Quality and public transport service contracts

Valerio Gatta [1*], Edoardo Marcucci [2]

[1] *Dipartimento di Statistica, Probabilità e Statistiche Applicate (DSPSA)*
Facoltà di Scienze Statistiche, Sapienza Università di Roma
[2] *Istituto di Scienze economiche, matematiche e statistiche (ISEMS)*
Facoltà di Economia, Università degli studi di Urbino

Abstract

Service contracts are the natural method to set bilateral commitments. In transport service context, public authorities and transport operators have different goals, therefore regulation plays an important role especially failing competition. After a brief description of the most important regulatory procedures, we focus our attention on the quality framework in service contracts. In recent years the inclusion of quality requirements in contracts is becoming common practice, especially when adopting price cap regulation. This paper suggests a criterion for service quality definition, measurement and integration in contracts for the production of socially valuable transport services. Using choice-based conjoint analysis to analyse customer preferences we estimate the passengers' evaluation of different service features and calculate a robust specification of a service quality index from the customers' point of view. A case study demonstrates the procedure to follow for measuring service quality in local public transport differentiated by geographical service segments.

Keywords: Service quality; Stated preferences; Service contracts.

Introduction

Public authorities and transport operators are both involved in the provision of public transport services. There is a contrast between the social goals of the former and the private ones of the latter. Private firms maximise profits without considering social welfare. Regulation plays an important role especially failing competition. Service contracts are the natural method to set bilateral commitments. A contract between the authority and the operator constitutes the instrument to induce firms in naturally non-competitive markets to act in line with social targets. Only in a few countries in Europe, the relation between authorities and transport operators are not regulated by a service contract.

The question of regulatory procedures has generated an extensive literature. The traditional rate-of-return (ROR) regulation has been examined by many authors (Averch

* Corresponding author: Valerio Gatta (valerio.gatta@uniromal.it)

and Johnson, 1962; Baumol and Klevorick, 1970; Bailey, 1973; Das, 1980 and others) who agreed that ROR induces firms to produce inefficiently causing damages to consumers. Other regulatory procedures have been developed. Sappington and Sibley (1988) proposed the "incremental surplus subsidy scheme" which induces a subsidized firm in a natural monopoly to price at marginal cost and use the cost-minimizing input mix. Various authors recognized the importance of billing algorithms as a potentially strategic key for increasing social welfare. Boiteux (1960), Williamson (1966) and others identified the optimal time-of-use prices. Willig (1978) and Panzar (1977) formalized a regulatory procedure including multipart and self-selecting tariffs. Another famous regulatory procedure is the price cap system where price is set by the regulator and is adjusted over time (Acton and Vogelsang, 1989) leading to more specific investment in cost-effective innovation (See Train, 1991 for a wide literature review).

Public transport has long been dominated by a production-oriented approach, but it is now progressively moving towards a more customer-oriented one and in recent years the inclusion of quality requirements in contracts is becoming common practice. In this paper we focus our attention on the quality framework in service contracts and following Hensher *et al.* (2003) we suggest a criterion for service quality definition, measurement and integration in contracts for the production of socially valuable transport services. This contractual context is becoming more relevant since in recent contributions (Bergantino *et al.*, 2006) there is a clear and specific reference to service quality factors in regulatory schemes based on price cap. Bergantino *et al.* (2006) specifically refer to a price-quality cap system.

This paper is structured as follows. Section 2 illustrates the various approaches developed to tackle the problem of quality definition and measurement, stressing the advantages connected to the approach adopted in this paper. Section 3 describes how the method proposed could be used in the context of contractual definition of quality when preparing a public transport service contract. Section 4 shows a case study that demonstrates the procedure to follow for calculating a service quality index (SQI). Finally section 5 proposes some concluding remarks.

Measuring Service Quality

The issue of quality is contentious. Although it is recognised as a key management tool, it still remains a fairly subjective concept. Quality is often related to the notion of standards, but in many cases the existing standards are linked to performance determinants which are not very important for the customer. We reject the resulting assumption of different kinds of quality, such as "expected" and "perceived" quality or "targeted" and "delivered" quality and believe that there is only one sort of quality and it must be strongly user-oriented, that is, based on customer preferences.

A second problem concerns the measurement method. Difficulties arise from the specific and subjective nature of services. The distinctive characteristics of intangibility, heterogeneity, inseparability and perishability make services unique and different from goods and thus rendering service quality evaluation more complicated than manufacturing quality control. The most popular tools are basically customer satisfaction surveys in which respondents are asked to evaluate quality factors one at a time. Data are generally analysed by multivariate statistical techniques like factor analysis, principal components, regression or structural equation models. SERVQUAL, proposed by Parasuraman, Zeithaml and Berry (1988), is the method that has attracted the greatest attention. It is a multiple-item scale for rating both the expectations and the perceptions of the service performance on a seven-point Likert scale. They measure service quality by means of the disconfirmation model, calculating the degree and direction of discrepancy between consumers' perceptions and expectations about different dimensions of the service. Other famous methods are, for example, SERVPERF (Cronin and Taylor, 1992), Normed Quality (Teas, 1993) and Zone Of Tolerance (Zeithaml *et al.*, 1993).

The intent to overcome some critical factors pertaining to the above methods like conceptual basis, psychometric problems or troubles with the usage of Likert scales such as the well-documented tendency for respondents to choose central response options rather than extreme ones, the impact of the number of scale points used, the influence of the format and the verbal labelling of the points and the transformation from ordinal data to cardinal data, induced us to search for a new approach for measuring service quality. Following Hensher *et al.* (2003) we adopt an alternative approach with the same level of general appeal (Gatta, 2006).

First of all, quality is linked with the concept of utility gained by the consumers. Every service implies a certain level of utility depending on its characteristics. The higher is the level of quality delivered, the greater is the corresponding utility. Another crucial point is the assumption that individuals' preferences are captured by utility functions. The higher is the utility level of a service, the greater is the probability that a consumer chooses that service.

In order to represent service quality as determined by consumers, we suggest to employ a stated preference (SP) survey in which individuals are asked to choose, according to their preferences, among a set of options. The basic idea is that users buy a package of service characteristics (attributes) when deciding to travel on a bus. In particular, we recommend choice-based conjoint analysis[1] (CBCA), a decompositional method that estimates the structure of consumers' preferences given their choices between alternative service options (Mc Fadden, 1974; Louviere and Woodworth, 1983). Such method was originally developed in marketing research field with the objective to identify the structure of customers' preferences for available or not yet available products on the market. Respondents typically observe profile descriptions of two or more products and pick the most preferred from the set. The flexibility and the rich information that can be gathered by using CBCA, have allowed its application also in transport, environment and medicine.

CBCA asks the agent to explicitly choose among the profiles, thus mimicking actual market choice, rather than rating or ranking alternatives. This is the characteristic distinguishing CBCA from other types of conjoint analysis. CBCA provides less information compared to the other two methodologies, but it is also easier for the agents to understand and respond to the choices proposed, since it reproduces a context similar to that they are, in reality, accustomed to. In fact they are asked to compare a set of alternatives and select the one providing the highest utility. Furthermore, this method does not require any assumptions to be made about order or cardinality measurement (Louviere, 1988).

[1] The seminal paper is Mc Fadden (1974).

The main drawbacks can be related to the increasing burden of information that respondents have to process before choosing, when the number of attributes rises[2] as well as to the reliance on people's stated intentions, as it occurs whenever making use of questionnaires. However, in our case, we are not precisely interested in estimating demand curves, rather to identify the relative weights of the various attributes in determining quality levels.

In CBCA, the options in each choice set are constructed in terms of levels of different service attributes and designed by the researcher. The package of service attributes with the highest utility is chosen. Therefore, through the users' conjoint evaluations of the attributes, and thus through their choices, we are able to estimate utility functions and identify the relative importance of the relevant quality attributes. Besides, by means of this method, we are able to determine the global satisfaction (or utility) that a passenger obtains from the actual service and how this might change under alternative service level delivered, as well as the contribution of each elemental attribute to the overall service quality level (Hensher and Prioni, 2002). This method is more reliable than those in which attributes are evaluated one at a time (e.g. SERVQUAL, SERVPERF, Normed Quality, Zone Of Tolerance) because the data gathered from the latter lack the information about trade-offs between attributes.

The major theoretical aspects are now briefly recalled. According to random utility theory (RUT) proposed by Thurstone (1927), utility is modelled as a random variable in order to reflect the assumption that the decision-maker has a perfect discriminative capability, while the analyst has incomplete information (Ben Akiva and Lerman, 1985) deriving from unobserved alternative attributes, unobserved individual characteristics or measurement errors (Manski, 1977). The utility that individual q associates with alternative i is given by

$$U_{iq} = V_{iq} + \varepsilon_{iq}, \tag{0.1}$$

where V_{iq} is the deterministic part of the utility and ε_{iq} is the random term. The deterministic term is a linear in the parameters function of the attributes of the alternatives

$$V_{iq} = \overline{\beta}\overline{X}_{iq}, \tag{0.2}$$

where $\overline{X}_{iq}$ is the vector of attributes as perceived by individual q for alternative i, and $\overline{\beta}$ is the vector of related parameters[3].

Mc Fadden (1974) supposed that an individual facing a finite choice set selects the alternative that maximizes utility. He proposed a probabilistic approach where the probability that individual q chooses alternative i from choice set C (J alternatives) is

---

[2] However, according to Hensher (2004), cognitive burden doesn't come from the increase of information that respondents have to process due to the product of the number of attributes and number of alternatives associated with each choice set, on the contrary limited information may in itself be especially burdensome where it is an incomplete representation of the attribute space that matters to an individual. He found that choice complexity is linked with the relevancy issue.

[3] The bar in $\overline{\beta}$ and $\overline{X}$ represents a vector, although such a bar usually indicates a mean value.

$$P_q(i \mid C) = P[(\varepsilon_{jq} - \varepsilon_{iq}) < (V_{iq} - V_{jq})], \quad \forall j \neq i$$
$$= \int_{\varepsilon} I(\varepsilon_{jq} - \varepsilon_{iq} < V_{iq} - V_{jq}; \forall j \neq i) f(\varepsilon_q) d\varepsilon_q \quad (0.3)$$
,

where $f(\varepsilon_q)$ is the joint density of the random vector $\varepsilon_q = (\varepsilon_{1q}, \ldots, \varepsilon_{Jq})$, $I(\bullet)$ is the indicator function equalling 1 when the expression in parenthesis is true and 0 otherwise. That probability is a multidimensional integral over the density of the unobserved portion of utility. Equation (0.3) is referred to as a random utility model (RUM) explanation of observed choices. Different assumptions about the distribution of the random term imply different discrete choice models that can be used to analyze the gathered choice data with the purpose of estimating the β-parameters and calculating a SQI.

Service quality, service contracts and incentives

A key element involves the inclusion of quality in contracting procedures. We assume that the service contract between the authority awarding it and the operator producing the service should explicitly foresee a minimum level of service quality measured in terms of a given level of SQI. The intention is to achieve the best possible service from the user's point of view according to operator's capabilities to fulfil the requirements specified by the authority.

The question of how to establish a minimum SQI level is not trivial. The task is complicated by the significant asymmetry in the available information. Usually, regulators dispose of less information about the firms' cost and demand function than do the firms themselves and hence do not know whether the firm is able to provide the targeted quality level or the exact cost it has to bear. However, regulators know that firms act in order to maximize their own profit (Train, 1991). When contracting there is an important distinction between *ex-ante* quality, as stipulated in the contract, and *ex-post* quality, after the contract is let. After fixing a SQI target, one has to start a monitoring system so that the quality of the service supplied can be kept under control. In order to get good results it is essential that the SQI monitoring system is assigned to an independent party.

Strong incentives result from the appropriate definition and measurement of a SQI. In fact, a better service quality produces higher user satisfaction and a more attractive service that implies new customers and consequently higher revenues for the operators who may invest them to improve service quality (QUATTRO Research Consortium, 1998). SQI targets should be modified over time according to individual preferences variations.

Empirical measurement of a SQI

Case study description

In this paragraph we illustrate a case study conducted in five geographical areas of the *Marche* a region in central Italy. We describe a procedure for measuring and integrating service quality in local public transport contracts. The project was carried out in collaboration with the local transport operator. The interview we prepared was composed of two sections: in the first one respondents were asked to provide information about their current trip, referred to as revealed preference (RP) data, and, in addition, about their socioeconomic characteristics; in the second one the interviewee had to make repeated choices between three alternatives, one representing her current trip (status quo) and two hypothetical trips (different bundles of trip attribute level), referred to as SP data. The literature identifies a set of attributes (see, for example, Hensher *et al.*, 2003; Friman *et al.*, 2001; Tranportation Research Board, 1999) and, in particular, we initially considered a group of 18 and we asked people to assign them a degree of importance[4]. Based on focus groups with customers and local operators we ended up with the following five attributes as the most appropriate dimensions to characterize service quality from a user's perspective[5]: bus fare (Cost); amount of delay at bus stop (Delay); bus travel time (Trip Length); bus frequency - number of buses per hour (Frequency); amount of time between service inception and service closure (Availability).

The attribute levels were selected as percentage changes from the status quo (figure 1). Even though this choice may introduce an endogeneity bias (De Palma and Picard, 2005) and implies complexity and cumulative cognitive burden for respondents, it anchors attributes to current experience. Respondents choose within a more realistic choice set and are not forced to choose one of the hypothetical alternatives they would never choose had they the opportunity to do so. Anchoring the SP exercise to the RP choice both avoids poor quality and inappropriate responses and provides an escape to the "no-reply" problem (Stopher, 1998; Hensher *et al.*, 2003; Bradley, 1993; Carson *et al.*, 1994; Louviere *et al.*, 2000)[6]. Through a formal experimental design, that is a full profile random design[7], the attribute levels were combined into bus options and we constructed 8 choice sets per interview. One of these choice sets had a control function, in fact it was formed by three fixed-design alternatives: the best possible one, the worst one and the current one. To allow for a rich variation in the combination of attributes

[4] The 18 attributes were: bus travel time, bus fare, bus delay, bus frequency, bus availability, information at bus stop, general cleanliness on board, seats at bus stop, cover bus stop, driver kindness, capability to cover the territory, availability of tickets, safety on board, seat availability on bus, opportunity to purchase different kinds of tickets, time to reach bus stop by car, access to bus facilities, time walking to bus stop.

[5] We considered only the first five attributes in terms of degree of importance in order to avoid cognitive burden and fatigue for the interviewee. According to Aaker and Day (1990) in SP experiments four or five attributes are usually selected.

[6] In order to avoid forcing respondents, alternatively, one should include a "no choice" or a "don't know" option in the choice set. However, it may lead to a serious loss of information.

[7] We use the "Shortcut" method of design generation included in Sawtooth Software's CBC product, in which profiles for each respondent are constructed using the least often previously used attribute levels for that respondent, subject to minimal overlap. Each one-way level frequency within attributes is balanced (Sawtooth Software, 1999).

levels we prepared 25 different versions of the survey form[8]. Overall for the five geographical service segments, we administered 264 interviews either on board or at the bus stops associated with the main routes. We used a random sampling strategy to select the sample[9]. The minimum acceptable sample size *n* is determined by the desired level of accuracy of the estimated probabilities about the proportion of decision makers that choose an alternative. In our case, we have $k = 3$ alternatives, therefore it requires for a simultaneous confidence statement of specified precision about the parameters of a multinomial population. According to Tortora (1978), the minimum sample size is given by the following formula:

$$n = \frac{Bp_i(1-p_i)}{\vartheta^2}, \quad (0.4)$$

where B is the upper $(\alpha / k)x100th$ percentile of the χ^2 distribution with 1 degree of freedom for a specified confidence level $\beta = 1-\alpha$; p_i is the true proportion closest to 0,5; ϑ is the percentage level of allowable deviation between the estimated and the true proportion (that is the precision level). If there is no prior knowledge about the values of p_i's, the most precautionary situation in terms of sample size can be made assuming some $p_i = 0,5$.

For $\beta = 95\%$ and $\vartheta = 6\%$, the minimum sample size suggested would be 398. However, each respondents contributes a number of observations, that is, each individual does $r = 8$ choice scenarios. In practice, we may obtain the required n choices from n/r respondents, but this holds only if decision makers treat each choice occasion as being an independent decision task. In the latter case, we would end up with about 50 individuals[10].

For econometric analysis, we ignored all the interviews in which the agents failed to answer correctly the control choice exercise (14 over 264, net used 250).

Delay	Cost	Trip length	Frequency	Availability
+100%	+50%	+50%	+50%	+20%
+50%	+25%	Same as now	Same as now	+10%
Same as now	Same as now	-50%	-50%	Same as now
-50%	-25%			-10%
-100%	-50%			-20%

Figure 1: attributes and their levels selected for the choice experiments.

[8] In this study we employed paper-and-pencil interviews. We are presently constructing computer-based interviews in which attribute levels are still selected as percentage changes from the status quo but they are expressed in the choice profiles as cardinal numbers. Basically, we are producing a software for customized interviews.

[9] We do not have information about the demand levels for each area because the service company itself does not know how many people are using the bus.

[10] The sample size for each area is equal or greater than 45, the exact amount requested in case of equidistribution of the proportions (when $p_i = 0,\overline{3} \quad \forall i = 1,2,3$).

Revealed Preference data

Table 1 provides information about the socioeconomic variables we included in the survey for each geographical segment.

Table 1: socioeconomic data by area.

SOCIOECONOMIC VARIABLES		***AREA***					*All*
		1	2	3	4	5	
GENDER	Female	60,0%	53,3%	55,3%	45,0%	55,3%	53,8%
	Male	40,0%	46,7%	44,7%	55,0%	44,7%	46,2%
AGE	<30	95,4%	55,6%	53,2%	71,7%	61,7%	69,7%
	30-50	3,1%	31,1%	29,8%	21,7%	27,7%	21,2%
	>50	1,5%	13,3%	17,0%	6,7%	10,6%	9,1%
NATIONALITY	Foreign	6,2%	-	12,8%	6,7%	14,9%	8,0%
	Italian	93,8%	100,0%	87,2%	93,3%	85,1%	92,0%
INCOME (€ annual)	<2500	72,2%	45,8%	37,0%	41,7%	42,3%	45,8%
	2500-5000	22,2%	12,5%	22,2%	27,8%	19,2%	21,4%
	5000-10000	-	12,5%	14,8%	8,3%	23,1%	12,2%
	10000-15000	5,6%	12,5%	11,1%	5,6%	3,8%	7,6%
	15000-20000	-	12,5%	3,7%	8,3%	7,7%	6,9%
	>20000	-	4,2%	11,1%	8,3%	3,8%	6,1%
MAIN OCCUPATION	Employed full time	-	33,3%	19,1%	16,7%	21,3%	16,7%
	Self-employed worker	3,1%	4,4%	6,4%	3,3%	4,3%	4,2%
	Student	75,4%	33,3%	36,2%	60,0%	38,3%	51,1%
	Student-worker	20,0%	8,9%	8,5%	13,3%	8,5%	12,5%
	Retired or pensioner	-	6,7%	12,8%	3,3%	4,3%	4,9%
	Unemployed	-	4,4%	4,3%	-	10,6%	3,4%
	Housewife	1,5%	8,9%	4,3%	3,3%	6,4%	4,5%
	Other	-	-	8,5%	-	6,4%	2,7%
FREQUENCY OF BUS USAGE	Almost every day	46,2%	37,8%	42,6%	8,3%	42,6%	34,8%
	1-3 times a week	29,2%	22,2%	27,7%	13,3%	25,5%	23,5%
	Once every two weeks	9,2%	13,3%	8,5%	58,3%	19,1%	22,7%
	Rarely	15,4%	26,7%	21,3%	20,0%	12,8%	18,9%
TRIP PURPOSE	Study	87,7%	28,9%	34,0%	38,3%	27,7%	46,2%
	Work	4,6%	26,7%	27,7%	15,0%	38,3%	20,8%
	Leisure	4,6%	28,9%	12,8%	33,3%	23,4%	20,1%
	Other	3,1%	15,6%	25,5%	13,3%	10,6%	12,9%
AVAILABILITY OF ALTERNATIVE MEANS OF TRANSPORT	None	63,1%	26,7%	31,9%	60,0%	44,7%	47,3%
	Car	18,5%	33,3%	44,7%	31,7%	25,5%	29,9%
	Motorcycle	1,5%	6,7%	2,1%	-	8,5%	3,4%
	Scooter	10,8%	13,3%	10,6%	3,3%	17,0%	10,6%
	Other	-	-	4,3%	-	-	0,8%
	More than one	6,2%	20,0%	6,4%	5,0%	4,3%	8,0%
TOTAL RESPONSES (N)		65	45	47	60	47	264

The sample is composed of 142 females and 122 males. In all but area 4 males are less than females. Although the average age is 30, the 69,7% of the sample is under 30 (95,4% in area 1). About half of the sample is composed by students and the annual income is under 2.500€ (about 75% in area 1). Almost all of the sample is made up of Italians. As regards the frequency of bus usage, the most representative class (34,8%) is the first one, that uses the bus "almost every day", except for area 4 where "once every two weeks" has a share of 58,3%. The data just described suggest that the sampled people have knowledge of the service delivered and the conclusions drawn from their responses should be considered reliable. Study is the main trip purpose (46,2%) while in half of the cases there is no availability of alternative means of transport so the bus is a forced choice.

Table 2 reports the average attribute levels associated with the current trip and the relative average cutoffs by segment.

Table 2: RP data on current trip: average attribute levels and relative average cutoffs.

	AREA					*All*
	1	2	3	4	5	
ATTRIBUTES IN CURRENT TRIP						
COST (€)	0,65	0,73	0,71	2,05	0,99	1,05
St. dev.	0,23	0,12	0,15	0,49	0,33	0,63
[min; max]	[0,13; 0,80]	[0,38; 0,80]	[0,19; 0,83]	[0,45; 3,20]	[0,40; 2,75]	[0,13; 3,20]
DELAY (minute)	1,80	2,79	1,74	1,84	1,49	1,92
St. dev.	1,67	2,22	1,78	3,26	2,20	2,34
[min; max]	[0; 6]	[0;10]	[0; 5]	[0; 15]	[0; 10]	[0; 15]
TRIP LENGTH (minute)	9,14	10,79	14,37	48,71	21,21	21,33
St. dev.	4,42	4,56	6,56	13,16	3,79	16,88
[min; max]	[2; 20]	[3; 20]	[5; 30]	[8; 60]	[10; 30]	[2; 60]
FREQUENCY (n°buses/h)	2,9	2,0	1,7	1,1	1,9	1,9
St. dev.	0,99	0,50	0,49	0,41	0,25	0,90
[min; max]	[0,67; 4]	[1; 4]	[1; 2]	[1; 4]	[1; 2]	[0,67; 4]
AVAILABILITY (minute)	796	816	772	853	857	818
St. dev.	59,30	70,14	40,38	95,57	49,83	89,44
[min; max]	[720; 1055]	[720 1055]	[615; 840]	[720 1080]	[690; 960]	[615; 1080]
ATTRIBUTES CUTOFF						
COST_Cutoff	0,98	1,10	1,11	2,86	1,41	1,52
	(+50%)	(+51%)	(+57%)	(+40%)	(+42%)	(+45%)
DELAY _Cutoff	10,87	9,35	11,13	12,86	10,71	11,07
	(+504%)	(+235%)	(+539%)	(+597%)	(+617%)	(+477%)
TRIP LENGTH _Cutoff	16,80	17,53	21,92	61,96	31,40	30,50
	(+84%)	(+63%)	(+53%)	(+27%)	(+48%)	(+43%)
FREQUENCY_Cutoff	1,5	1,1	0,9	0,8	1,2	1,1
	(-47%)	(-42%)	(-47%)	(-21%)	(-37%)	(-41%)
AVAILABILITY_Cutoff	628	677	551	685	664	641
	(-21%)	(-17%)	(-29%)	(-20%)	(-23%)	(-22%)

The fractions of the variables expressed by minutes are rounded as general numeric type variables.
In brackets the percentage variations with respect to the actual levels.

Before commenting the results in Table 2, we briefly clarify the meaning of cutoffs. A cutoff is a self-imposed constraint by the decision maker. According to Swait (2001) we have extended the traditional compensatory utility maximization framework incorporating attribute cutoffs into the decision problem formulation. We asked people to state their upper bounds for variables which have a negative impact on utility and lower bounds for those with positive impact on utility. Then allowing respondents to violate the self-imposed constraints at a potential cost leads to the formulation of a penalized utility function and implies non-linearities in the preference function.

As we can see in Table 2, the average bus fare is approximately 1€ while on average the maximum level of bus fare that respondents are willing to pay is about 1,50€. The high fare experienced in area 4 is due to the fact that user in that area don't often buy monthly tickets[11]. The average delay is about 2 minutes while the cutoff is about 11 minutes. If we look at the percentage increase of delay and compare it with the other percentage variations, we find a surprising result: while agents are willing to accept an increase of 477% of current delay, on the other hand, the increase (or decrease) in other attributes is never greater than 50%. This fact reflects an underestimation of the actual delay because of the large number of interviews that were administered at the terminus. Bus travel time is around 21 minutes while the cutoff is about 30 minutes. The number of buses per hour is about 2 while the cutoff is not much greater than 1. Finally, the time interval between the first and the last bus is 818 minutes and users are willing to accept a decrease till 614 minutes.

Econometric results

Now we turn our attention to the issue of parameter estimation. We may obtain information about the relative importance of the attributes using discrete choice models. Multinomial logit (ML), is derived from the assumption that the error terms of the utility functions are independent and identically Gumbel distributed. The choice probabilities of ML are expressed as follows

$$P_q(i \mid C) = \frac{e^{\lambda \bar{\beta} \bar{X}_{iq}}}{\sum_{j=1}^{J} e^{\lambda \bar{\beta} \bar{X}_{jq}}} \quad , \tag{0.5}$$

where λ is the scale parameter inversely related to the variance of the error term[12].

We have data from 5 different areas so we cannot simply estimate 5 different MLs. In order to make meaningful comparisons between attributes parameters associated with different geographical segments we follow Hensher *et al.* (2003); we pool the data and use the nested logit (NL) structure as a trick to reveal differences in scale. Normalising the scale parameter for one segment (area 1) allows variation for the other four segments.

[11] In fact, when users bought a monthly ticket, the bus fare was calculated dividing the price for that type of ticket by the average number of trips made in a month.

[12] In most cases, the arbitrary decision about λ does not matter and can be safely ignored, but since the scale factor affects the values of the estimated taste parameters (the larger the scale, the bigger the coefficients), one should never directly compare the coefficients from different choice models (Adamowicz et al., 1998).

The econometric results with the scaled coefficients are summarised in Table 3. Besides the five attributes we included the cutoff violations (e.g. COST_VC), that is the positive amount by which the lower and upper cutoffs for each attribute are violated, as explanatory variables. Therefore the parameters associated to these quantities should be negative, representing marginal disutilities. In the final model we included all the variables which have significant parameter[13].

Table 3: final model with the scaled coefficients.

VARIABLE	*AREA*				
	1	2	3	4	5
	$\tilde{\beta}$ -COEFFICIENT				
COST	-2,8287	-0,8358	-1,6680	-0,7472	-1,4181
	(0,3048)	(0,2642)	(0,2299)	(0,1201)	(0,2055)
DELAY	-0,2414	-0,1094	-0,0492	-0,1212	-0,0886
	(0,0439)	(0,0245)	(0,0327)	(0,0300)	(0,0345)
TRIP LENGTH	-0,0332	-0,0261	-0,0303	-0,0168	-0,0292
	(0,0167)	(0,0133)	(0,0119)	(0,0038)	(0,0072)
FREQUENCY	0,2251	0,5747	0,3757	0,6132	0,1588
	(0,0564)	(0,0809)	(0,0905)	(0,2133)	(0,0872)
AVAILABILITY	0,0034	0,0040	0,0027	0,0033	0,0029
	(0,0007)	(0,0006)	(0,0006)	(0,0007)	(0,0006)
COST_VC	-	-6,0148	-	-0,8768	-1,6492
		(1,3018)		(0,3818)	(0,7550)
TRIP LENGTH _VC	-	-	-0,1153	-	-
			(0,0572)		
FREQUENCY _VC	-	-	-	-2,1301	-0,7756
				(0,5669)	(0,2956)
SCALE PARAMETER	1,0000	0,6378	0,7653	0,9529	0,6802
	(fixed)	(0,0584)	(0,0760)	(0,0822)	(0,0638)

Log likelihood function	-1598.892
Restricted log likelihood	-5394.436
Chi squared	7591.089
Degrees of freedom	35
Prob[ChiSqd > value] =	.0000000
R-sqrd = 1-LogL/LogL* =	.70360

In brackets the standard errors for the parameter estimates.

[13] We investigated a number of interactions between attributes and socio-economic variables, but they did not add significantly to the overall goodness-of-fit. It was possible to estimate interactions coefficients only for pooled data since interactions within nests could not be estimated due to limited availability of data. Even if this problem could potentially be overcome by increasing the number of observations, the results obtained with pooled data indicate limited explanatory power achievable through this method. However given a sufficiently high number of observations one could avoid using a nested logit specification since the scale parameters may not prove statistically different from 1 when individual characteristics are introduced in the specification. Moreover, we implicitly assume that the residuals corresponding to different questions for the same individual are independent, which may not hold. A possible extension could be the use of a random coefficient model, with random terms specific to the individual and identical across questions. These specific issues remain to be tested in future research endeavours.

The overall explanatory power of this non-linear model is very high, a pseudo-R^2 of 0,7 is equivalent to approximately 0,9 for a linear model (Domencich and Mc Fadden, 1975) and this is in line with a similar study conducted in a different environment (Hensher *et al.*, 2003). The interpretation is very interesting since the weights of the attributes vary between areas. Area 1 is the most price sensitive while area 4 is the least price sensitive. Area 2 is characterised by a very high cutoff of bus fare. Area 1 shows the highest coefficients for delay and trip length. People in area 4 are the most sensitive to bus frequency and the associated cutoff has a significant impact. People in area 2 are the most responsive to service availability. Finally, the scale parameters are all statistically significant and different from one (except for area 4) indicating that the data cannot be pooled.

Calculating a Service Quality Index

In drafting contracts, it is crucial to take into account the local conditions and the distinctive characteristics of the public transport system considered. Hence, setting the minimum SQI level should be context-specific.

In order to calculate a SQI for each area we first calculate the SQI measure for each user through the formula

$$SQI_q = \sum_{k=1}^{K} \tilde{\beta}_k X_{kq} \; . \qquad (0.6)$$

The SQI for user q is obtained by multiplying the RP attribute levels, as perceived by user q, by the appropriate scaled $\tilde{\beta}$-parameter in Table 3 and summing across the k attributes (in this case five). Then for each geographical segment s the overall SQI is measured by taking the individual SQI average for the sampled users in each area:

$$SQI_s = \frac{\sum_{q=1}^{n_s} SQI_q}{n_s} \; . \qquad (0.7)$$

Table 4 shows the overall SQIs and the contributions of each attribute by area.

The various SQIs assume values between 0,69 (area 5) and 3,18 (area 2) and the mean is 1,24. As expected, bus fare, delay and travel time are sources of negative utility while service frequency and service availability offer positive contributions. In particular service availability is the most important attribute in explaining user satisfaction in each segment. Increasing the amount of time between service inception and service closure has the greatest effect in improving the SQI.

Moreover, SQI measures for different scenarios of public transport service can be calculated, since different mixes of attribute levels produce different SQI indexes.

Table 4: SQI and attributes contributions per area.

ATTRIBUTES	*AREA*					
	1	2	3	4	5	All
SQI_COST*	-1,84	-0,61	-1,19	-1,53	-1,40	-1,36
SQI_DELAY*	-0,43	-0,31	-0,09	-0,22	-0,13	-0,25
SQI_TRIP LENGTH*	-0,30	-0,28	-0,44	-0,82	-0,62	-0,49
SQI_FREQUENCY*	0,66	1,12	0,62	0,66	0,31	0,67
SQI_AVAILABILITY*	2,70	3,25	2,11	2,80	2,53	2,68
SQI	0,78	3,18	1,02	0,88	0,69	1,24

*Contributions account for cutoffs' influence

Concluding remarks

This paper has analysed service quality measurement and its integration in service contracts so to provide correct regulatory incentives via the introduction of the proposed quality specification and measurement. A case study illustrates the mechanism. In order to obtain reliable results, in the future a carefully structured sampling plan is needed.

Using SP methods and CBCA we estimate passengers' evaluation of different bus service features which users perceive to be the sources of utility and via discrete choice models we calculate a SQI.

Future research will pursue three different goals, one more strictly related to the methodological issues and the remaining two both related to the practical impact that SQI measurement might have.

As it is for further methodological investigation, as a referee mentioned, one should also test other possible explanations of the results obtained. In particular, one could consider the effects of random sampling variation (were this strong enough, it could explain the variability of the coefficients that are now imputed to differences across regions), individual characteristics' heterogeneity and observed attributes' endogeneity.

Whereas for the practical impact that SQI measurement might have, we would like to explore the role SQI might have within a service quality contract based on a price-quality cap as recent contributions underline (Bergantino *et al.*, 2006; Billette de Villemeur *et al.*, 2003; Cremer *et al.*, 1997) as well as to study the potential applications a SQI might have in defining a marketing strategy aimed at increasing profits. In fact, from the supplier's point of view, there is a need to establish the optimum trade-off between the service quality and its supply cost. The proposed method may also provide a useful performance assessment tool, in fact the operators may well understand where to focus their investment in order to reach a high level of service quality and increase their competitive advantage without wasting financial resources in relatively less important attributes amelioration.

The focus on quality should be a shared goal by the authorities and operators involved in the provision of transport services and the adoption of the suggested framework could prove a first step in this direction.

References

Aaker, D. and Day, G. S. (1990) *Marketing research*, John Wiley, New York.

Acton, J. and Vogelsang, I. (1989) "Introduction", Symposium on Price Cap Regulation, *Rand Journal of Economics*, 20(3), pp. 369-372.

Adamowicz, W., Louviere, J. and Swait, J.(1998) "Introduction to Attribute-Based Stated Choice Methods", Final report to Resource Valuation Branch, Damage Assessment Center, NOAA, U.S. Department of Commerce, January.

Averch, H. and Johnson, L. (1962) "Behavior of the Firm Under Regulatory Constraint", *American Economic Review*, 52(5), pp. 1053-1069.

Bailey, E. (1973) *Economic Theory of Regulatory Constraint*, Lexington Books, Lexington, MA.

Baumol, W. and Klevorick, A. (1970) "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion", *Bell Journal of Economics and Management Science*, 1(1), pp. 162-190.

Ben-Akiva, M. and Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge.

Bergantino, A. S., Billette de Villemeur, E. and Vinella, A. (2006) "A Model of Partial Regulation in the Maritime Ferry Industry", VIII SIET conference, 29th of June, Trieste, Italy.

Billette de Villemeur, E., Cremer H., Roy, B. and Toledano, J. (2003) "Optimal Pricing and Price-Cap Regulation in the Postal Sector", *Journal of Regulatory Economics*, 24(1), pp. 49-62.

Boiteux, M. (1960) "Peak-Load Pricing", *Journal of Business*, 33(2), pp. 157-179.

Bradley, M. A., (1993) "Behavioral models of airport choice and air route choice", Sixth International Conference On Travel Behaviour Research, Santiago, Chile, June 1993.

Carson R., Louviere J., Anderson D., Arabie P., Bunch D., Hensher D., Johnson R., Kuhfeld W., Steinberg D., Swait J., Timmermans H., Wiley B., (1994), "Experimental Analysis of Choice", Marketing Letters, 5(4), pp. 351-368.

Cronin, J. J. and Taylor, S. A. (1992) "Measuring service quality: a reexamination and extension", *Journal of Marketing*, 56(3), pp. 55-68.

Cremer, H., De Rycke, M. and Grimaud, A. (1997) "Service Quality, Competition, and Regulatory Policies in the Postal Sector", *Journal of Regulatory Economics*, 11(1), pp. 5-19.

Das, S. (1980) "On the Effect of Rate of Return Regulation under Uncertainty", *American Economic Review*, 70(3), pp. 456-460.

De Palma, A. and Picard, N. (2005) "Route choice decision under travel time uncertainty", *Transportation Research A: Policy and Practice*, 39(4), pp. 295-324.

Domencich, T. and Mc Fadden, D. (1975) *Urban travel demand, a behavioural analysis*, North-Holland Publishing, Oxford.

Friman, M., Edvardsson, B. and Garling, T. (2001) "Frequency of negative critical incidents and satisfaction with public transport services, I", *Journal of Retailing and Consumer Service*, 8(2), pp. 95-104.

Gatta, V. (2006) "Evaluating Service Quality: a Stated Preferences approach", unpublished Ph.D. Thesis, Department of CNAPS, Faculty of Statistics, *La Sapienza* University of Rome.

Hensher, D. A. and Prioni, P. (2002) "A service quality index for area-wide contract performance assessment regime", *Journal of Transport Economics and Policy*, 36(1), pp. 93-113.

Hensher, D. A., Stopher, P. and Bullock, P. (2003) "Service quality - developing a service quality index in the provision of commercial bus contracts", *Transportation Research A*, 37(6), pp. 499-517.

Hensher, D. A. (2004) "How do Respondents Handle Stated Choice Experiments? – Information processing strategies under varying information load", Working Paper ITS-WP-04-14.

Louviere, J. and Woodworth, G. G. (1983) "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data", *Journal of Marketing Research*, 20(4), pp. 350-367.

Louviere, J. (1988) "Conjoint analysis modelling of stated preferences. A review of theory, methods, recent developments and external validity", *Journal of Transport Economics ad Policy, 22*(1), pp. 93-119.

Louviere J., Hensher D., Swait J., (2000), "Stated Choice Methods: Analysis and Applications", Cambridge University Press, UK.

Manski, C. (1977) "The structure of random utility models", *Theory and Decision*, 8, pp. 229–254.

Mc Fadden, D. (1974) "Conditional logit analysis of qualitative choice behaviour", in Zarembka P., ed., *Frontiers in Econometrics*, Academic Press, New York, pp. 105-142.

Panzar, J. (1977) “The Pareto Dominance of Usage Insensitive Pricing”, in Dorick H., ed., Proceedings of the Sixth Annual Telecommunications Policy Research Conference, Lexington Books, Lexington, MA.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1988) “SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality”, *Journal of Retailing*, 64(1), pp. 12-37.

QUATTRO Research Consortium, (1998) “Quality approach in tendering urban public transport operations”, Transport Research - Fourth Framework Programme Urban Transport – Final Report, Office for Official Publications of the European Communities, Luxembourg.

Ramsey, F. (1927) “A Contribution to the Theory of Taxation”, *Economic Journal*, 37(1), pp. 47-61.

Sappington, D. and Sibley, D. (1988) “Regulating without Cost Information: The Incremental Surplus Subsidy Scheme”, *International Economic Review*, 29(2), pp. 297-306.

Sawtooth Software Inc., (1999) “CBC User Manual”, Sequim: Sawtooth Software.

Swait, J. (2001) “A Non-compensatory Choice Model Incorporating Attribute Cutoffs”, *Transportation Research B*, 35(10), pp. 903-928.

Stopher, P. R. (1998) “A review of separate and joint strategies for the use of data on revealed and stated choices”, *Transportation*, 25(2), pp. 187-205.

Teas, R. K. (1993) “Expectations, performance evaluation, and consumers’ perceptions of quality”, *Journal of Marketing*, 57(4), pp. 18-34.

Thurstone, L. (1927) “A law of comparative judgment”, *Psychological Review*, 34(4), pp. 273-286.

Tortora, R. D. (1978) “A Note on Sample Size Estimation for Multinomial Populations”, *The American Statistician*, 32(3), pp. 100-102.

Train, K. E. (1991) *Optimal Regulation: The Economic Theory of Natural Monopoly*, The MIT Press, Cambridge.

Transportation Research Board (1999) “A handbook for measuring customer satisfaction and service quality”, Transit Cooperative Research Program, Report 47, National Academy Press, Washington, D.C.

Williamson, O. (1966) “Peak-Load Pricing and Optimal Capacity under Indivisibility Constraints”, *American Economic Review*, 56(4), pp. 810-827.

Willig, R. (1978) “Pareto Superior Non-Linear Outlay Schedules”, *Bell Journal of Economics*, 9, pp. 56-69.

Zeithaml, V. A., Berry, L. L. and Parasuraman, A. (1993) “The Nature and Determinants of Customer Expectations of Service”, *Journal of the Academy of Marketing Science*, 21(1), pp. 1-12.

Acknowledgements

Financial help form CIPE 17/2003 contract act N. 1804, 23/12/2003 titled “Interdisciplinary analysis for the definition of urban mobility policies in the Marche Region aimed at containing individual transport and increasing local public transport” is gratefully acknowledged. We would also like to thank the participants to the SKIM Software & Sawtooth Software conference on “Design & Innovations” held in Munich (Germany), the 3rd International Kuhmo Conference and Nectar Cluster 2 meeting on “Pricing, Financing, and Investment in Transport” held in Tuusula (Finland), the SIEP seminar on “Public services: the new trends in regulation, production and financing” held in Pavia (Italy), as well as the valuable suggestions and comments of Nathalie Picard and the anonymous referees for their insightful and helpful comments.