Determining a public transport satisfaction index from user surveys

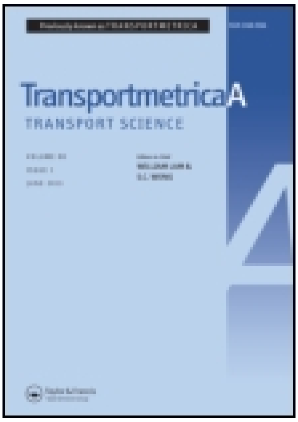
Article *in* Transportmetrica · January 2012

DOI: 10.1080/18128602.2011.654139

2 authors:

J. M. Del Castillo
Universidad de Sevilla

Francisco G Benitez
Universidad de Sevilla



Transportmetrica A: Transport Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ttra21

Determining a public transport satisfaction index from user surveys

J. M. del Castillo [a] & F. G. Benitez [b]

[a] Department of Transportation Engineering, Faculty of Engineering , University of Seville , Seville , Spain

[b] Camino de los Descubrimientos , Sevilla , E-41092 , Spain

Accepted author version posted online: 04 Jan 2012.Published online: 31 Jan 2012.

To cite this article: J. M. del Castillo & F. G. Benitez (2013) Determining a public transport satisfaction index from user surveys, Transportmetrica A: Transport Science, 9:8, 713-741, DOI: 10.1080/18128602.2011.654139

To link to this article: http://dx.doi.org/10.1080/18128602.2011.654139

Transportmetrica A: Transport Science, 2013
Vol. 9, No. 8, 713–741, http://dx.doi.org/10.1080/18128602.2011.654139

Determining a public transport satisfaction index from user surveys

J.M. del Castillo[a]* and F.G. Benitez[b]

[a]*Department of Transportation Engineering, Faculty of Engineering, University of Seville, Seville, Spain;* [b]*Camino de los Descubrimientos, Sevilla E-41092, Spain*

(*Received 24 February 2011; final version received 29 December 2011*)

The quality of public transport systems may be directly observed through user surveys by collecting ratings given by the users to specific aspects such as network coverage, transfers between lines and waiting time, among others. Besides these specific satisfaction ratings, an overall global satisfaction rating of the service is required. This way of proceeding, by asking questions on a limited number of disaggregated aspects, makes it easier to analyse the factors involved in the quality of a means of public transport and to rank these factors according to their contribution to the global satisfaction. This article presents a methodology to determine the relationship between the global satisfaction rating and the specific satisfaction ratings. This methodology employs three types of models for such a relationship: weighted means, a multivariate discrete distribution and a generalised linear model. These models allow the identification of the contribution of the specific satisfaction ratings to the global satisfaction rating. This information may be used by transit companies to improve their service quality.

Keywords: public transport; user satisfaction; quality survey; weighted mean; exponential family distribution; pseudolikelihood; generalised linear model

1. Introduction

Most public transport organisations have increasingly come to understand the importance of customer satisfaction. It is understood that it is less costly to keep existing customers than to gain new ones. Under the greatly restricted budgets available for increasing the quality of public transport systems, agencies and administrations must identify priorities for increasing the users' global satisfaction. The relationships between quality and satisfaction and between quality judgements and satisfaction judgements are complex. Nevertheless, in the service industry or tertiary sector, quality and satisfaction are intimately related. In the context of transportation, this implies that the identification of the factors reflecting the proper functioning of a transportation system is of key importance. Namely, the standard level of such factors should be improved in order to maximise the global satisfaction perceived by the users.

The evaluation of the significant factors in users' satisfaction is a priority issue for public transportation operators. The user satisfaction in a public transport system is defined as 'the overall level of attainment of a customer's expectation, measured as the percentage of the expectations actually fulfilled' (Tyrinopoulos and Antoniou 2008).

*Corresponding author. Email: delcastillo@us.es

This satisfaction level is an aggregate measure of the satisfaction perceived by the user for different aspects of the transportation system. In what follows, the overall or aggregate satisfaction will be called 'global satisfaction', whereas the satisfaction with respect to specific features of the transportation system will be called 'specific satisfactions'. A literature review was conducted to identify the most relevant features of the transportation system regarding the user satisfaction. It was found that trip duration, accessibility, fare, network connectivity, information, comfort, safety and employees' kindness were among the most investigated ones. Besides those, environmental impacts and sustainability have regularly been considered recently.

The contributions of this work are twofold: first, several models are proposed in order to estimate the global satisfaction from the specific satisfactions. Namely, three types of models are presented in this work: models based on weighted means, a model based on a multivariate discrete distribution and a generalised linear model. The multivariate discrete distribution has been derived from a wide class of such distributions that have a useful property relative to the conditional expectation. This derivation is an original contribution itself. A particular kind of models based on means are the block models, which are introduced in this work. Additionally, although the generalised model and the means employed in this work are well-known, their application in the field of transportation for the sake of modelling a satisfaction index is new. Finally, these three models are applied to a sample obtained from users of the public bus company of the Spanish city of Bilbao.

Second, the joint analysis of the quantitative relevance given to the specific items by the different models, allows the identification of the most relevant factors for the global satisfaction. This information may be used for setting a prioritisation of the transportation system features in accordance with their influence on the users' global satisfaction. Namely, the objective of the prioritisation is to improve those aspects that contribute to the global satisfaction most positively.

There are several reasons for this choice of models. The models based on means are simple and a natural choice when the dependent variable should lie between the minimum and the maximum of the independent variables. This is what happens with the global satisfaction and the specific satisfaction ratings. The multivariate discrete distribution and the generalised linear model have been selected because they not only give an estimate of the global satisfaction rating but also its entire probability distribution. In this sense, these models are more ambitious than the model based on means and it seems reasonable to compare the former with the latter.

The usefulness of finding an aggregated measurement of user satisfaction goes beyond having a single measure of the service quality perceived by the users. This type of measure may also be incorporated into a discrete choice model in order to account for the user satisfaction in the process of transport mode choice making. This has been done by Tam *et al.* (2010) who have proposed a discrete choice model for the ground access modal split pattern at the Hong Kong International Airport by means of the inclusion of a 'satisfaction' latent variable related to the perceived service quality by the air passenger. Another interesting research field is the study of the relationship between the quality perception from consecutive service stages. This has been investigated by Chiou and Chen (2011) for a low-cost airline in China.

This article is organised as follows. Section 2 reviews some recent developments in the field of public transport user satisfaction measurements. Section 3 explains how the transit survey was carried out. Section 4 presents the models based on the weighted power mean

(WPM) and on block means. In Section 5 the multivariate discrete distribution is derived and fitted to the sample of items ratings. In Section 6 a generalised linear model is employed for modelling the user satisfaction as a function of the specific items ratings. The estimation of the models yields a quantitative assessment of the influence on the global satisfaction of the factors whose quality levels are measured through each specific satisfaction. Finally, the results given by the above models are compared and analysed in Section 7. This analysis reveals that users' satisfaction with the bus service is concentrated on a few factors.

2. Brief state-of-the-art

Many efforts have been made to evaluate customer satisfaction in the last decade by assessing the quality of the service (TRB 1999, CEN 2002, TRB 2004) and identifying users' priorities. Some researchers concentrated on the features that were supposed to reflect the appropriateness of different significant factors, such as accessibility. This factor has been measured through various indices such as the distance or time to urban facilities and/or to working places (Nil and Naciye 2004, Mizuki and Akiko 2005). Other surveys on public bus services are those of Stradling *et al.* (2007) and Fellesson and Friman (2008); the first one conducted a survey measuring 68 items related to likes/dislikes of the bus network in Edinburgh, and the second compared the customers' perceived satisfaction with public transport services in nine European cities. The results showed that significant aspects contributing to users' satisfaction are timetable adequacy and service frequency, service reliability and information, bus stop design, staff skill and safety, among others. Former studies have also examined consumers' evaluation from two points of view, one related to the travel mode choice (Mann and Abraham 2006, Krizek and El-Geneidy 2007) and the second focusing on the mass transit market (Eboli and Mazzulla 2007). In this last work, a model based on structural equations examines the contribution that the service quality attributes to the global user satisfaction in a similar approach to that applied to public transport services earlier by Andreassen (1995) and Karlaftis *et al.* (2001). Other noteworthy works in this field are those of Friman (1998), Friman and Gärling (2001a, b) and Friman *et al.* (2001) in Sweden, Agarwal (2008) in India, Budiono (2009) in Indonesia, Wu *et al.* (2009) and Ji and Gao (2010) in China and Dell'Olio *et al.* (2010) in Spain, among others.

The problem of finding an aggregate measurement of a series of observations, measurements or scores appears in many scientific and engineering fields. That is why, it is hardly surprising that many different methods have arisen to tackle the aforementioned problem. In principle, these methods could be divided into two types depending on whether certain statistical hypotheses are made on the observations or not. The methods that are not derived from statistical hypotheses are quite diverse, ranging from methods based on the use of aggregation functions or operators to methods relying on fuzzy logic and neural networks, which are used by Wu *et al.* (2009). Beliakov (2003) carried out a brief review of the main aggregate operators. Means are a type of aggregation operators that is often employed. The study by Eboli and Mazzulla (2009) used this type of function to create a global satisfaction index. Another approach to find an aggregate measurement is that of Perng *et al.* (2010), who used grey relational analysis to measure satisfaction in airport retailing. Eboli and Mazzulla (2009) provide a very up-to-date list of works that use non-statistical methods to obtain global indices.

The methods based on statistical hypotheses are also relatively diverse. The majority of the works in this category rely on structural equations modelling as a method for obtaining the global satisfaction index. This is the case of the works by Friman (1998), Friman and Gärling (2001a, b), Friman *et al.* (2001), Karlaftis *et al.* (2001) and Eboli and Mazzulla (2007). Methods that use different types of regression analyses are also encountered: Agarwal (2008) and Budiono (2009) employed linear regression whereas Ji and Gao (2010) used a multilevel logistic regression. Other types of modelling, such as an ordered choice model (Dell'Olio *et al.* 2010), have also been proposed. Besides trying to find the aggregate measurement, a usual task among statistical methods is the reduction of the sample dimensionality. Principal component analysis and factor analysis are widely used for this task. This approach to dimensionality reduction has been followed by Friman and Gärling (2001a, b), Friman *et al.* (2001) Eboli and Mazzulla (2007), Krizek and El-Geneidy (2007), Agarwal (2008) and Budiono (2009).

Three models are proposed in this study to tackle the problem of finding a measurement of users' global satisfaction with public transport. The first model can be classed within the models that use an aggregation function to find a global index value. The aggregation functions employed are different types of means. The second model proposed is a new multivariate discrete distribution that aims at describing the sample consisting of the ratings assigned to the specific satisfactions and to the global satisfaction. The third model is a generalised linear model that relates, within a statistical framework, the global satisfaction to the specific satisfactions. The model based on means is simple and intuitive. That is why it can be used as a benchmark with which to compare the results of the second and third models. Due to the fact that the second and third models are based on statistical distributions, they not only have the advantage of providing not only the global index estimate, but also the distribution of this index and, particularly, confidence intervals for this estimate. To the authors knowledge, the three models presented in this work have never been applied for modelling a user satisfaction index.

3. Transit satisfaction survey

In May 2010 a survey was conducted among 1508 users of the public bus company in Bilbao, a city located in the north of Spain, in order to determine the quality of the bus service. The survey was carried out by boarding operative buses on working days and interviewing some randomly selected passengers. Therefore, the respondent population corresponds to bus users. The survey included most of the lines of the bus network.

The questionnaire was divided into three sections. The first one collected personal information related to the user: age, sex, driving license, access to private transport mode and recommendation to use public bus transport and demographic sector. The second section collected information related to the passenger behaviour such as commuting pattern, number of commute days in a week, purpose, travel time, distance of travel, transport choice, public bus transport use pattern and fare and passes class. The third section consisted of 35 questions associated with various aspects of the service, such as frequency, travel time, punctuality, price, information, cleanliness, staff behaviour, comfort, bus stop security and condition, safety and information. In what follows, each of these aspects will be referred to as an item.

People were asked to choose from 11 levels of satisfaction in the integer interval {0–10}, with a value of 10 denoting the highest level of satisfaction. These ratings will be referred to as specific satisfactions (SSs) in the rest of this article. The specific satisfactions are grouped into eight categories or blocks. The list of blocks and their corresponding service aspects are given in Table 1. Finally, the users were asked to rate their overall satisfaction with the bus service, employing the same values and ordering, that is, 0 to 10.

Table 1. List of items rated by the users.

Block	Item
1. Connectivity	1. Connection to lines of the same operator
	2. Connection to lines of other operators
	3. Line diversity (number of lines of the transit network)
2. Accessibility	4. Accessibility of the bus network (number of bus stops)
	5. Reduced mobility users' accessibility
	6. Adequacy of the most used bus stop location
3. Information	7. Service information availability
	8. Availability of timetables and line plans
	9. Line information explicitness
	10. Information panels on terminals and bus stops
	11. Information panel on next stop
	12. Information on passes and tariffs
4. Time satisfaction	13. Bus punctuality
	14. Service frequency
	15. Trip duration
	16. Line reliability
	17. Service time window
5. User Attendance	18. Driver kindness
	19. Staff kindness
6. Comfort	20. Physical state of vehicles (quality, conservation, new/old)
	21. Bus cleanliness
	22. Bus comfort
	23. Bus illumination
	24. Bus temperature adequacy
	25. Average user volume (occupancy)
	26. Professionality/caution/driver skillfulness
	27. Bus stop coziness (weather conditions)
	28. Bus stop conservation and cleanliness
	29. Bus stop illumination
	30. Adequate visual arrival of buses at bus stops
7. Security/Safety	31. Bus safety (vehicles)
	32. Security on buses
	33. Bus stop safety
8. Environmental impact	34. Noise
	35. Bus contribution to traffic fluidity

The given rating is then the global satisfaction (GS). A Cronbach's alpha value of 0.94667 was indicative of the internal consistency of the responses. This value is due to the high correlation between the SS ratings. Namely, the linear correlation coefficients between the specific satisfactions lie between 0.193 and 0.694.

4. Models of global satisfaction based on means

4.1. *Weighted power means*

For all the observations of the sample, the ratings of the GS lie between the minimum and the maximum of the SS ratings. Obviously, this fact stems from using similar value scales for the specific satisfactions and the global satisfaction and if individuals did not comply with them, this would mean that their answers lacked consistency. For this reason, the use of a mean as a model for the global satisfaction is called for. It should be mentioned that models based on means may not be appropriate if one wants to include socioeconomic variables in the global satisfaction index. If such variables need to be considered, one could segment the sample according to the different categories of individuals. In this case, one should have some caution with the resulting sample size.

Before introducing the models employed, we will briefly summarise several properties of a mean. A mean M of a set of real numbers $x_1, x_2 \ldots, x_d$ is a function that satisfies the following property (Bullen 2003):

$$\min(x_1, x_2 \ldots, x_d) \leq M(x_1, x_2 \ldots, x_d) \leq \max(x_1, x_2 \ldots, x_d)$$

Furthermore, a mean is homogeneous if

$$M(\lambda x_1, \lambda x_2 \ldots, \lambda x_d) = \lambda M(x_1, x_2 \ldots, x_d)$$

The choice of a model based on a homogeneous mean makes sense since it is reasonable to assume that when the rating of the questions is scaled by a constant, the global satisfaction rating is also scaled by the same scalar. Finally, for the sake of flexibility of the predictive models and with the likely fact that individuals place higher relevance on certain factors rather than others, it is important for the model to contain relative weightings for each of the items. A simple mean model that fulfils the previous requirements is the WPM:

$$M(x_1, x_2 \ldots, x_d) = \left(\sum_{k=1}^{d} w_k x_k^{\beta} \right)^{1/\beta}$$

where the weights w_k should sum one: $\sum_{k=1}^{d} w_k = 1$. In principle, the exponent β can take positive and negative values. The WPM has the following limit expressions:

$$\lim_{\beta \to 0} M(x_1, x_2 \ldots, x_d) = \prod_{k=1}^{d} x_k^{w_k}$$

$$\lim_{\beta \to \infty} M(x_1, x_2 \ldots, x_d) = \max(x_1, x_2 \ldots, x_d)$$

$$\lim_{\beta \to -\infty} M(x_1, x_2 \ldots, x_d) = \min(x_1, x_2 \ldots, x_d)$$

The appeal of the weighted power mean, apart from being homogeneous, is the fact that it incorporates as particular cases the arithmetic mean (for $\beta = 1$), the geometric mean (for $\beta = 0$) and the harmonic mean (for $\beta = -1$). Additionally, it includes, as limit cases, similar valid models such as the minimum and maximum, which might be potential representatives of the reality. The application of the former mean to predict the global satisfaction rating by the *i*-th individual produces the following predictive models:

(a) ***WPM model***. This corresponds to a WPM with a structure defined by

$$\hat{y}_i = \left(\sum_{k=1}^{d} w_k x_{ik}^{\beta} \right)^{1/\beta}$$

where the weights w_k and the exponent β are the parameters to be estimated. The prediction of the global satisfaction for the *i*-th individual is $\hat{y}_i$ and x_{ik} are the specific satisfaction ratings for the k item given by the *i*-th individual.

(b) ***WAM model***. This is a weighted arithmetic mean arising from the WPM model when taking $\beta = 1$.

(c) ***EAM model***. This is an equally weighted arithmetic mean with an expression derived from the WPM model with $\beta = 1$ and identical weights: $w_k = 1/d, k = 1, \ldots, d$.

Recalling that the questionnaire comprises 35 items, that is, $d = 35$, the WPM has a total of 36 parameters and a linear constraint for 35 of them. The estimation of these parameters was carried out by minimising the mean absolute error ε_{abs} obtained by calculating the absolute value of the deviation between the estimated and observed values of the global satisfaction rating. The mean absolute error is given by

$$\varepsilon_{\text{abs}} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{\text{abs}_i} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{1}$$

where y_i is the global satisfaction rating given by the *i*-th individual and n is the number of individuals who answered the survey. The error form defined in expression (1) was chosen as it best reflects the distance between the global satisfaction rating given by the individual and the estimated one. A quadratic error was rejected so as not to place too much weighting on the individuals whose global satisfaction rating is further away from its estimate.

The results of fitting the previously defined models (WPM, weighted arithmetic mean (WAM) and EAM) to the sample reveal that the best model, for the global satisfaction as a function of the single satisfactions, corresponds to the WPM with an exponent $\beta = 1.0122$. This model yields a minimum value of 0.85359 for the mean absolute error. Figure 1 shows the mean absolute error versus the exponent β for the weights values that minimise it.

Table 2 lists the mean absolute error for each model. Table 3 lists the weights (in %) for the WPM and WAM models, associated with the index or item number of the specific satisfaction. This number ranges from 1 to 35, and the corresponding items are shown in Table 1. The model WPM yields the best results among all the models, followed very closely by the WAM model. These two models provide almost identical rankings of the relevance of the SS independent variables. A negligible difference can be appreciated in the

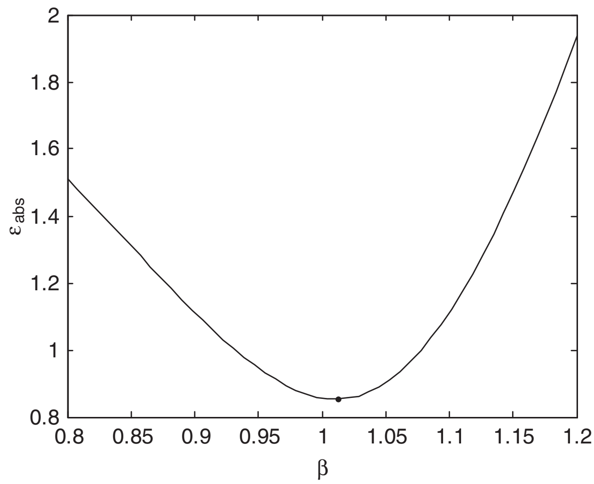


Figure 1. Mean absolute error ε_{abs} as a function of the exponent β in the WPM model.

Table 2. Errors given by the WPM, WAM and EAM models.

Model	Absolute mean error ε_{abc}
WPM with optimal $\beta = 1.0122$	0.85359
WAM	0.85625
EAM	0.95593

alternating positions of questions 3 and 4, and 5 and 15, as Table 3 shows. Additionally, the weight values are almost identical for the WPM and WAM models, since the optimal value of $\beta = 1.0122$ for the former is close to unit. It can be seen that from the original 35 items only 18 are present in Table 3. Namely, the items whose weights are smaller than 10^{-3} or 0.1% have been discarded from the model. The right-hand column of Table 3 shows the cumulative weight up to the corresponding item, with the weights given by the optimal WPM model.

Table 4 shows the grouping of the relevant specific satisfactions into the blocks set out in the questionnaire. In each block, the items have been arranged in decreasing order according to the weights obtained with the WPM and WAM models. It is remarkable that from the original eight blocks listed in Table 1, only five remain as the most relevant ones. Furthermore, the fourth column shows the percentage of the weight of each question within its block. This block weight has been obtained from the sum of the individual weights of the corresponding questions. It can be observed that the block 'information' hardly contributes to the global satisfaction. Additionally, the blocks not listed in Table 4 are irrelevant. These are 'user attendance', 'security/safety' and 'environmental impact'.

Table 3. Weights for the items given by the WPM and WAM models, listed in decreasing order.

Index of item	Weights ($w_k x$ 100) WPM model ($\beta = 1.0122$)	WAM model ($\beta = 1.0$)	Cumulative weight (%) for WPM model
16	17.0340	17.0370	17.0
6	12.2920	12.3090	29.3
23	10.2860	10.3480	39.7
13	9.9989	9.9815	49.7
2	9.6967	9.7140	59.4
14	8.0060	8.0172	67.4
3 (WPM) or 4 (WAM)	5.1693	5.1656	72.6
4 (WPM) or 3 (WAM)	5.1664	5.1317	77.7
17	4.7043	4.6899	82.4
24	4.0592	4.0158	86.4
1	3.2048	3.1920	89.6
21	3.1873	3.1761	92.8
25	2.0468	2.0311	94.8
26	1.3519	1.3795	96.2
20	1.2660	1.2759	97.5
5 (WPM) or 15 (WAM)	1.0070	1.0258	98.5
15 (WAM) or 5 (WPM)	0.9861	0.9904	99.5
7	0.5370	0.5196	100.0

Table 4. Weight of the most significant blocks given by the WPM model.

Block	Block weight (×100)	Item	Weight within block (×100)
4. Time satisfaction	40.73	16	41.82
		13	24.55
		14	19.66
		17	11.55
		15	2.42
6. Comfort	22.20	23	46.34
		24	18.29
		21	14.36
		25	9.22
		26	6.09
		20	5.70
2. Accessibility	18.47	6	66.57
		4	27.98
		5	5.45
1. Connectivity	18.07	2	53.66
		3	28.61
		1	17.73
3. Information	0.53	7	100.00

4.2. *Definition of block models*

Based on the analysis of the previous section, and in order to reduce the number of parameters of the model, a family of new models named 'block models' is now introduced. The formulation is also based on terms of a mean, but in this new approach, the mean is not for all the items ratings but only for particular means of each block. In other words, the block model is a mean of means. This formulation aims at reducing the number of parameters that should be estimated by considering the influence of each block as a whole. Namely, the global satisfaction prediction is given by a WPMof the block ratings:

$$\hat{y}_i = \left(\sum_{b=1}^{B} w_b z_{ib}^{\beta} \right)^{1/\beta} \tag{2}$$

where z_{ib} is the mean rating corresponding to block b, for individual i-th, B is the number of blocks and w_b is the weight associated with block b. In the above model, the block weights are parameters to be estimated and should comply with

$$\sum_{b=1}^{B} w_b = 1.$$

The exponent β must also be estimated and captures the model nonlinearity.

In expression (2), the block means z_{ib} are not fully defined. In principle, we could also consider that each block mean is the power-weighted mean of the block items' ratings. However, this would dramatically increase the number of parameters and could easily lead to overfitting. In order to simplify the formulation, two alternatives for the block means have been taken into account. These are the maximum and the arithmetic mean of the ratings given to the block items.

For the first choice, the rating of block b would be given by

$$z_{ib} = \max(x_{ik},\ k \in K(b)) \tag{3}$$

where $K(b)$ is the set of items of b-th block. For the second alternative, the rating of block b is simply the arithmetic mean of all the ratings given to the block items

$$z_{ib} = \frac{1}{k_b} \sum_{k \in K(b)} x_{ik} \tag{4}$$

where k_b stands for the number of items in block b.

The use of the maximum for the block ratings may be justified by the existence of a kind of behaviour that might be defined as a 'low level of exigency'. Users showing this behaviour could also be viewed as optimistic users. For this kind of user, the rating of the block would be the same as that assigned to the best rated item. In other words, the user's overall satisfaction level for the block items would depend on the ratings given to them, but it is just equal to the maximum rating. Therefore, it would be insensitive to the rating given to the rest of the items of the block provided that they are below the maximum rating. For this reason, this kind of behaviour would be characteristic of a low level of exigency of the user, since the lowest ratings would not affect the final block rating. Finally, the alternative to the maximum for the block rating is the arithmetic mean, which seems to be a natural choice.

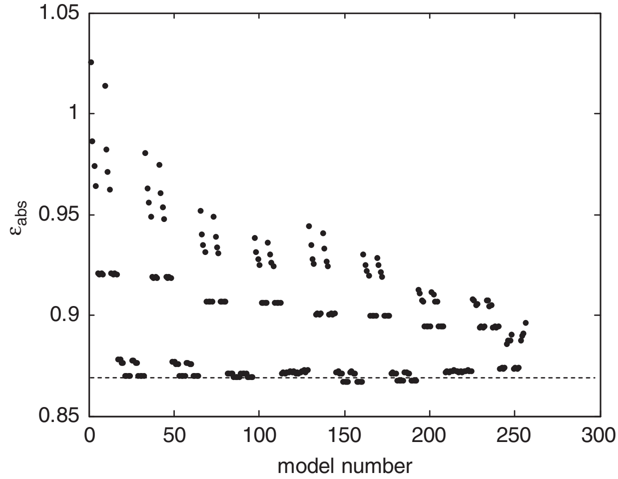


Figure 2. Mean absolute error given by the 256 models of type MMB8 with $\beta = 1$.

Based on expressions (2)–(4), a model for the eight blocks of the questionnaire has been estimated. This model has been called MMB8, an abbreviation for maximum–mean–block. For each block, the two former alternatives can be chosen to characterise its overall rating. Thus, a total of $2^B = 2^8 = 256$ MMB8 models are possible. The fitting of any of those models requires the estimation of 8 weights and the exponent β. For the sake of simplicity, the value of the parameter β was chosen from the values 0.9, 0.95, 1, 1.05 and 1.1, resulting in a final number of 5×256 MMB8 models. The goodness-of-fit is again measured by the mean absolute error, which is given by expression (1). The minimum mean absolute error of all the 256 models for a fixed value of β is obtained for $\beta = 1$. Therefore, this value has been considered in what follows.

Figure 2 shows the absolute residual mean error, for case $\beta = 1$, obtained with the MMB8 model. In order to depict the outcomes, the 256 models have been codified with a number of 8 binary digits. Each digit corresponds to one of the eight blocks. A value of 0 for the digit representing block b indicates that the chosen model for the block satisfaction is the maximum rating assigned to the questions of the block, that is, expression (3) gives the block rating. On the contrary, a value of 1 means that the arithmetic mean given by (4) has been adopted for the block rating. Finally, the model code is transformed into a decimal number and a unit added, such that the model numbering varies from 1 to 256. This number is the variable depicted along the horizontal axis of Figure 2. One can observe that there are 16 models that yield an error smaller than 0.87. These models are listed in Table 5. The errors of the best eight models (those numbered 149, 150, 151, 152, 157, 158, 159 and 160) are identical up to the fifth decimal cipher. On the other hand, the errors produced by models 181, 182, 183, 184, 189, 190, 191 and 192 are the same up to the third decimal cipher. To sum up, any of those models could be adopted as a valid one in practice. Moreover, the generated error is only slightly larger than that of the WPM model with $\beta = 1.0122$ (0.85359) and that of the WAM model (0.85625). Therefore, any of the MMB8 models with $\beta = 1$ listed in Table 5 constitutes a valid alternative to the WPM and WAM models. The WPM model yields a slightly smaller error, but requires the estimation

Table 5. Mean absolute error given by the best 16 models of type MMB8 with $\beta = 1$.

Model number	Model code	Error	Model number	Model code	Error
151	10010110	0.86722	183	10110110	0.86768
159	10011110	0.86722	189	10111100	0.86768
157	10011100	0.86722	181	10110100	0.86768
149	10010100	0.86722	191	10111110	0.86768
160	10011111	0.86722	182	10110101	0.86769
158	10011101	0.86722	184	10110111	0.86769
150	10010101	0.86722	192	10111111	0.8677
152	10010111	0.86722	190	10111101	0.8677

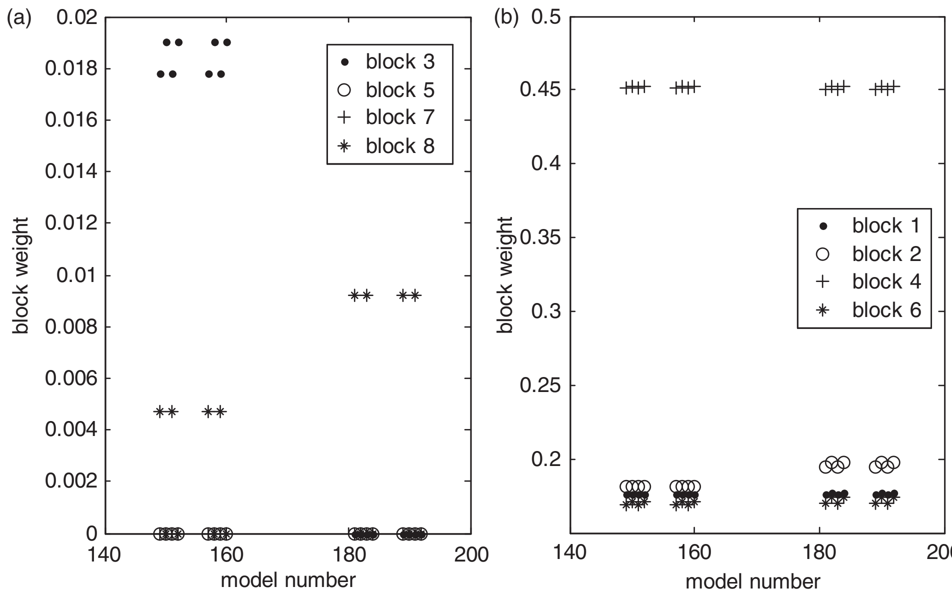


Figure 3. Weights of the least significant blocks (a) and of the most significant blocks (b) given by the 16 best MMB8 models with $\beta = 1$.

of 35 parameters versus the 7 parameters (weights of the eight blocks of questions minus 1 due to the unit sum constraint) of the MMB8 models.

The block weights w_b of the best 16 models MMB8 with $\beta = 1$ of Table 5 are represented in Figure 3. By examining the values of these weights, we can draw the conclusion that the weights of blocks 3, 5, 7 and 8 are much less significant than those of the remaining blocks. In fact, as Figure 3(a) shows, most of the weights of these blocks are smaller than 0.01 and only the weights of block 3 are slightly below 0.02. These values are much reduced in comparison with an equally weighted distribution of 0.125 among the 8 blocks. On the other hand, the weights of blocks 1, 2, 4 and 6 are always above 0.15, as can be seen from Figure 3(b). In particular, block 4 is the most significant block since it accounts for almost half of the weighting. This is not surprising since the valuation of the

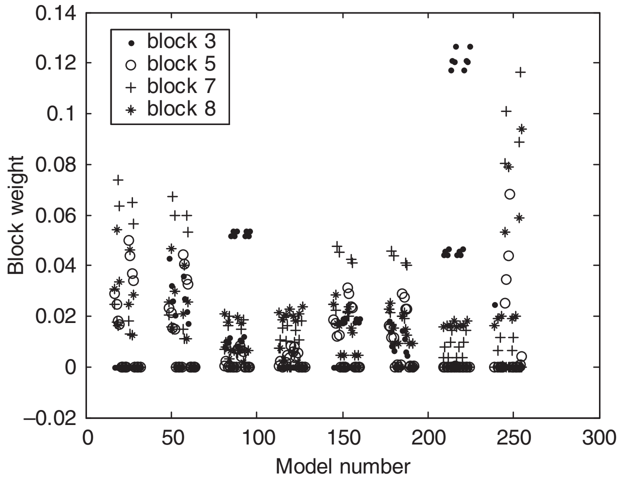


Figure 4. Weights of the least significant blocks given by the 128 best MMB8 models with $\beta = 1$.

overall satisfaction is likely to be highly sensitive to items within this block, such as service punctuality, frequency and reliability and trip duration (Table 1).

Figure 4 confirms the fact that blocks 3, 5, 7 and 8 are less significant, not only for the 16 best models but for the whole set of possible MMB8 models. In Figure 4, the weights of blocks 3, 5, 7 and 8 corresponding to the MMB8 models with $\beta = 1$ are displayed along the vertical axis, whereas the horizontal axis identifies the model by its number. For the sake of clarity, only the best 128 models have been represented from a total amount of 256. The vertical bands without points correspond to those 128 models with higher errors, which have been discarded. In conclusion, Figure 4 clearly indicates that blocks 3, 5, 7 and 8 are not significant due to their weights being much smaller than those yielded by the equally weighted model (0.125). The conclusions drawn are essentially the same for the other values of the exponent β and have not been included in order to make the exposition brief.

4.3. *Results obtained with the four block models*

The analysis carried out in the previous section suggests that MMB models should be fitted with a smaller number of parameters by reducing the number of significant blocks. Based on the previous results, a new family of MMB models was tested in which the intervening blocks are the significant ones (blocks 1, 2, 4 and 6). The number of fitted models is then $2^4 = 16$. For this family, named MMB4, the expression of the mean is given by (2), with $B = 4$. The model MMB4 was also fitted by considering the following values for β: 0.9, 0.95, 1.05 and 1.1. The minimum mean absolute error of all the 16 models for a given value of β is achieved with $\beta = 1$ and consequently this value was finally selected.

The errors obtained after the model fitting are listed in increasing order in Table 6 for the 16 MMB4 models with $\beta = 1$. The code numbering follows the same rule as in the

Table 6. Mean absolute error given by the 16 models of type MMB4 with $\beta = 1$.

Model number	Model code	Error	Model number	Model code	Error
12	1011	0.86770	16	1111	0.89728
4	0011	0.87046	10	1001	0.90063
7	0110	0.87175	6	0101	0.90745
8	0111	0.87292	13	1100	0.92125
11	1010	0.87332	2	0001	0.92132
15	1110	0.87423	5	0100	0.96599
3	0010	0.88018	9	1000	0.96698
14	1101	0.89482	1	0000	1.11900

Table 7. Block weights for the best six models of type MMB4 with $\beta = 1$ model code.

Block	1011	0011	0110	0111	1010	1110	Average
1. Connectivity	0.17675	0.14986	0.17284	0.23215	0.21626	0.18750	0.18923
2. Accessibility	0.19739	0.15108	0.20688	0.16796	0.16425	0.18309	0.17844
4. Time satisfaction	0.45194	0.48971	0.53335	0.45600	0.55447	0.47680	0.49371
6. Comfort	0.17392	0.20935	0.08693	0.14390	0.06503	0.15261	0.13862

previous cases. The smallest error corresponds to model 12, with a code 1011, indicating that the prediction follows the expression

$$\hat{y}_i = \frac{w_1}{k_1} \sum_{k \in K(1)} x_{ik} + w_2 \max(x_{ik}, k \in K(2)) + \frac{w_3}{k_3} \sum_{k \in K(3)} x_{ik} + \frac{w_4}{k_4} \sum_{k \in K(4)} x_{ik} \tag{5}$$

where, in this case, $K(1) = \{1, 2, 3\}$ are the questions of block 1, $K(2) = \{4, 5, 6\}$ those of block 2, $K(3) = \{13, 14, 15, 16, 17\}$ those of block 4 and $K(4) = \{20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30\}$ those of block 6. From here, $k_1 = k_2 = 3$, $k_3 = 5$ and $k_4 = 11$ in expression (5).

The block weights estimated for the best 6 models are listed in Table 7. One can observe that these weights do not vary noticeably among the 6 models. Only the weight of block 6 for the models coded 0110 and 1010 is almost half that of the rest of the models. Due to the consistency of the weight values across these 6 models, they can be considered as robust models. Additionally, as can be observed in Table 6, the errors of these 6 models are smaller than 0.875. This value can be compared with those errors yielded by the MMB8 with $\beta = 1$ models (Table 5), deducing finally that MMB4 models with $\beta = 1$ are a preferable alternative to MMB8 as they contain a smaller number of parameters (3 versus 7) and lead to similar errors. In summary, the answers of the individuals to questions of blocks 3, 5, 7 and 8 can be discarded to predict the global satisfaction rating.

In relative terms, the final choice of the model is somewhat arbitrary, as the 6 best models of Table 6 are equally valid and lead to an error not larger than 0.875. A particular choice can be made based on the model robustness. This factor can be evaluated through the frequency of a binary digit appearing in the model code. For instance, if the choice has to be made among the 3 best models, in the first binary digit, the 0 appears in two models, 4 and 7, whereas the unit appears only in model 12; this makes the 0 the most robust (or representative) choice. Applying the same reasoning to the 3 following binary digits, one can obtain that for the second digit, the value that appears in most models is 0. For the third and fourth binary digits, the most repeated value is 1 in both cases. Therefore, the most robust model would be the one coded as 0011. By repeating this procedure and selecting the more representative digits among the first 5 models, it is easy to see that the same result emerges. Therefore, it is sensible to choose model 0011 as the one that fulfils this robustness criterion and that has a number of parameters and error in a most balanced way. For this model the estimated global satisfaction would be given by

$$\hat{y}_i = w_1 \max(x_{ik}, k \in K(1)) + w_2 \max(x_{ik}, k \in K(2)) + \frac{w_3}{k_3} \sum_{k \in K(3)} x_{ik} + \frac{w_4}{k_4} \sum_{k \in K(4)} x_{ik}$$

The results presented in this section correspond to a unit exponent in (2). Similar conclusions can be drawn values for the remaining values of β.

5. Modelling the survey using a discrete distribution

5.1. *Formulation of the model*

In the previous sections a series of models has been proposed in which the global satisfaction rating was predicted using a model based on the power weighted mean of the specific satisfaction ratings or one of the blocks' average ratings. In the following sections a radically different way of facing the problem of predicting the GS rating is presented. This approach is considerably more ambitious than the previous one since not only can the GS rating be predicted, but the objective is also to fit a multivariate discrete distribution to the sample made up of the users' answers in the survey. Specifically, the sample is comprised of the GS and SS ratings in such a way that the i-th row of the $\mathbf{n}$ sample matrix is made of the i-th individual's ratings and can be written as

$$\mathbf{n}_i = (y_i, x_{i1}, x_{i2}, \ldots, x_{id})$$

The same notation has been used in the above expression as in the previous sections: x_{ik} is the rating given to the specific satisfaction related to item k and y_i is the rating given to the global satisfaction by individual i.

The aim of the new approach is to find the expression of the probability that an individual answers the survey in a certain way. Namely, the answers or ratings are considered as random variables that can take the values of the scale employed in the survey, which are the integers from 0 to 10. Therefore, the value of the probability mass function (PMF) will be

$$f(y_i, x_{i1}, x_{i2}, \ldots, x_{id}) = \mathrm{Prob}(Y = y_i, X_1 = x_{i1}, X_2 = x_{i2}, \ldots, X_d = x_{id}) \tag{6}$$

Following the hypothesis that the survey carried out on n individuals is the result of n independent and identically distributed random variables, the sample log-likelihood will be

$$\ell = \sum_{i=1}^{n} \log(f(y_i, x_{i1}, x_{i2}, \ldots, x_{id})) \tag{7}$$

The prediction of any variable when the remaining ones are known is given by the conditional expectation of that variable conditioned on the rest of the variables. In particular, the prediction of the GS rating of the i-th individual when the SS ratings are known will be expressed as

$$\hat{y}_i = E(Y|X_1 = x_{i1}, X_2 = x_{i2}, \ldots, X_d = x_{id}) = \frac{\sum_{y=0}^{10} yf(y, x_{i1}, x_{i2}, \ldots, x_{id})}{\sum_{y=0}^{10} f(y, x_{i1}, x_{i2}, \ldots, x_{id})} = m(x_{i1}, x_{i2}, \ldots, x_{id}) \tag{8}$$

In short, it is obvious that the validity of this process as a method for predicting the GS rating will depend on the goodness-of-fit obtained by adopting a specific discrete distribution and fitting it to the available sample. Therefore, the prediction accuracy rests entirely on choosing the correct distribution. One could argue that multinomial distribution would be the best choice, given that it is the most conventional discrete multivariate distribution with bounded support. However, this would be an incorrect choice since the covariance between any of the variables would always be negative. On the contrary, the sample covariance between any pair of variables of the sample is positive for all the possible pairs. This result was expected and it is due to the fact that the satisfaction with any two specific aspects is positively correlated. Therefore, the distribution to be adopted must not restrict the variables' sign of the covariance between the variables, meaning that this can be positive as well as negative.

A rather general result was put forward by Becker and Utev (2002), concerning the analytical form of the conditional expectation of a function of a variable conditioned on the rest of the variables. The result is valid in certain discrete multivariate distributions that are characterised by the fact that the dependency between the variables arises from the cross product terms. Specifically, Becker and Utev (2002) demonstrate the aforementioned result for those distributions that have the following form:

$$\mathrm{Prob}(N_1 = n_1, N_2 = n_2, \ldots, N_d = n_d) = f(n_1, n_2, \ldots, n_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = C(\boldsymbol{\alpha}, \boldsymbol{\theta}) \exp(-\mathbf{nAn^T}) \prod_{i=1}^{d} f_i(n_i; \theta_i) \tag{9}$$

where $\mathbf{n} = (n_1, n_2, \ldots, n_d)$ is the vector of the d variables. Moreover, $f_i(n_i; \theta_i)$ are univariate distributions that depend only on the parameter θ_i and have the same support as that of the i-th variable. $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ is the vector of d parameters that appears in the distributions f_i, and $\mathbf{A}$ is a symmetric matrix with elements α_{ij} and a diagonal of zeroes, i.e. $\alpha_{ii} = 0$. The elements of matrix $\mathbf{A}$ that are not zero are grouped together in the $\boldsymbol{\alpha}$ vector of $d(d-1)/2$ elements. The factor C depends on all the parameters and is a normalisation factor needed so that the sum of all the possible values of the probability mass function is equal to one. In the following the notation $\mathbf{n}_{-i}$ is employed for the vector of the $d-1$ random variables resulting from the elimination of the variable i, that is

$$\mathbf{n}_{-i} = (n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_d)$$

Becker and Utev (2002) show that the following relationship is satisfied by the distributions whose form corresponds to (9):

$$\mathrm{E}(H(N_i)|\mathbf{n}_{-i}) = \mathrm{E}(H(N_i)|N_k = n_k; k = 1, \ldots, d; k \neq i) = h\left(\sum_{k=1}^{d} \alpha_{ik} n_k\right) \tag{10}$$

In other words, the conditional expectation of a function of a variable is not an arbitrary function of the values that the rest of the variables take, but instead a function of the linear combination $\sum_{k=1}^{d} \alpha_{ik} n_k$. In (10), $\alpha_{ii} = 0$ must be considered so that the conditional expectation depends only on the conditioning variables. In principle and without giving more details of the expression of the distribution, it is not possible to establish any relationship between the functions H and h. For the survey object of this work, the prediction of the GS rating in accordance with the SS ratings can be expressed simply according to the previous result, assuming that the distributions of the global and specific ratings are of the type defined in (9).

5.2. *Choice of the discrete distribution*

In the following section we shall assume that the variables can take integer values between 0 and m, although the assumption that the variable N_i would take integer values between 0 and m_i could be made without any additional complications. In accordance with the generic expression of distribution (9), the variables are independent if the elements of matrix A are all zero. In this case the joint probability distribution is simply the product of the distributions $f_i(n_i; \theta_i)$. Therefore, a natural choice for these distributions is the binomial distribution of parameters m and θ_i, or in other words

$$f_i(n_i; \theta_i) = \binom{m}{n_i} \theta_i^{n_i} (1 - \theta_i)^{m - n_i} \tag{11}$$

By introducing the previous expression into (9), the following expression is obtained for the joint probability distribution:

$$f(n_1, n_2, \ldots, n_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = C(\boldsymbol{\alpha}, \boldsymbol{\theta}) \exp(-\mathbf{nAn}^{\mathbf{T}} + \boldsymbol{\pi} n^{\mathbf{T}}) \prod_{i=1}^{d} \binom{m}{n_i} (1 - \theta_i)^m \tag{12}$$

where the elements of the vector of parameters $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_d)$ are

$$\pi_i = \log\left(\frac{\theta_i}{1 - \theta_i}\right) \tag{13}$$

The expression of the joint probability with the new parameters is finally

$$f(n_1, n_2, \ldots, n_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = D(\boldsymbol{\alpha}, \boldsymbol{\pi}) \exp(-\mathbf{nAn}^{\mathbf{T}} + \boldsymbol{\pi} n^{\mathbf{T}}) E(m, \mathbf{n}) \tag{14}$$

In this expression the factor E does not depend on the parameters π_i and α_{ik} but only on the values of the random variables and the maximum values that they can take, in this case m.

The choice of distribution (14) is quite natural since it is an exponential family distribution whose sufficient statistics are the sample moments of N_i and $N_i N_j$. The distribution has a clear interpretation in terms of information content. Indeed, if only the means of each variable of the sample are known and we do not want to make any additional hypotheses on the data, the distribution would be chosen by considering that the variables N_i are independent and binomial of parameters m and θ_i. This is the same as considering that $\alpha_{ik} = 0$ in (14). If more information on the sample becomes available and in particular the sample values of the cross-moments of $N_i N_j$ are known, the least prejudiced choice for the distribution would be the distribution given by (14). The dependency between the variables is reflected here by the new information available.

The distribution given in (14) provides results that can be handled analytically at least in what concerns the conditional distribution of a variable given the others. Namely, the conditional distribution of the variable *i*-th may be written as follows:

$$\mathrm{Prob}(N_i = n_i|\mathbf{n}_{-i}) = \frac{\binom{m}{n_i}(1-\tilde{\theta}_i)^m \exp(n_i\tilde{\pi}_i)}{(1-\tilde{\theta}_i)^m \sum_{n=0}^{m}\binom{m}{n}\exp(n\tilde{\pi}_i)} = \binom{m}{n_i}(1-\tilde{\theta}_i)^m \exp(n_i\tilde{\pi}_i) \qquad (15)$$

Expression (15) results from the consideration that it is simply a binomial distribution of parameter $\tilde{\theta}_i$ and that parameters $\tilde{\pi}_i$ and $\tilde{\theta}_i$ are related to the original parameters in the following way:

$$\tilde{\pi}_i = \pi_i - \sum_{k=1}^{d}\alpha_{ik}n_k = \log\left(\frac{\tilde{\theta}_i}{1-\tilde{\theta}_i}\right) \qquad (16)$$

and

$$\log(1-\tilde{\theta}_i) = -\log(1+e^{\tilde{\pi}_i}) = -\log\left(1+\exp\left[\pi_i - \sum_{k=1}^{d}\alpha_{ik}n_k\right]\right) \qquad (17)$$

Therefore, the mean of conditional distribution (15) is the mean of a binomial distribution of parameter $\tilde{\theta}_i$ and this finally leads to

$$E(N_i|\mathbf{n}_{-i}) = m\tilde{\theta}_i = \frac{m}{1+e^{-\tilde{\pi}_i}} = \frac{m}{1+\exp\left[-\pi_i + \sum_{k=1}^{d}\alpha_{ik}n_k\right]} \qquad (18)$$

To the authors' knowledge, distribution (14) has never been used before in the field of transportation. The majority of applications of (14) to model discrete and bounded jointly distributed random variables have been carried out exclusively for binary variables, or in other words, for cases where $m = 1$ and the variables only take the values 0 or 1. In such cases, the distribution is called a quadratic binary distribution. This distribution has been used by Cox and Wermuth (2002) in studies in the area of social sciences. Another important use of the distribution in the binary case is for modelling random graphs (Van Duijn *et al.* 2009).

In the binary case ($m = 2$) the parameter $\tilde{\pi}_i$ is the logit or log odds ratio of the binary variable conditioned on the rest of the variables. Expression (16) shows that the proposed distribution leads to a linear regression model for the logit of the conditional distributions. Arnold *et al.* (1999) showed that the distribution expressed in (14) is the only one where the logit of the conditional distribution depends linearly on the rest of the variables. Moreover, Joe (1997) studied certain properties of the particular case of the binary distribution where the variables can be interchanged and the distribution is completely specified by only two parameters.

Finally, another way to predict the value of a variable when the value of the rest of the variables is known is by using the mode of the conditional distribution. The mode of the binomial distribution with parameters m and θ is $\lfloor (m+1)\theta \rfloor$ as long as this value is not an integer. Otherwise, the distribution has two modes: the previous one and $\lfloor (m+1)\theta \rfloor - 1$. Assuming that the bimodal case is not representative, the variable i-th would therefore be estimated using the mode of the conditional distribution:

$$\text{Mode}(N_i|\mathbf{n}_{-i}) = \lfloor (m+1)\tilde{\theta}_i \rfloor = \left\lfloor (m+1) \Bigg/ \left(1 + \exp\left[-\pi_i + \sum_{k=1}^{d} \alpha_{ik} n_k\right]\right) \right\rfloor \tag{19}$$

In this section it has been assumed that the variables take values on the same scale, namely from 0 to m. However, it is important to remark that the developments of this section are still valid if each variable N_i has a different scale and takes the integer values between 0 and m_i. In this case, the parameter m should be replaced by m_i in expressions (11), (12), (14), (15), (18) and (19).

5.3. *Estimation of the parameters of the distribution*

The distribution given by expression (14) is defined by specifying the values of parameters π_i and α_{ij} and these have to be estimated from the data. If the sample is $N_{ji}, i = 1, \ldots, d,\ j = 1, \ldots, n$, with n being the size of the sample, the parameters may be estimated by maximising the likelihood of the sample. In addition, the distribution given for expression (14) is an exponential family distribution whose sufficient statistics are $\sum_{j=1}^{n} N_{ji}$ and $\sum_{j=1}^{n} N_{ji}N_{jk}\ i, k = 1, \ldots, d,\ k \neq i$. Therefore, the parameters can be estimated simply by knowing the value of the sufficient statistics of the sample. In practice, this procedure leads to a system of non-linear equations in the parameters. These are the equations that are obtained by maximising the sample likelihood. Furthermore, knowledge of the analytical expression of the normalisation factor D in (14) is needed for this method. For the distribution in question, the analytical expression of this factor is the sum of the $(m+1)^d$ possible values of the random variables. Due to the fact that this number can be huge, the maximum likelihood approach is in practice unfeasible since the numerical evaluation of the normalisation factor would be unacceptable in terms of computing time.

To avoid this problem and due to the particular form of the distribution, the parameters can be estimated using the maximum pseudolikelihood estimators. The concept of pseudolikelihood was introduced by Besag (1974) and it has been used to estimate the parameters of distributions where the numerical evaluation of the normalisation factor is not feasible. The pseudolikelihood is simply the likelihood of the conditional distributions. In the discrete and bivariate case with $\mathbf{n} = (n_1, n_2)$, the log-pseudolikelihood is simply

$$PL = \log(\text{Prob}(N_1 = n_1|n_2)) + \log(\text{Prob}(N_2 = n_2|n_1))$$

In the multivariate case, the pseudolikelihood can be defined in different ways, although the most common is the so-called full pseudolikelihood (FPL) by Mardia *et al.* (2007). In the case of a discrete distribution it would be

$$FPL = \sum_{i=1}^{d} \log(\text{Prob}(N_i = n_i|\mathbf{n}_{-i}))$$

Therefore, the full pseudolikelihood of the sample $N_{ji}, i = 1, \ldots, d,\ j = 1, \ldots, s$ will be

$$\ell = \sum_{j=1}^{n} \sum_{i=1}^{d} \log(\text{Prob}(N_{ji} = n_{ji} | \mathbf{n}_{-ji})) \tag{20}$$

The pseudolikelihood is also called composite likelihood (Mardia *et al*. 2009). Using the distribution given by (14), and the fact that the conditional distributions are binomial distributions and that they correspond to expression (15), the full pseudolikelihood of the sample would be

$$\ell = \sum_{j=1}^{n} \sum_{i=1}^{d} \log\binom{m}{N_{ji}} + m \sum_{j=1}^{n} \sum_{i=1}^{d} \log(1 - \tilde{\theta}_{ji}) + \sum_{j=1}^{n} \sum_{i=1}^{d} N_{ji} \tilde{\pi}_{ji}$$

The notation $\tilde{\pi}_{ji}$ and $\tilde{\theta}_{ji}$ is used for the parameters of the conditional distributions in the previous expression. These are linked to the parameters of the distribution π_i and α_{ik} and to the sample values. Specifically, by recalling expression (16), the following is obtained:

$$\tilde{\pi}_{ji} = \pi_i - \sum_{k=1}^{d} \alpha_{ik} N_{jk} = \log\left(\frac{\tilde{\theta}_{ji}}{1 - \tilde{\theta}_{ji}}\right) \tag{21}$$

In addition, taking into account that the parameters of the conditional distribution are linked by (17), the following expression is finally obtained:

$$l(\boldsymbol{\alpha}, \boldsymbol{\pi}) = \sum_{j=1}^{s} \sum_{i=1}^{d} \log\binom{m}{N_{ji}} - m \sum_{j=1}^{s} \sum_{i=1}^{d} \log(1 + e^{\tilde{\pi}_{ji}}) + \sum_{j=1}^{s} \sum_{i=1}^{d} N_{ji} \tilde{\pi}_{ji}$$

For a given sample, the first term is constant and the other two terms depend through expression (21) on the values of the parameters of the distribution, π_i and α_{ik}. The function $l(\boldsymbol{\alpha}, \boldsymbol{\pi})$ is concave in these parameters, which ensures that the parameter estimates are unique.

5.4. *Fitting the distribution to the survey data*

The previous distribution has been fitted to the ratings obtained from the survey. The rating ranges from 0 to 10 and consequently $m = 10$. The global satisfaction rating has been taken as the first variable of the distribution and the remaining variables are the ratings of the SS, that is: $N_1 = Y$ and $N_{k+1} = X_k, k = 1, \ldots, d$. The number of variables of the sample is $d = 36$ and the number of parameters to be estimated is 666 (=36 + 36*35/2). It is obvious that this number is too high for the size of the sample available, which is 1508. That is why, it is necessary to fit a distribution with considerably fewer parameters so that the model is not overfitted. Therefore, it is necessary to select among all the variables those that contribute most to the likelihood of the distribution. This is a typical problem of variable selection.

We have decided to include the variables in the distribution in the sequential order of relevance with which they appear in the optimal model of the weighted mean. The problem of variable selection can, therefore, be solved in a simple way. In this model, the relevance of the variable is obviously proportional to its weight. Table 3 shows the weight of each of the variables for the optimal models based on the WAM and the WPM. It can be seen that the order of relevance of the variables is more or less the same in the two models. The only two differences in the order in pairs 3 and 4 and 5 and 15 are due to the fact that the weights of these questions are practically identical. Therefore, and in order to compare the model based on the distribution with the models based on means, we have decided to select the variables in descending order of their weights in the optimal model of the WAM. Finally, the following two models based on distribution (14) have been chosen so that the results are comparable with those of the 4- and 8-block models (models MMB4 and MMB8):

- Model MD4: Distribution fitted to the sample made up of the GS rating and the 4 most significant variables, which are those corresponding to items 16, 6, 23 and 13 (Table 3).
- Model MD8: Distribution fitted to the sample made up of the GS rating and the 8 most significant variables, those corresponding to items 16, 6, 23, 13, 2, 14, 4 and 3 (Table 3).

With the above choice, the expression that predicts the global satisfaction rating defined by the mean of the conditional distribution (18) has four and eight parameters, the same number of parameters as the four- and eight-block models. This choice of the number of variables of the sample seems to be reasonable. Once the distribution has been fitted, the GS rating may be estimated by the mean of its conditional distribution. Since the aim of the fitting is actually to predict the global satisfaction rating, the mean of the absolute errors has been used as a proxy for the goodness-of-fit achieved by the estimated distribution. The absolute error is the absolute difference between the GS rating and its conditional expectation:

$$\varepsilon_{\text{abs,mean}} = \frac{1}{n}\sum_{j=1}^{n} \left| Y_j - E(Y_j\,|\mathbf{X_j}) \right| \tag{22}$$

where $\mathrm{E}(Y_j\,|\mathbf{X_j})$ is the mean of the conditional distribution of the distribution fitted to the corresponding sample. Besides the mean of the conditional distribution, the GS rating may also be forecasted by the mode of the conditional distribution. Recalling that the expression for the mode is given by (19), the error given by this GS rating estimate is

$$\varepsilon_{\text{abs,mode}} = \frac{1}{n}\sum_{j=1}^{n} \left| Y_j - \mathrm{Mode}(Y_j\,|\mathbf{X_j}) \right| \tag{23}$$

The results of the distribution fitting by means of models MD4 and MD8 are presented and analysed in Section 7.

6. A generalised linear model for the GS rating

In the previous sections a model has been proposed for the survey based on a discrete multivariate distribution whose conditional distributions are binomial. In theory, we are only interested in the prediction of the global satisfaction rating according to the ratings given to the items of the survey. Therefore, a model that is not as ambitious as the distributional model should also be considered. Such a model should aim only at describing the variable of interest, which is the global satisfaction rating. Due to the fact that this rating takes only integer values between 0 and 10, an appropriate model for this type of variable would be a generalised linear model in which the dependent variable follows a binomial distribution whose logit varies linearly with the independent variables. This is the most common form of a binomial generalised linear model (Hardin and Hilbe 2007), in which the dependent variable is $Y \sim B(m, \theta)$ and the log odds ratio or logit depends on the independent variables $X_k,\ k = 1, \ldots, q,$ as follows:

$$\log\left(\frac{\theta}{1-\theta}\right) = \pi = \beta_0 + \sum_{k=1}^{q} \beta_k X_k = \boldsymbol{\beta} X$$

where the coefficients $\beta_k, k = 0, \ldots, q,$ must be estimated from the sample. An extension of this type of model for describing ratings on items has been proposed by D'Elia and Piccolo (2005).

Taking into account that the expression of the probability mass function is

$$f(y) = \binom{m}{y}(1-\theta)^m \exp(\pi y)$$

its logarithm will be proportional to

$$\log(f(y)) \propto m \log((1-\theta)) + \pi y = -m\ \log(1+e^{\pi}) + \pi y = -m\ \log(1+e^{\boldsymbol{\beta} X}) + \boldsymbol{\beta} X y$$

If the sample available is $Y_j, X_{ji}, i = 1, \ldots, q, j = 1, \ldots, n,$ the likelihood of the generalised linear model described above will be

$$\ell = \text{constant} - m \sum_{j=1}^{n} \log(1+e^{\pi_j}) + \sum_{j=1}^{n} Y_j \pi_j$$

where

$$\pi_j = \beta_0 + \sum_{k=1}^{q} \beta_k X_{jk} = \boldsymbol{\beta} \mathbf{X_j}$$

and $\mathbf{X_j} = (1, X_{j1}, X_{j2}, \ldots, X_{jq})'$. The global satisfaction rating can be predicted with the GLM described in this section using the mean or even the mode of the distribution. In the first case the predicted value would be

$$\hat{Y}_i = E(Y_i|\mathbf{X}_i) = \frac{m}{1+e^{-\pi_i}} = \frac{m}{1+\exp\left[-\beta_0 - \sum_{k=1}^{q} \beta_k X_{jk}\right]} \tag{24}$$

and in the second case

$$\hat{Y}_i = \text{Mode}(Y_i|\mathbf{X}_i) = \left\lfloor \frac{m+1}{1+e^{-\pi_i}} \right\rfloor = \left\lfloor (m+1) \Big/ \left(1+\exp\left[\beta_0 + \sum_{k=1}^{q} \beta_k X_{jk}\right]\right) \right\rfloor \tag{25}$$

The vector of coefficients $\boldsymbol{\beta}$ of the GLM has been estimated for samples made up of the ratings of the 4 and 8 most significant items and the GS rating, just as for the model based on the distribution (14). The two models are called GLM4 and GLM8. The generalised linear model has also been fitted to the complete sample, i.e. to the 35 items, to gain an idea of the improvements in prediction when considering all the items. This model is the GLM35 and it has 36 parameters. The results are presented and discussed in the next section.

7. Results and discussion

7.1. *Results given by the models*

This section presents the results given by the models MD4 and MD8, which are based on the multivariate discrete distribution. The results given by the generalised linear models GLM4, GLM8 and GLM35 are also presented. Table 8 shows the errors obtained with the above-mentioned models. Namely, the error given in the second column is the mean of the absolute differences between the GS rating and its conditional expectation, that is, the error given by (22). For the MD4 and MD8 models, the conditional expectation in (22) is obtained from (18), whereas for the generalised linear models (GLM4, GLM8 and GLM35) it is given by (24).

The third column of Table 8 shows the errors given by forecasting the GS rating through the mode of the conditional expectation. This error is calculated according to (23) and the mode of the conditional distribution for the MD4 and MD8 models is obtained from (19). For the generalised linear models GLM4, GLM8 and GLM35, (25) yields the mode. As it can be deduced by comparing the values of the second and third columns of Table 8, slightly smaller errors are obtained by employing the conditional mode. Additionally, the fourth column of Table 8 lists the percentage of right predictions of the GS rating with the mode of the conditional distribution. This is the percentage of observations whose GS rating coincides with the mode of the conditional distribution. The percentage of correct predictions increases slightly as more items are included in the model.

By comparing the errors given by models GLM4 and GLM8 with those given by models MD4 and MD8, one can observe that the errors are slightly larger in the latter two with the same number of variables. This happens for both types of GS rating forecasts, that is, for the conditional expectation and conditional mode. One can see that the error decreases as more items are added. We may compare the error given by the MD and GLM

Table 8. Results given by the distribution and by the generalised linear models.

Model	Mean absolute error ($\varepsilon_{\text{abs,mean}}$) with mean	Mean absolute error ($\varepsilon_{\text{abs,mode}}$) with mode	Percentage of right predictions with mode
MD4: Distribution of GS rating + 4 most relevant items	0.90728	0.89589	35.41
MD8: Distribution of GS rating + 8 most relevant items	0.88704	0.87997	37.20
GLM4: Generalised linear model with 4 most relevant items	0.89783	0.87997	37.00
GLM8: Generalised linear model with 4 most relevant items	0.86495	0.85743	37.40
GLM35: Generalised linear model with all items	0.85767	0.84947	37.20

models with that given by the best 4-block models of Table 6, which is about 0.87 and by the best 8-block models, which is 0.86722 (Table 5).

We see from Table 8 that the errors given by the MD4, MD8 and GLM4 models are slightly greater than those produced by the block models, although the difference is barely relevant. On the other hand, it is necessary to take into account that the distributional approach and the generalised linear model are more ambitious approaches than the fitting of a model focused solely on predicting the global satisfaction rating. Namely, when fitting a distribution and a generalised linear model, one obtains not only the mean of the conditional distribution but also the whole conditional distribution, which gives the probability that an individual rates the global satisfaction when he/she has rated the specific satisfaction in a given way. This may also be used for obtaining confidence intervals for any rating. In particular, the multivariate distribution fit also gives the conditional distribution of any other rating and not only that of the GS rating. This may be an interesting issue for certain applications.

The model GLM35 considers all the items of the survey for the distribution of the GS rating and requires the estimation of 36 parameters. We may compare this model with the WPM model which requires the estimation of 35 parameters, since the weights sums one. The error given by the WPM model is 0.85359 (Table 2) which is comparable to those given by the GLM35 model with the conditional mean as well as with the conditional mode. As it has been previously mentioned, the advantage of the generalised linear model over the WPMmodel stems from the fact that the former gives the complete distribution of the GS rating.

Finally, another remarkable aspect of the model based on the distribution and of the generalised linear model is the fact that, for each individual of the sample, the predicted values of the GS rating with both types of predictors, conditional mean and mode, are always located between the minimum and maximum rating values given to the items by that individual. In other words, the mean as well as the mode of the conditional distribution behave as a mean in the sense of the models introduced in Section 4. It is important to take into account that in a mean model, the value predicted by the mean is always between the minimum and the maximum of the components of the mean, whatever the values of the parameters that define the mean are. However, this property does not necessarily hold for the mean and/or mode of the conditional distribution.

7.2. *Selection of the most relevant items*

As it has been explained in Section 1, the identification of the most important aspects for the users is a key issue for the improvement of the service. In other words, the public administrator would like to know on which aspects it would be most efficient to act so that people have a better perception of the service. These aspects may be chosen from those items that have the coefficients with the highest value. The mean of the conditional distribution decreases with coefficients α_{ij} for models MD4 and MD8 based on distribution (14). The aspects that the users perceive to be more important are those whose coefficients α_{ij} have a lower value. The opposite occurs for the generalised linear model, since the conditional mean increases as the coefficient β_k increases in accordance with expression (27). Then, the most important aspects according to this model would be those with the highest coefficient values. Table 9 shows the values of the coefficients obtained after fitting models MD4, MD8, GLM4, GLM8 and GLM35. The first column shows the number of the item of the survey to which the coefficient corresponds. The *p* values of the coefficients of models GLM4, GLM8 and GLM35 that are greater than 0.01 are given in the table within brackets. Particularly, variable 4 in model GLM8 is not significant and neither are variables 3, 4 and 17 for the complete model GLM35. It is interesting to see that in the complete model, item 23 (bus illumination) is not as significant as in the model with 8 variables, since it has a lower coefficient value and a higher *p*-value. Finally, the GLM35 model allows the identification of item 17 (service time window) as significant. This item was excluded for the rest of the models. However, this item has a coefficient with a relatively low value.

It can be clearly seen that the most important aspect concerns item 16 (reliability of the service), which appears in the first position in all the models except MD4, where it appears in the second place. Another relevant item in accordance with models GLM4 and GLM8 would be 23 (bus illumination), an item that appears in the third and the second place, respectively. However, this aspect is not very relevant in MD8 and GLM35. In fact, item 6 (adequacy of the bus stop location) is relevant in nearly all the models; it is the second/third most relevant in the distribution-based models, and the corresponding coefficient has a significant value for the generalised models. Furthermore, items 13 (punctuality), 2 (connection to lines of other operators) and 14 (service frequency) can also be considered relevant since their coefficients have significantly high values in both model types. Service frequency is slightly more relevant than connection to other transit lines. In summary, items 16, 6, 13, 2 and 14 are clearly the most relevant across all the models, and in general terms, the first one is more or less twice as relevant as the others, if we draw this conclusion in accordance with the coefficient values given by models GLM8 and GLM35. Finally, items 17 (service time window) and 3 (line diversity or number of lines of the network) are found to be slightly relevant in models GLM35, GLM8 and MD8, respectively.

The relative magnitude of the parameters of the function that forecasts the global satisfaction rating allows the identification of those items that most influence the users'

Table 9. Coefficient values given by the distribution and by the generalised linear model.

Model					
Item number	MD4	GLM4	MD8	GLM8	GLM35
16	−0.09677	0.13228	−0.15677	0.09882	0.08620
6	−0.08387	0.07566	−0.09176	0.04987	0.04859
23	−0.04474	0.08491	−0.00280	0.06734	0.04516 (0.03276)
13	−0.14651	0.10073	−0.02661	0.05463	0.04422
2			−0.04487	0.05742	0.04740
14			−0.06726	0.06583	0.05838
4			0.04529	0.02537 (0.08610)	0.02187 (0.16100)
			−0.05604	0.03200 (0.01280)	0.01864 (0.16530)
17					0.02861 (0.02730)

overall satisfaction. For the Bilbao survey, the fitting of the models presented in this work has led to the identification of the following six most relevant items: line reliability, adequacy of bus stop location, punctuality, connectivity to lines of other transit companies and service frequency. Not surprisingly, the users' satisfaction with the bus service is concentrated on few factors, which have a major impact on the overall satisfaction. By comparing the weight that each model gives to the different items, one may discard some items whose weighting is not homogeneous across the models. This way of proceeding guarantees a robust selection of the most significant items in terms of user global satisfaction. Then, it seems clear that those items that have a great significance in all the models are finally the items that we are looking for. On the contrary, one should be more reluctant to select an item that is relevant only in some models. The final decision on these items should probably also take into account more subjective considerations.

The WPM and WAM models have identified the following six items as the most significant: 16, 6, 23, 13, 2 and 14. The service aspects corresponding to these items are: line reliability, bus stop location adequacy, bus illumination, bus punctuality, connections to lines of other operators and service frequency. The weights of these items, given by the WPM and WAM models, sum up to 67%, that is, two-thirds of the total. InSection 4, the block models have been introduced and models MMB4 and MMB8 with four and eight blocks, respectively, have been estimated. The main conclusion from this analysis is the fact that the four-block model has essentially the same accuracy as the eight-block model. Therefore, the only relevant blocks are 1, 2, 4 and 6 and the contribution of the remaining blocks in the MMB model is negligible, as shown in Figure 3(a). From the relevant blocks, block 4 (time satisfaction) weights about three times more than blocks 1 (connectivity), 2 (accessibility) and 6 (comfort), as indicated in Table 7. It is interesting to observe from Table 1 that three of the five items of block 4 are included in the six most relevant items of the WPM and WAM models. These are items 13, 14 and 16. In a second level of significance would be blocks 1 and 2, because only one of their three items is also identified as very significant by the WPM and WAM models. These are items 2 of block 1 and 6 of block 2. This fact is in accordance with the average weight of this block (see the last column of Table 7). Finally, the relatively low relevance of block 6, whose average weight is the smallest one (0.13862), is in concordance with the fact that only one of its 11 items (item 23) is in the group of the six most important items according to the WPM and WAM models.

The results given by the distributional models MD4 and MD8 and by the generalised linear models GLM4, GLM8 and GM35 basically confirm the results summarised in the previous paragraph. Namely, the most relevant items are 16, 6, 13, 2 and 14, whereas some care must be exercised with item 23 (bus illumination). The forecasting performance of the models as measured by the mean absolute error is in a way rather similar. A minimum error of 0.85359 is achieved by the WPM model with $\beta = 1.0122$, and a maximum of 0.9 by the MD4 model. The block models and the GLM4 and GLM8 models yield errors of about 0.87. The greater error given by the models based on the distribution should not be considered as a poor mark since these models fit the overall sample. The final choice of a particular model is a somewhat personal question and each model has its own advantages and disadvantages. The models based on means are simple and intuitive but they only give an estimate of the GS rating. On the contrary, the generalised linear model and the multivariate distribution also give the distribution of the rating in the form of a binomial distribution.

8. Conclusions

Transportation agencies and administrations are becoming more customer-oriented and are striving to increase user satisfaction with the main purpose of attracting a larger number of users. For this reason, much emphasis is being placed on satisfaction considerations. Periodic surveys on user satisfaction related to different issues of the services are conducted in order to provide quantitative assessment of the most important aspects from the users' point of view. Most surveys employ a scale of ratings so that users can value the quality of the different service items under investigation. Additionally, these item-related or specific ratings are combined into a single index or global satisfaction rating.

This work presents different methods to express the global satisfaction ratings as a function of the specific satisfaction ratings. Namely, three types of models have been presented: models based on means (WPM and block means), a model based on a multivariate discrete distribution and a generalised linear model. The parameters of these models have been estimated by fitting them to the sample of global and specific ratings collected in a survey made for the Bilbao public bus company. The forecasting performance of the global satisfaction rating given by the models is quite satisfactory in all the cases. The models based on means give as a direct result the function that forecasts the global satisfaction rating. The generalised linear model and the multivariate distribution give such a function as the conditional expectation of the global satisfaction rating.

In summary, the contribution of this work is twofold. First, several models for estimating the global satisfaction rating have been introduced. The models based on weighted means are particularly suited for constructing a global satisfaction index from specific satisfaction ratings.

In particular, as far as the authors know, the generalised linear model and the model based on the multivariate distribution have never been applied in the field of transportation for the purpose of forecasting a satisfaction index or rating. These models have the advantage of yielding not only an estimate of the GS rating but also its conditional probability. This is a fact that has not been previously addressed in other works. Additionally, the derivation of the multivariate discrete distribution, its conditional probability distributions and the pseudolikelihood have also been presented in this article. Previous developments and uses of the distribution are confined to its simplest version, which is that corresponding to binary data. Secondly, a methodology to select the most important items influencing the overall satisfaction has been exposed. This methodology helps to compare the results given by several models in order to select the most relevant items in a robust way. Namely, it has been found that the most influencing items are the following: line reliability, adequacy of the bus stop location, punctuality, connection to lines of other operators and service frequency. These findings are in accordance to those obtained by Eboli and Mazzulla (2007), Fellesson and Friman (2008), Tyrinopoulos and Antoniou (2008), Budiono (2009) and Ji and Gao (2010).

Acknowledgements

This research was carried out with the collaboration of ESL Consulting for the market exploitation agency of the transit service. The contents of this article reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the City service, the Department of Transportation or the Regional Transportation Administration.

References

Agarwal, R., 2008. Public transportation and customer satisfaction: the case of Indian Railways. *Global Business Review*, 9, 257–272.

Andreassen, T.W., 1995. (Dis)satisfaction with public services: the case of public transportation. *Journal of Services Marketing*, 9, 30–41.

Arnold, B., Castillo, E., and Sarabia, J.M., 1999. *Conditional specification of statistical models*. New York: Springer.

Becker, N.G. and Utev, S., 2002. Multivariate discrete distributions with a product-type dependence. *Journal of Multivariate Analysis*, 83, 509–524.

Beliakov, G., 2003. How to build aggregation operators from data. *International Journal of Intelligent Systems*, 18, 903–923.

Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series. B*, 36, 192–236.

Budiono, O.A., 2009. *Customer satisfaction in public bus transport*. Ms. Thesis, Karlstad University.

Bullen, P.S., 2003. *Handbook of means and their inequalities*. New York: Springer.

Chiou Y.C., Chen, Y.H., 2012. Service quality effects on air passenger intentions: a service chain perspective. *Transportmetrica*, in press, DOI:10.1080/18128602.2010.548837.

CEN, 2002. *Transportation-Logistics and Services – Public Passenger Transport-Service Quality Definition, Targeting and Measurement*. European Committee for Standardization.

Cox, D.R. and Wermuth, N., 2002. On some models for multivariate binary variables parallel in complexity with the multivariate Gaussian distribution. *Biometrika*, 89, 462–469.

D'Elia, A. and Piccolo, D., 2005. A mixture model for preferences data analysis. *Computational Statistics & Data Analysis*, 49, 917–934.

Dell'Olio, L., Ibeas, A., and Cecín, P., 2010. Modelling user perception of bus transit quality. *Transport Policy*, 17, 388–397.

Eboli, L. and Mazzulla, G., 2007. Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation*, 10 (3), 21–34.

Eboli, L. and Mazzulla, G., 2009. A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation*, 12 (3), 21–37.

Fellesson, M. and Friman, M., 2008. Perceived satisfaction with public transport service in nine European cities. *Journal of the Transportation Research Forum*, 47 (3), 93–103.

Friman, M., 1998. *Satisfaction with public transport services*. Diss. Karlstad, Högskolan i Karlstad.

Friman, M. and Gärling, T., 2001a. Frequency of negative critical incidents and satisfaction with public transport services. II. *Journal of Retailing and Consumer Services*, 8 (2), 105–114.

Friman, M., Gärling, T., 2001b. Satisfaction with public transport related to service performance. *In:* Travel Behavior Research, The Leading Edge, No. 845854. International Association for Travel Behavior Research. Oxford, UK: Pergamon, Elsevier.

Friman, M., Edvardsson, B., and Gärling, T., 2001. Frequency of negative critical incidents and satisfaction with public transport services. I. *Journal of Retailing and Consumer Services*, 8 (2), 95–104.

Hardin, J.W. and Hilbe, J.M., 2007. *Generalized linear models and extensions*. College Station: Stata Press.

Ji, J. and Gao, X., 2010. Analysis of people's satisfaction with public transportation in Beijing. *Habitat International*, 34, 464–470.

Joe, H., 1997. *Multivariate models and dependence concepts*. London: Chapman & Hall.

Karlaftis, M.G., Golias, J., and Papadimitriou, E., 2001. Transit quality as an integrated traffic management strategy: measuring perceived service. *Journal of Public Transportation*, 4 (1), 27–44.

Krizek, K.J. and El-Geneidy, A., 2007. Segmenting preferences and habits of transit users and non-users. *Journal of Public Transportation*, 10 (3), 71–94.

Mann, E. and Abraham, C., 2006. The role of affect in UK commuters' travel mode choices: an interpretative phenomenological analysis. *British Journal of Psychology*, 97, 155–176.

Mardia, K.V., Hughes, G., Taylor, C.C., 2007. *Efficiency of the pseudolikelihood for multivariate normal* and von Mises distributions. Research Report 07-02, Dept. of Statistics, University of Leeds.

Mardia, K.V., *et al*., 2009. Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, 96, 975–982.

Mizuki, K. and Akiko, T., 2005. Spatial dimensions of job accessibility by commuting time and mode in the Tokyo metropolitan area. *Journal of Theory and Application of GIS*, 13 (2), 139–148.

Nil, P. and Naciye, D., 2004. Measuring accessibility and utilization of public spaces case study: Famagusta. *Cities*, 21 (3), 225–232.

Perng, S.W., Chowa, C.C., and Liao, W.C., 2010. Analysis of shopping preference and satisfaction with airport retailing products. *Journal of Air Transport Management*, 16, 279–283.

Stradling, S., *et al*., 2007. Passenger perceptions and ideal urban bus journey experience. *Transport Policy*, 14, 283–292.

Tam, M.L., Lam, W.H.K., and Lo, H.P., 2010. Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica*, 6 (1), 3–17.

TRB, 1999. A handbook for measuring customer satisfaction and service quality. *TRCP Report 47*. Transportation Research Board.

TRB, 2004. Transit capacity and quality of service manual. TRCP Report 100, 2nd edition. Transportation Research Board.

Tyrinopoulos, Y. and Antoniou, C., 2008. Public transit user satisfaction: variability and policy implications. *Transport Policy*, 15 (4), 260–272.

Van Duijn, M.A.J., Gile, K.J., and Handcock, M.S., 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31, 52–62.

Wu, R.-.Z., *et al*., 2009. Evaluation model of satisfaction degree for urban public transit service. *Journal of Traffic and Transportation Engineering*, 9 (4), 65–70.