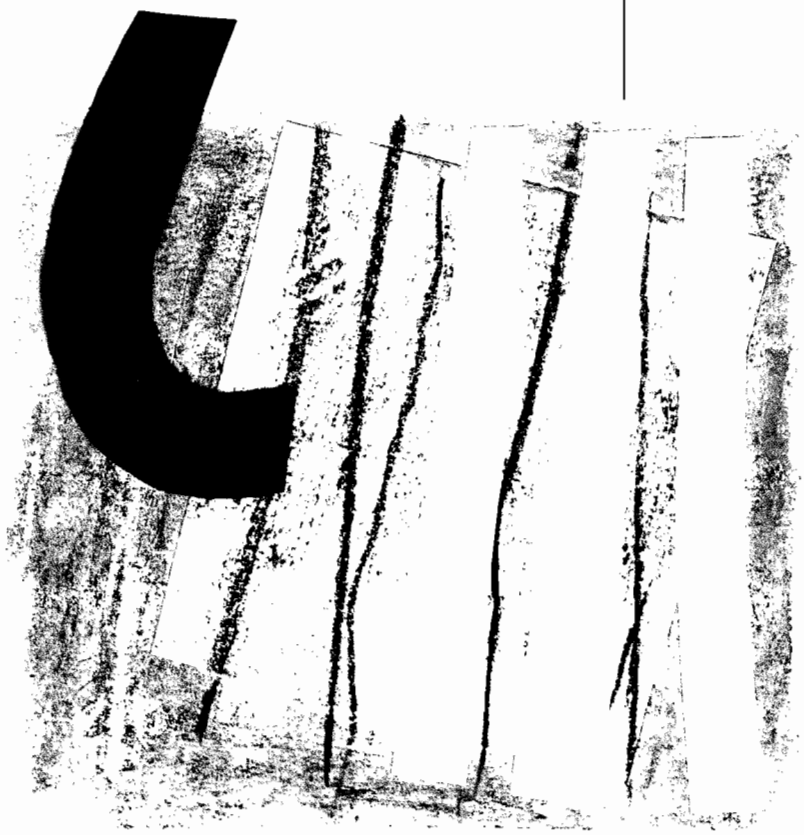
MEASURING SERVICE QUALITY BY LINEAR INDICATORS

Daniel Peña

96-46



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 96-46
Statistics and Econometrics Series 17
September 1996

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

MEASURING SERVICE QUALITY BY LINEAR INDICATORS

Daniel Peña *

Abstract

This paper presents a methodology for building linear indicators of the quality of a service. A regression model is presented in which customers evaluate the overall service quality and a set of dimensions or attributes that determine this service quality. The model assumes that overall service quality is determined by a linear combination of attribute evaluations with some unknown weights. As different customers may have different weights, the estimation of the parameters of the distribution of weights in the population is carried out by generalized constrained least squares. The model is applied to the measurement of the quality of the education provided by an university and the quality of the Spanish railroad system.

Key Words

Constrained Regression. Multiattribute model. Weight Assessment.

* Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

MEASURING SERVICE QUALITY BY LINEAR INDICATORS

Daniel Peña
Universidad Carlos III de Madrid

December 1996

Summary

This paper presents a methodology for building linear indicators of the quality of a service. A regression model is presented in which customers evaluate the overall service quality and a set of dimensions or attributes that determine this service quality. The model assumes that overall service quality is determined by a linear combination of attribute evaluations with some unknown weights. As different customers may have different weights, the estimation of the parameters of the distribution of weights in the population is carried out by generalized constrained least squares. The model is applied to the measurement of the quality of the education provided by an university and the quality of the Spanish railroad system.

Key Words. Constrained Regression. Multiattribute model. Weight Assessment.

INTRODUCTION

Any process of measurement requires the comparison with some standard. When measuring product quality these standards can be objective (as rate of failure, dimensions, strength, time and so on) but in services quality these standards are generally subjective and depends on each customer 's experiences with this and similar services. Also these standards are changing over time. It is generally accepted that the quality of a service is a function of the comparison between the perceived quality and the customer's expectations. Some authors have proposed a model in which service quality depends on the difference between these two quantities (Parasuraman et al., 1988, 1991, 1994a, 1994b; Zeithaml et al., 1990) whereas others have advocated models in which expectations play a different role (Cronin and Taylor, 1992, 1994; Teas, 1993, 1994). A survey of approaches to service quality can be found in Morgan and Murgatroyd (1994).

Perceived quality is usually a function of several quality factors, dimensions or attributes, and a key step in the measuring of service quality is determining the relative weight of each factor or attribute. Methods oriented to multidimensional quality measurements are usually based on Conjoint Analysis (Luce and Tukey, 1965) . See Carroll and Green (1995) for a recent survey of the present state of this methodology and Lynch et al (1994) , Wedel and DeSarbo (1994)

and Ostrom and Iacobucci (1995) for interesting applications to the evaluation of service quality. In these procedures the customers are given several hypothetical services defined by certain levels of the quality attributes and are asked to provide quality evaluation or preferences for these services. The method assumes that the quality attributes can be given an objective interpretation and then the levels of the attributes have, when presented to the customers for evaluation, a clear meaning to them.

However, these methods are less useful in situations in which the quality attributes do not have objective standards, and therefore it is very difficult to create a series of hypothetical quality situations for the customers to evaluate. This paper presents a procedure to determine a linear indicator quality index that can be applied to any type of quality attributes. The paper is organized as follows. Section 2 presents the role of linear quality indicators within an operational definition of service quality. Section 3 discusses the problems with previous approaches to estimate the weights in linear quality indicators. Section 4 presents a new model for the quality evaluations, its main hypothesis and the estimation of the weights in the population. Section 5 describes an application of this model to measuring the quality of education in a Spanish university. Section 6 applies this model to the measurement of the railroad service quality in Spain. The technical details of the model are presented in the appendix.

LINEAR QUALITY INDICATORS

Suppose that we have a population of customers. This population includes our present customers but it can also includes the future or potential customers. We assume that the ith customer from this population has an evaluation Q_i for the perceived quality of a given service. This evaluation will be the result of some comparison of his/her expectations about the service and the perceived quality. This customer's evaluation will be a function of several attributes $X_1, \ldots, X_k$ which determined the global evaluation of the service. Let us call $X_{i1},\ldots, X_{ik}$ to the evaluations of these attributes made by the ith customer. Then,

$$Q_i = f(X_{i1},\ldots X_{ik}) \tag{2.1}$$

We define the *service quality* as the expected value of Q_i in the customer's population

$$Q = E[Q_i] = \frac{\sum_{i=1}^{N} Q_i}{N} \tag{2.2}$$

where we assume that the size of the customer's population, N, is large. A linear quality indicator assume that the function (2.1) can be approximated by

$$Q_i = \sum_{j=1}^{k} w_{ij} X_{ij} \tag{2.3}$$

where the weights w_{ij} are measuring the relative importance of attribute X_j to the quality of the service for the ith customer. Assuming that all the customer evaluations, X_{ij} and Q_i, are made in the same scale, for instance (0,10), the weights w_{ij} can be taken as :

(1) $w_{ij} \geq 0$ (2.4)

(2) $\sum_{j=1}^{k} w_{ij} = 1$ (2.5)

We also assume that the weights w_{ij} used by the ith customer for the j attribute are independent of the evaluation made by this customer for this attribute X_{ij}. As explained before, the evaluation of an attribute represents how the performance of the service in this attribute compare to an ideal or standard performance. For instance, suppose that the service is a restaurant and the attribute is the speed of the service measured by the time the customer has to wait to receive his/her order. Then, the evaluation of the waiting time depends on previous experiences of the customer on similar situations and will normally depend on the type of restaurant. We assume that the evaluation of this attribute in a particular restaurant is independent from the importance that the speed in the service has in his/her judgement of the quality of the service. The latter determine what we want and the former depends on what we get.

The service quality can be readily be obtain from equations (2.2) and (2.3) using the independence of the variables w_{ij} and X_{ij}. This quality service will be

$$Q = E[Q_i] = \sum_{j=1}^{k} E[w_{ij}]\, E[X_{ij}]$$

$$= \sum_{j=1}^{k} w_j\, m_j \tag{2.6}$$

where we have called m_j to the average evaluation of attribute jth in population, and w_j the average weight of this attribute in this population.

Fitting a model like (2.6) implies estimating the mean weights of each attribute in the population, and the average of the evaluations for the attributes. The advantages of working with a model of this type are as follows:

1) a quality index allows comparing our service to the one provided by other companies and can reveal our relative strengths and weaknesses.

(2) the knowledge of the weights allows the ordering of the attributes according to their relative importance to the customer, and shows the key factors in order to improve quality.

(3) the customers can be segmented by their weighting function, obtaining a market segmentation function directly linked to our quality objectives.

(4) if the evaluations of the attributes is related to some objective measures of performance it is possible to substitute the subjective evaluations for these objective measurements allowing a simple monitoring of the quality index.

The operational definition of service quality presented in this section has some limitations. We may have a good service quality on average but a very bad service quality for some groups of customers. This may happen either because some segments of the customers have a very different weighting function for the quality attributes or because they have a different evaluation of the attributes. These two situations should be identified because we can provide a better service if we identify clusters of customers with different values or opinions about quality. Then, it is more informative to measure service quality in these different populations. It must be remembered that the mean is only a good descriptive measure when we have an homogeneous sample and that it can be very non representative when the data comes from a mixture of very different populations.

THE PROBLEM OF ESTIMATING THE ATTRIBUTES WEIGHTS

Several authors have recommended to estimate the weights in a linear indicator of quality like (2.3) by asking directly to the customers. For instance, Zeithamel et al.(1990) in their conceptual model of service quality SERVQUAL identify five attributes of service quality (see Table 1) and determine the weights of these attributes by asking to a sample of customer to distribute 100 points among these five attributes. the result of this evaluation is given as column M3 in Table 1. They also asked to people in the sample for the relative importance of each attribute in a 0-10 scale (see the first column of Table 1). Finally, they compute the percentage of answers in which each attribute was considered the most important for the previous evaluations. We can use these three pieces of evidence to compute the weights, and the results for the SERVQUAL model are presented in columns M1 to M3 of Table 1.

Table 1. Computation of weights by different methods for the SEVQUAL model.

Attribute	Importance	M1	M2	M3
Tangibles	8.56	18.68	1.1	11
Reliability	9.44	20.60	42.1	32
Responsiveness	9.34	20.38	18.0	22
Assurance	9.18	20.03	13.6	19
Empathy	9.30	20.30	25.1	16

The first column of the Table presents the importance of the attribute in a 0-10 scale given directly by people in the sample. The second includes the weights derived from a procedure that will be called from now on M1. This procedure uses the result of dividing the importance of the attribute in the first column of Table 1 by the sum of the Importance of the five attributes. For instace, 18.68= 8.56 / (8.56+9.44+9.34+9.18+9.30) . The next column corresponds to a method that will be called M2, in which the weights are taken equal to the percentage of answers that indicated that the attribute is the most important. The fourth corresponds to a method that will be called M3, in which the customers estimate directly the weights. The main conclusions from this Table are: (1) procedure M1 leads to a similar and almost uniform weighting for all the attributes; (2) procedure M2 leads to a very asymmetric distribution of weights; (3) procedure M3 leads to values that are approximately half way between the other two.

We have also found in the cases considered in sections 4 and 5 that procedure M1 leads to a very close to uniform weighting of the attributes. This effect increases with the number of attributes. For instance, suppose that we consider only the two first attributes of Table 1, which have importance evaluation of 8.56 and 9.44. Then the weights obtained by M1 will be 47.5 and 52.5, with a difference of 5 points. If now we add the three other attributes, the difference of weights is reduced to 1.5. In the limit, if we add more and more attributes, the distribution of weights tends to be uniform.

The asymmetric distribution from procedure M2 is to be expected. For instance, consider the case in which all the members of the population agree that attribute 1 is the most important, 2 is also important and all the others are not. Then we will get a weight of 100 for attribute 1 and zero for the rest that is clearly unsatisfactory. A modification of this method will be to assign a rank order to the attributes, obtain the mean of these orders and try to use these mean rank to build weights. The problem with this procedure is again that it does not take into account that a rank scale will not define well in general an interval or continuous scale for the weights. The difference in importance, and therefore in weights, between the 1st and the 2nd attribute will not be in general the same than the one between attributes 3rd and 4th.

The third procedure has the problem that many customers do not have the habit of making these kind of assessment and the results may be very unreliable. In my experience, the weights obtained in this way have a large variability and the same person can give a different

distribution of weights depending of the order of the presentation. Also this procedure is very difficult to execute when the number of attributes is large.
In the next section we present a different procedure to compute the weights that has several advantages. First, it can be applied for any number of attributes. Second, it is based on the kind of evaluations customers are more familiar with. Third, it is based on a statistical model for the population and can be tested and checked with the data. Fourth, the procedure seems to work well in all the cases we have considered.

A MODEL FOR THE WEIGHTS AND ITS ESTIMATION

We assume that a random sample of size n from the customers' population has provided evaluations for the global service quality, y_i, $(i=1,...,n)$ as well as evaluations of the attributes that determine the quality of the service, x_{ij}, for certain well defined attributes X_j, $(i=1,...,n; j=1,...k)$. From now on and without loss of generality we assume that this evaluations are done in a 1-10 scale. The following hypothesis are made :

(H1) The quality of the service for the ith customer is an unknown variable that is measured in an imperfect way by the evaluation y_i provided by him/her. This means that the evaluation y_i is related to the service quality Q_i by

$$y_i = Q_i + u_i \qquad (4.1)$$

where u_i is the measurement error which we assume follows a normal $N(0, \sigma_u^2)$ distribution among the population of customers.

Note that in this assumption the variable u_i includes all the factors which determine that the same customer asked about the quality of the service may give different evaluation in different moments of time. It also includes the error due to the scale of measurement. This variable will change from customer to customer, but assuming that is due to many independent factors we may suppose, by the central limit theorem, that follows a normal distribution in the population. We also assume that this random error has a zero mean, that is, there are not systematic biases on the evaluation and that the variability is roughly the same for all customers in the population. We believe that H1 is quite general and it can be considered to hold in most situations.

(H2) The evaluation $\mathbf{x}_i = (x_{i1},...,x_{ik})$ made of the service quality attributes by the ith customer is made without error.

This assumption will be approximately true when the error in evaluating the attributes are small compared to the error in evaluationg the service quality. In practice there will always be some measurement error that, besides, can be different for different attributes. However, we assume this hypothesis as a first approximation and for simplicity. Dropping it increases very much the technical complication of the model because then it is transformed into an error in variables model.

(H3) The service quality Q_i is a linear function of the attributes $\mathbf{X}_i$, as

$$Q_i = \mathbf{w}_i' \mathbf{x}_i \tag{4.2}$$

where $\mathbf{w}_i = (w_{i1},\ldots, w_{ik})'$ are the weights defined by (1.4) and (1.5). As explained is the previous section we assume that the variables weights, w_{ij}, and attribute evaluations, x_{ij}, are independent variables.

The assumption of linearity is a strong one, but it can be tested after estimating the model. In some cases some nonlinear effects and interaction between attributes are expected. For instance, a bad performance in two important attributes can lead to a lower perceived quality that the one implied by adding up the effects of each attribute. Then we say that there is interaction between these two attributes and this feature should be included in the model as a product term. Again this hypothesis can be tested when the model is estimated.

(H4) The weights w_{ij} are random variables in the customer's population and follow a normal distribution with expected value w_j and variance σ_w^2, that is the same for the k attributes.

This hypothesis is rather restrictive because the variability of the weights in the population will be in many cases different for some attributes. It can be eliminated but again the complexity of the model increases. In this paper we present the analysis in this basic case.

With these four assumptions the distribution of the random variables y_i (i = 1,...,n) given the **X** will be normal with mean

$$E(y_i) = E(Q_i) = \mathbf{w}'\mathbf{x}_i \tag{4.3}$$

and, as shown in the appendix, the variance, $\sigma_i^2=\sigma_u^2(\theta k s_i^2+1)$, depends on the variability of the evaluations made by the ith customer, s_i^2, the measurement error σ_u^2, and the ratio $\theta = \sigma_w^2/\sigma_u^2$ between the common variability of the weights in the population and the measurement error. Therefore we have a heterokedastic regression model subject to linear restriction over the parameters $\hat{w}$. It is shown in appendix 1, using maximum likelihood methods, that the estimation of parameters $\hat{w}$, the vector of mean weights in the population, is given by

$$\hat{w} = \hat{w}_0 - (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1} \mathbf{1}\, a\, (\mathbf{1}' \hat{w}_0 - 1)$$

where **1** is a vector k-dimensional of ones, the constant is a scaling factor equal to $(\mathbf{1}'(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{1})^{-1}$, so that the restriction is verified. Note that if we multiply (4.4) by 1' we obtain that $\mathbf{1}'\hat{w} = 1$. The value $\hat{w}_0 = (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D} \mathbf{Y}$ is the generalized least squares estimate without restrictions, and D is the variance covariance matrix of the observations y_i that is given in the Appendix. Equation (4.4) shows that if $\hat{w}_0' \mathbf{1} = 1$ no correction is applied. Otherwise, the estimator is corrected to fulfill this restriction. Note that this estimator is different from the trivial one, $\hat{w}_0/(1'\hat{w}_0)$, that will be obtained by dividing each component of $\hat{w}_0$ by the sum

of all the components in order to fulfill the restriction. In (4.4) each component of $\hat{w}_0$ is corrected by an amount that depends on its variance and its covariances with the other components, as measured by the matrix $(\mathbf{X'DX})^{-1}$.

Using this estimator, and assuming that θ is known, σ_u^2 can be estimated, as shown in the appendix, by

$$\hat{\sigma}_u^2 = \frac{1}{n}\Sigma\frac{(y_i - \hat{w}'x_i)^2}{\theta k s_i^2 + 1}. \tag{4.5}$$

Usually θ is unknown, and the Appendix presentes a method to estimate it. However, in some applications we have a priori a set of possible values for this parameter. Then, the simplest way to deal with it is to compute the residual variance (4.3) with different values of this parameter and take as estimate the value that minimizes (4.3). Note that the estimate (4.2) depends also on θ, and so the procedure requires to compute first (4.2) for each value of the parameter and then compute (4.3). The advantage of this procedure is that it does not require a special software and it can be carried out with any standard statistical package that includes weighted least squares. In the examples presented in the next sections θ has been set equal to .01 and a sensitivity analysis has been carried out to check if the results depend on the θ value assumed. In all cases we have found that the result are quite robust to the particular value chosen in the interval (.05 to .001).

Note that we can use most of the standard regression methods to check the validity of this random coefficients model. In particular the restriction of the weights adding up to one can be tested by comparing the constrained and the unconstrained estimates. The key assumption of equal variances in the distribution of the weights can also be tested by estimating a model without these restrictions.

A point of special interest is determining groups of customers with different weighting structure. For instance, sometimes the distribution of customers weights can be thought of as a mixture of two or more distributions corresponding to two or more different type of customers. This should be taken into account to avoid serious misspecification errors in the model. For instance, a small set of customers with evaluations very different from the others may determine completely the weighting function if they have more extreme (either good or bad) evaluations that the bulk of the other customers. See Peña and Yohai (1995) for an analysis of this problem. Wedel and DeSarbo (1994) have developed procedures to deal with this problem

MEASURING THE QUALITY OF EDUCATION

This methodology was applied to measure the quality of the education at the universidad Carlos III de Madrid. The University ask the students every semester to make judgments with respect to the courses they have followed in the semester and the teachers that have taught them. The questionnaire they fill was obtained from a factor analysis on a large battery of questions that were made to the students (Peña, 1995). In this preliminary study the five factors presented in Table 1 were identified. The students made evaluations of their overall satisfaction with the

instructor and then evaluate each of these five factors. We have used data from academic year 95/96 from the two semesters to estimate the relative weight of the attributes in determining the global student satisfaction. The sample includes 7, 253 students of the Faculty of Social Science of the Universidad Carlos III de Madrid.

We have used data from the second semester of the academic year 94/95 to estimate the relative weight of the attributes in determining the global student satisfaction. The samples includes 7497 students of the Faculty of Social Science of the Universidad Carlos III de Madid, and we have the overall evaluation of the instructor as well as the evaluation of the five dimension or attributes presented in Table 2. With this information we have estimated the model described in section 4. That is, we regress (by generalized least squeares) the overall satisfaction with the instructor to the five attibutes and use equation (4.4) to determine the implicit weights in the answer of the students. Note that, as explained in the appendix, the linear restriction over the weights is taken into account in the estimation. The five variables have significant t values and the R^2 of the regression was .73.

Table 2. Weights of teaching attributes and t-values for students in Universidad Carlos III de Madrid

	w	t
1. Class organization and clarity in lecturing	.58	65.9
2. Enthusiasm and interest	.30	30.7
3. Stimulate classroom participation	.06	7.6
4. Usefulness of teaching material, books and notes	.04	5.9
5. Punctuality	.02	3.0

Table 2 shows that the most important attribute to explain the overall instructor evaluation if the dimension of class organization and clarity in lecturing, that has a weight of 58%. The dimension enthusiasm and interest shown by the instructor is another key attribute to determine students' satisfaction with teaching and has a weight of 30%. The other three attributes have a much maller importance.

The overall students' satisfaction with their education depend on their evaluation of teaching but also on some other factors that affect the quality of their education. We had a group discussion in two classrooms of business students, that were following an elective quality course, and as a result other eight attributes were judged to be potentially important in the quality of the education. These attributes are presented in Table 3 together with their weights estimated by the model explained in section 4 for in four different samples.

The first column correspond to the sample from classroom A which includes data from 58 students. The third corresponds to classroom B, with 67 students. Table 3 shows that the estimated weights are very different in the two classrooms. In classroom A it is surprising the small weight of the teaching evaluations and the large weight of the library. Analyzing these data with more detail we found that the low value for the teaching evaluation is due to 3 atypical

data points. Two of these three students gave the same evaluation for his/her global satisfaction (6 in a 0-10 scale) but were very different in their attributes evaluation. For instance, in the first two attributes the first put (9,9), the second (1,2). The third student shows up specially for the high evaluation of the library service: 10 out of ten, whereas the global evaluation was 5. The second column of the Table 3, the one labeled A^*, presents the weights estimated from 55 data from classroom A in which the three influential points have been deleted. It can be seen that then the weights in the two groups are more similar. Finally, the forth column correspond to the complete sample of students from both classrooms.

This example shows the importance of a careful analysis of the data to identify customers that have a different value system and, as a consequence they have a different weighting scheme. Although these points can be identified using influence measures (see Cook and Weisberg, 1982) and residual analysis, it would be very convenient to have automatic methods that produce a segmentation of the sample into homogeneous sets. Procedure for finding groups of outliers in regression as the one developed by Peña and Yohai (1995) may be useful, but research is needed in how to carry out this computation in a simple an efficient way.

Table 3. Weights of quality of education attributes for two groups of Business students. The weights have been multiplied by 100 so that they add up to 100.

	A	A^*	B	TOTAL
1. Teaching Evaluations	.18	.33	.66	.54
2. Fairness of the exams	.21	.14	.12	.13
3. Quality of Administrative services	.00	.00	.00	.00
4. Quality of the library	.35	.36	.08	.19
5. Quality of Computer rooms	.00	.00	.00	.00
6. Campus Facilities	.00	.00	.00	.00
7. Teaching Organization (schedule of classes,exams,..)	.06	.05	.06	.02
8. Student participation in decisions	.00	.00	.00	.00
9. Sport and cultural activities	.18	.10	.07	.12

MEASURING THE QUALITY OF THE RAILROAD SYSTEM IN SPAIN

The previous model is applied to build a quality index of the railroad system in Spain. The procedure to build this index can be summarized as follows: (1) identifying the quality attributes, (2) taking a random sample of 4000 customers and obtaining their evaluations; (3) estimating the parameters of the model and making segmentation of the customers by their quality evaluations, (5) controlling the quality of the service.

Starting with the first step, the identification of the attributes was the result of several group sessions with customers of RENFE (the public railroad system in Spain). Initially 52 attributes were identified. 28 out of these 52 attributes correspond to the pre-journey (information, ticket office, railroad station), 20 to the journey and 6 to the post-journey (claims and so on). A small random sample of 165 customers was taken to check these attributes and as a result of the statistical analysis of the questionnaire the number of attributes was reduced to 25 : 10 for the pre-journey, 12 for the journey and 3 for the post-journey.

In order to obtain a representative sample of the railroad service random sample by train type and day of the week was de designed. The sample size was 4 000, and 2 600 customers werw selected from long distance whereas 1 400 were for short distance traveling. We will report here, for the sake of brevity, only the results for the long distance pasasengers. The interviews were made during the journey by train. The passengers were requested: (1) to evaluate in a 0-10 scale the importance of these 25 attributes ; (2) to evaluate in a 0 10 scale the real situation of this quality attribute in RENFE and to give an overall evaluation of the service.

The mean attribute importance obtained by the first method show a small variability, as shown in Table 4. It was found a slightly higher value for main train characteristics and smaller values for the pre-journey aspects. If we try to use these evaluations to build weights by dividing each attribute mean importance by the sum of all the others we obtained that the weights vary between a maximum of 4.2% and a minimum of 3.75%. The mean weight of each attribute is presented in Table 4, in which two examples are also given of the type of attribute that is included in each bracket of values. We believe that the small variability of the weights does not correspond to the truth and it is due to the difficulty which customers have of expressing in an abstract way their preferences.

Table 4. Importance scores of customers in Spain

Braket of values	Number	Attributes
7.1 and 7.2	4	Information pre-journey, train cafeteria
7.3 and 7.4	3	Ticket offices, room in station
7.5 and 7.6	8	Signs in station, train speed
7.7 and 7.8	6	Toilets cleanness in station, train security
7.9 and 8.0	4	Punctuality, train cleanness

Applying the method developed in previous section to the attribute evaluation and to the overall evaluation we obtained the result presented in Table 5. Only 14 out of the 2 attributes seems to affect the global evaluation of the services and now their weights vary between 10.4% and

2.1%. All the coefficients given in Table 5 were significant (with t values going from 3.0 to 7.3) and the R^2 of the regression was .623.

Table 5. Weights of quality attributes for the railroad service system in Spain determined by the constrained regression model.

Attribute	Mean weight	Intercity	Express
Cleanness in the train	10.4	11.9	9.4
Comfort	10.3	7.8	11.5
Punctuality	10.3	10.7	14.9
Speed	9.5	8.5	7.9
train security	8.6	10.0	7.7
Cancellations	8.0	7.7	6.3
Claims	7.8	7.5	6.1
Responsible person in the train	7.5	9.8	7.3
Cleanness station	6.7	6.5	5.2
Information in station	6.3	5.8	4.0
Frequency	6.1	6.5	9.0
Ticket offices	3.9	3.0	4.5
Pre-Journey Information	2.5	2.8	4.3
Cafeteria	2.1	1.5	1.7

It should be stress that the second column in Table 5 represents mean values (as estimated by regression) for long distance passengers. A cluster analysis showed that the customers have a very different preference structure when they are stratified by train type. For instance, columns three and four represent the weights for Intercity trains (high quality trains usually faster than the average train which link large cities) and for express (a medium quality train which transports the mail and travels at night). It can be seen that the weights depend on the type of train.

CONCLUSIONS

The Knowledge of the relative importance of quality attributes for customers is key for any process of sevice quality improvement. The procedure presented in this paper seems to be a useful way to estimate the implicit weights used by customers in their overall evaluation of service quality.

APPENDIX

In order to estimate the model we require the covariance matrix of the vector of variables $\mathbf{w}_i$. This vector will follow a multivariate normal singular distribution with expected value $\mathbf{w} = (w_1,\ldots, w_k)$ and covariance matrix

$$\Sigma_w = \sigma_w^2 \begin{bmatrix} 1 & -\frac{1}{k-1} & \cdots & -\frac{1}{k-1} \\ -\frac{1}{k-1} & 1 & \cdots & \cdots \\ \vdots & \vdots & & \\ -\frac{1}{k-1} & \cdots & \cdots & 1 \end{bmatrix} = \sigma_w^2 . A$$

where the A matrix is given by (1/(k-1))(kI - 1 1'), I is the identity matrix and 1 = (1,...,1) is a vector of ones. This covariance matrix can be easily obtained from the assumption that all marginal variables have the same variance, the same covariances and they add up to one. As :

$$(w_{i1} - w_1) + \ldots + (w_{ik} - w_k) = 0$$

taking the square of this expression and then expected values

$$k\,\sigma_w^2 + 2\binom{k}{2}\gamma_w = 0$$

where γ_w are the covariances between the attributes that are assumed to be equal. Then

$$k\,\sigma_w^2 + 2\,\frac{k(k-1)}{2}\,\gamma_w = 0$$

which implies that the covariances are equal to $-\sigma^2_w/(k-1)$, and the covariance matrix is obtained. Then, the distribution of the random variables y_i (i = 1,...,n) given the **X** variables in normal with expected value $E(y_i) = \mathbf{w}'\mathbf{x}_i$ and variance

$$\sigma_i^2 = \mathrm{Var}(y_i) = \mathrm{Var}(Q_i + u_i) = \mathbf{x}_i'\Sigma_w\,\mathbf{x}_i + \sigma_u^2.$$

that can be written, calling $\theta = \sigma_w^2/\sigma_u^2$, as

$$\sigma_i^2 = \sigma_u^2(\theta r_i + 1) \qquad (A.1)$$

where $r_i = \mathbf{x}_i'\,\mathbf{A}\,\mathbf{x}_i$. Using the expression for A, $r_i = (1/(k-1))(k\,\mathbf{x}_i'\mathbf{x}_i - (\mathbf{x}_i'1)(1'\mathbf{x}_i))$ and calling $\hat{x}_i = (1'\mathbf{x}_i)/k$ to the average of evaluations for the ith customer, we have

$$r_i = \frac{k^2}{k-1}\left(\sum\frac{x_{ij}^2}{k} - \bar{x}_i^2\right) = k\, s_i^2$$

where s_i^2 is the variance of the evaluation of the k attributes by the ith customer. The estimation requires that the likelihood must be maximized with the restrictions $\sigma_w^2 \geq 0$, $\sigma_u^2 \geq 0$, $\mathbf{w} \geq 0$, $\mathbf{w}'\,\mathbf{1} = 1$. Including a lagrange multiplier for the linear restriction, the likelihood is

$$M(w, \sigma_u^2, \theta, \lambda) = -\frac{n}{2}\log\sigma_u^2 - \frac{1}{2}\sum\log(\theta r_i + 1) - \frac{1}{2\sigma_u^2}\sum\frac{(y_i - w'x_i)^2}{(\theta r_i + 1)} + \frac{\lambda}{2}(w'1 - 1) \quad \text{(A.2)}$$

The parameter values that maximize this function are obtained by

$$\frac{\partial M}{\partial w} = -\sum\frac{(y_i - \hat{w}'x_i)}{\theta r_i + 1}x_i + \lambda\, 1 = 0 \quad \text{(A.3)}$$

$$\frac{\partial M}{\partial \lambda} = 0 = 1'\hat{w} - 1 \quad \text{(A.4)}$$

$$\frac{\partial L}{\partial \sigma_u^2} = -\frac{n}{2}\frac{1}{\sigma_u^2} + \frac{1}{2}\frac{1}{\sigma_u^4}\sum\frac{(y_i - w'x_i)^2}{\theta r_i + 1} \quad \text{(A.5)}$$

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2}\sum\frac{r_i}{1 + \theta r_i} + \frac{1}{2}\sum\frac{r_i (y_i - w'x_i)^2}{\sigma_u^2 (\theta r_i + 1)^2} \quad \text{(A.6)}$$

The solution of these equations requires a nonlinear optimization algorithm. Note that the structure of the system is simple because we can fix θ and determine $\hat{w}$ and $\hat{\sigma}_u$. Then we compute a new value for θ, which will lead to new estimates for $\hat{w}$ and $\hat{\sigma}_u$, and so on.

Let us consider the estimation of $\hat{w}$ given σ_u^2 and θ. Then, calling $\mathbf{Y}$ to the vector of observations $(y_1, \ldots y_n)$, $\mathbf{X}$ to the attributes evaluation matrix and $\mathbf{D}$ to the diagonal matrix with elements $(\theta r_i + 1)^{-1}$, we can write (A.3) as

$$-\mathbf{X'\,D\,Y} + \mathbf{X'D\,X}\,\hat{w} + \lambda\,\mathbf{1}\,\sigma_u^2 = 0.$$

Then, $\hat{w} = (\mathbf{X'\,D\,X})^{-1}(\mathbf{X'\,D\,Y} - \lambda\,\sigma_u^2\,\mathbf{1})$, and $\mathbf{1'}\,\hat{w} = \mathbf{1'}\,(\mathbf{X'\,D\,X})^{-1}(\mathbf{X'\,D\,Y} - \lambda\,\sigma_u^2\,\mathbf{1})$.

For (A.4) λ is given by

$$\lambda = \sigma_u^{-2}\,(\mathbf{1'}\,(\mathbf{X'\,D\,X})^{-1}\,\mathbf{1})^{-1}\,(\mathbf{1'}(\mathbf{X'\,D\,X})^{-1}\,\mathbf{X'D\,Y} - 1) \quad \text{(A.5)}$$

and calling $a = \sigma_u^{-2} (1'(X'D\,X)^{-1}\,1)^{-1}$, and using (A.5) we finally have.

(A.6)

$$\hat{w} = \hat{w}_0 - (X'\,D\,X)^{-1} 1\, a(1'\,\hat{w}_0 - 1)$$

where $\hat{w}_0 = (X'\,D\,X)^{-1}\,X'\,D\,Y$ is the unrestricted generalized least squares estimate. Note that this estimates do not depend on the scale parameter σ_u^2. Once a first estimate of $\hat{w}$ is obtained and given θ, σ_u^2 can be estimated by

$$\hat{\sigma}_u^2 = \frac{1}{n} \Sigma \frac{(y_i - \hat{w}' x_i)^2}{\theta\, r_i + 1} \tag{A.7}$$

and finally θ is estimated by solving:

$$\Sigma \frac{r_i + \theta\,(r_i^2 - e_i^2)}{(1 + \theta\, r_i)^2} = 0 \tag{A.8}$$

where $e_i = (y_i - w'x_i)/\sigma_u$ is the standardized residual. This last equation requires numerical methods. The values $\hat{\theta}$ will lead to a new value of $\hat{w}$ and the procedure is iterated until convergence.

In summary given θ the problem can be solved by generalized least squares. A simple procedure is to assume this parameter as known, and make afterwards a sensitivity analysis of the result to a set of possible values.

Acknowledgment

This work has been supported by the Cátedra BBV de Calidad, Universidad Carlos III de Madrid. The results presented for the Spanish railroad system are part of a larger study made for RENFE jointly with Jesus Juan and Teresa Villagarcía. Juan Antonio Gil was very helpful in computing some of the results in section 5. Alberto Maydeu and Pilar Rivera made many useful comments to a first draft of this paper.

References

Carroll, J. D. and Green, P. E., Psychometric Methods in Marketing Research: Part I, Conjoint Analysis. *Journal of Marketing Research.* XXXII, 1995, 385-391.

Cook R. D. and Weisberg, S, *Residuals and Influence in regression.* Chapman and Hall, 1982.

Cronin, J. and Taylor S., Measuring Service Quality: A Reexamination and Extension *Journal of Marketing*, 56, July, 1992, p. 55-68.

Cronin, J. and Taylor, S., SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality. *Journal of Marketing*, 58, January, 1994, p. 125-131.

Luce, D. R. and Tukey J. W.,Simultaneous Conjoint Measurement: A new type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1, 1964, p. 1-27.

Lynch, J. G., Bouzas T. E. y Berg, S. V., Regulatory Measurement and Evaluation of Telephone Service Quality. *Management Science*, 40, 2, 1994, p. 168-194

Ostrom A. and Iacobucci, D., Consumer Trade-Offs and the Evaluation of Services. *Journal of Marketing*, 59, 1995, p.17-28.

Morgan C. and Murgatroyd S. *Total Quality in the Public Sector*. Open University Press. Buckinham , UK, 1994.

Parasuraman, A. Zeithaml, V.A. and Berry, L.L., SERVQUAL:A multiple Item scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing* , 64, 1988, p. 12-40.

Parasuraman, A. Zeithaml, V.A. and Berry, L.L,. Refinement and Reassessment of the SERVQUAL Scale. *Journal of Retailing*, Vol. 67, 4, 1991, p. 420-450.

Parasuraman, A. Zeithaml, V.A.and Berry, L.L., Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research. *Journal of Marketing*, Vol. 58, 1994, p. 111-124.

Parasuraman, A. Zeithaml,V.A. and Berry, L.L, Alternative Scales for Measuring Service Quality: A Comparative Assessment Based on Psychometric and Diagnostic Criteria. *Journal of Retailing*, Vol. 70, N° 3, 1994, p. 201-230.

Peña, D. Experiencias de mejora de la Calidad en la Universidad. *Calidad*, Diciembre 1995.

Peña, D. And Yohai, V.,The detection of influential subsets in linear regression using an influence matrix". *Journal of the Royal Statistical Society B*, 57,1, 1995, p. 145-156.

Teas, R. K., Expectations, Performance Evaluation, and Consumers´ Perceptions of Quality. *Journal of Marketing*, Vol. 57, 1993, p. 18-34.

Teas, R. K., Expectations as a Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment. *Journal of Marketing*, Vol. 58, January, 1994,p. 132-139.

Wedel, M. and DeSarbo, W.,A Review of Recent Developments in Latent Class Regression Models. En BAGOZZI, R. (Editors) Advanced Methods of Marketing Research, I-II. Blackwell Publishers, Massachusetts, 1994, p. 353-388.

Zeithaml, V. A. Parasuraman, A. and Berry, L. L., *Delivering Quality Service* Macmillan Inc. 1990.