C.; Rodriguez, A.; dell'Olio, L. Modelling user satisfaction in public transport systems considering missing information. *Transportation (Amst.)* **2019**. [CrossRef]

57. van Buuren, S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **2007**, *16*, 219–242. [CrossRef] [PubMed]

58. Vuong, Q.H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **1989**, *57*, 307. [CrossRef]

59. Greene, W.H.; Hensher, D.A. *Modeling Ordered Choices: A Primer*; Cambridge University Press, Ed.; Cambridge University Press: Cambridge, UK, 2010; ISBN 9780521194204.

60. Bech, M.; Gyrd-Hansen, D. Effects coding in discrete choice experiments. *Health Econ.* **2005**, *14*, 1079–1083. [CrossRef] [PubMed]